A COMBINATORIAL LEMMA AND ITS APPLICATION TO PROBABILITY THEORY

BY

FRANK SPITZER[1]

1. Introduction. To explain the idea behind the present paper the following fundamental principle is emphasized. Let $X=(X_1, \cdots, X_n)$ be an n-dimensional vector valued random variable, and let $\mu(x)=\mu(x_1, \cdots, x_n)$ be its probability measure (defined on euclidean n-space E_n). Suppose that X has the property that $\mu(x)=\mu(gx)$ for every element g of a group G of order h of transformations of E_n into itself. Let $f(x)=f(x_1, \cdots, x_n)$ be a μ-integrable complex valued function on E_n. Then the expected value of $f(x)$ is

$$Ef(X) = \int f(x)d\mu(x) = \int \bar{f}(x)d\mu(x), \tag{1.1}$$

where

$$\bar{f}(x) = \frac{1}{h} \sum_{g\in G} f(gx). \tag{1.2}$$

This principle will be fruitful when it is possible to write $\bar{f}(x)$ in a form which is simpler to integrate than $f(x)$.

We shall consider only the case when G is the symmetric group of permulations σ on n symbols, so that

$$\sigma x = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma_1 & \sigma_2 & \cdots & \sigma_n \end{pmatrix} x = (x_{\sigma_1}, x_{\sigma_2}, \cdots, x_{\sigma_n}).$$

Vector random variables $X=(X_1, \cdots, X_n)$ with the property that $\mu(x)=\mu(\sigma x)$ for every permutation are called symmetrically dependent and have been treated at length by E. Sparre Andersen [1; 2; 3]. For the most interesting applications of the present theory it will be necessary to require in addition that $X_1, \cdots, X_n$ be identically distributed and independent.

For the function $f(x)$ in equation (1.1) we shall take

$$f(x) = \max\,[0, x_1, x_1 + x_2, \cdots, x_1 + x_2 + \cdots + x_n].$$

The fundamental combinatorial result of this paper (Theorem 2.2) identifies

Presented to the Society, September 2, 1955 under the title *On the maximum of partial sums of random variables*; received by the editors September 2, 1955.

(1) The research of this author was supported in part by the United States Air Force under Contract No. AF18(600)-685 monitored by the Office of Scientific Research.

the set of numbers $[f(\sigma x)]$, as σ ranges over G, with a set of numbers $[g(\sigma x)]$ which have very desirable properties with regard to integration. The author is indebted to H. F. Bohnenblust for valuable discussions. The proof of Theorem 2.2 is due to him.

When applied to the case of independent identically distributed random variables $X_1, X_2, \cdots$ and their partial sums $S_k = X_1 + \cdots + X_k$, Theorem 2.2 and the principle of equation (1.1) produce Theorem 3.1:

For $|t| < 1$,

$$\sum_{n=0}^{\infty} \phi_n(\lambda) t^n = \exp\left[\sum_{n=1}^{\infty} \frac{\psi_n(\lambda)}{n} t^n\right], \tag{1.3}$$

where $\phi_n(\lambda)$ *is the characteristic function of* max $[0, S_1, S_2, \cdots, S_n]$, *and* $\psi_n(\lambda)$ *the characteristic function of* max $[0, S_n]$. Equation (1.3) generalizes results of M. Kac and G. A. Hunt (Theorem 4.1 in [7]), and of E. Sparre Andersen (Theorem 1 in [3]). In §6 equation (1.3) is generalized in Theorem 6.1 which gives the joint characteristic function of S_n and max $[0, S_1, \cdots, S_n]$.

§§4, 5, and 7 contain a number of new results concerning the limiting behavior of the random variables max $[0, S_1, \cdots, S_n]$ and N_n = the number of positive S_k, $k = 1, 2, \cdots, n$. These results are chosen to illustrate the power of the combinatorial method, without in any way exhausting its possibilities. The case of sums of nonidentically distributed independent random variables clearly is beyond the scope of this method. The same remark applies to continuous parameter stochastic processes and consequently the present method is a natural one for the class of processes with stationary independent increments.

2. Combinatorial considerations.

The following proposition will be very useful.

THEOREM 2.1. *Let* $x = (x_1, \cdots, x_n)$ *be a vector such that* $x_1 + x_2 + \cdots + x_n = 0$, *but no other partial sum of distinct components vanishes. Let* $x_{k+n} = x_k$, *and* $x(k) = (x_k, x_{k+1}, \cdots, x_{k+n})$, $k = 1, 2, \cdots, n$. *Then, for each* $r = 0, 1, \cdots, n-1$, *exactly one of the cyclic permutations* $x(k)$ *of* x *is such that exactly* r *of its successive partial sums are positive.*

Proof. Let $s_k = x_1 + x_2 + \cdots + x_k$, $s_0 = 0$, $s_{k+n} = s_k$. Then the successive partial sums of the components of $x(k)$ are $s_k - s_{k-1}, s_{k+1} - s_{k-1}, \cdots, s_{k+n-1} - s_{k-1}$. By assumption the s_k are all distinct for $k = 1, 2, \cdots, n$, so that the number r of positive terms in the above sequence equals the number of positive terms among $s_j - s_{k-1}$, $j = 1, 2, \cdots, n$. Hence r may be given any value between 0 and $n-1$ in one and only one way, i.e. by choosing k so that s_{k-1} is $(r+1)$st from the top in order of magnitude among the s_k, $k = 1, 2, \cdots, n$.

One can obtain an interesting version of Theorem 2.1 which does not require the assumption that $s_k=0$. We call the polygon connecting the points $(0, 0)$, $(1, s_1), \cdots, (k, s_k), \cdots, (n, s_n)$ the sum polygon of the vector x. The sum polygon for the cyclically permuted vector $x(k)$ is defined the same way. Then the numbers $s_k-(k/n)s_n$ represent the vertical distances from the vertices of the sum polygon to its chord, the line connecting $(0, 0)$ with (n, s_n). Now we may apply Theorem 2.1 to the numbers x_k-s_n/n, whose sum vanishes, if they satisfy the incompatibility assumption. (It is clearly sufficient that the x_k be rationally independent.) The result is that if we consider a sum polygon and its n cyclic permutations, and prescribe an integer r between 0 and $n-1$, then exactly one of the cyclic permutations of the sum polygon will have the property that exactly r of its vertices lie strictly above its chord.

The geometric meaning of this theorem will provide the clue to Theorem 2.2. Its proof depends on the possibility of finding a unique cyclic permutation of certain subsets of components, such that their sum polygon (the polygon connecting the points (k, s_k)), lies entirely below its chord. Rational independence of the components will be assumed there.

Now we introduce certain notations. For any real a

$$a^+ = (|a| + a)/2 = \max [0, a],$$

so that

$$\max [0, a_1, \cdots, a_n] = \max_{1\leq k\leq n} \overset{+}{a_k}.$$

$$\sigma x = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma_1 & \sigma_2 & \cdots & \sigma_n \end{pmatrix} x = (x_{\sigma_1}, x_{\sigma_2}, \cdots, x_{\sigma_n}), \tag{2.1}$$

$$s_k(\sigma x) = x_{\sigma_1} + x_{\sigma_2} + \cdots + x_{\sigma_k},$$

$$S(\sigma x) = \max_{1\leq k\leq n} \overset{+}{s_k}(\sigma x) = \max_{1\leq k\leq n} \left(\sum_{i=1}^{k} x_{\sigma_i}\right)^+. \tag{2.2}$$

Consider the permutation τ represented as a product of cycles, including the one-cycles, and with no index contained in more than one cycle. For example suppose that $n=7$, and that

$$\tau = (14)(2)(3756). \tag{2.3}$$

Then we define

$$T(\tau x) = (x_1 + x_4)^+ + \overset{+}{x_2} + (x_3 + x_7 + x_5 + x_6)^+. \tag{2.4}$$

In formal notation, let

$$\tau = (\alpha_1(\tau))(\alpha_2(\tau)) \cdots (\alpha_{n(\tau)}(\tau)), \tag{2.3'}$$

where the $\alpha_i(\tau)$, $i=1, 2, \cdots, n(\tau)$, are disjoint sets of integers whose union is the set $[1, 2, \cdots, n]$. Then define

$$T(\tau x) = \sum_{i=1}^{n(\tau)} \left(\sum_{k \in \alpha_i(\tau)} (\tau) x_k \right)^+ . \tag{2.4'}$$

Now it is claimed that

THEOREM 2.2. *For an arbitrary fixed vector $x=(x_1, x_2, \cdots, x_n)$ the sets $[S(\sigma x)]$ and $[T(\tau x)]$, which are generated by letting σ and τ run through all of the $n!$ permutations, are identical sets.*

Proof. If the theorem is proved for a set of x which is dense in E_n, then its truth follows for arbitrary x, since the numbers $S(\sigma x)$ and $T(\tau x)$ are continuous functions of x. Therefore the proof is given for an arbitrary fixed x with rationally independent components. (For rational r_i, $r_1x_1+r_2x_2+ \cdots +r_nx_n=0$ if and only if each $r_i=0$.)

It is planned to exhibit between permutations of the form σ, defined in (2.1), and permutations of the form τ, defined in (2.3′) a one to one correspondence $\sigma_x(\tau)$. This mapping will depend on x and will have the property that, for each τ and for each x with rationally independent components,

$$T(\tau x) = S(\sigma_x(\tau)x). \tag{2.5}$$

The proof will then be complete.

Suppose a permutation τ is given in the form of (2.3′). The order of the indices in each set $\alpha_i(\tau)$ is then prescribed up to an arbitrary cyclic permutation. In accordance with the remarks following the proof of Theorem 2.1 we choose that unique cyclic permutation of the indices which ensures that the sum polygon lies below its chord. In terms of the example of equation (2.3),

$$\tau = (14)(2)(3756)$$

is rewritten as

$$\tau = (14)(2)(5637),$$

if it turns out that

$$x_1 < \frac{x_1 + x_4}{2}; \quad \max\left[x_5, \frac{x_5+x_6}{2}, \frac{x_5+x_6+x_3}{3}\right] < \frac{x_5+x_6+x_3+x_7}{4}. \tag{2.6}$$

In formal notation, suppose that the cycle $(\alpha_\nu(\tau))=(j_1, j_2, \cdots, j_k)$. Then without changing that cycle as a permutation we can rewrite it in one and only one way, as $(a_\nu(\tau))=(i_1, i_2, \cdots, i_k)$ so that

$$\max\left[x_{i_1}, \frac{x_{i_1}+x_{i_2}}{2}, \cdots, \frac{x_{i_1}+\cdots+x_{i_{k-1}}}{k-1}\right] < \frac{x_{i_1}+\cdots+x_{i_k}}{k}, \tag{2.6'}$$

where $(i_1, i_2, \cdots, i_k)$ is a cyclic permutation of $(j_1, \cdots, j_k)$. This is done for each cycle.

Finally, the cycles $\alpha_i(\tau)$ may be permuted among themselves. Again using the rational independence of the components of x, there is a unique permutation (relabeling) of the $\alpha_i(\tau)$, after which

$$(2.7') \qquad \frac{\sum_{k\in\alpha_1(\tau)} x_k}{\sum_{k\in\alpha_1(\tau)} 1} > \frac{\sum_{k\in\alpha_2(\tau)} x_k}{\sum_{k\in\alpha_2(\tau)} 1} > \cdots > \frac{\sum_{k\in\alpha_{n(\tau)}(\tau)} x_k}{\sum_{k\in\alpha_{n(\tau)}(\tau)} 1}.$$

Thus in our example τ should now be rewritten as

$$\tau = (5637)(2)(14),$$

if it turns out that

$$(2.7) \qquad \frac{x_5 + x_6 + x_3 + x_7}{4} > x_2 > \frac{x_1 + x_4}{2}.$$

Finally we define

$$\sigma_x(\tau) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 6 & 3 & 7 & 2 & 1 & 4 \end{pmatrix},$$

or in general

$$(2.8) \qquad \sigma_x(\tau) = \begin{pmatrix} 1 & 2 \cdots n \\ i_1 & i_2 \cdots i_n \end{pmatrix},$$

where the indices $i_1, \cdots, i_n$ are the integers $[1, 2, \cdots, n]$ in that unique order in which they now appear in the successive sets $\alpha_i(\tau)$.

It remains to define $\tau_x(\sigma)$ as a function of σ, and to show that $\sigma_x[\tau_x(\sigma)] = \sigma$, and finally that equation (2.5) holds. Given

$$\sigma = \begin{pmatrix} 1, & 2, & \cdots, & n \\ i_1, & i_2, & \cdots, & i_n \end{pmatrix},$$

consider the sum polygon through the points $(0, 0)$, $(1, x_{i_1})$, $\cdots$, $(k, x_{i_1} + \cdots + x_{i_k})$, $\cdots$, (n, s_n). Now we define the lowest convex majorant of the sum polygon as that unique polygon which goes through $(0, 0)$ and (n, s_n) in such a way that all its vertices are also vertices of the sum polygon and that it always lies above or coincides with the sum polygon. (Uniqueness follows from the fact that the x_i are rationally independent.) Suppose now that the lowest convex majorant constructed for the permutation σ has the vertices $(0, 0)$, $(k_1, x_{i_1} + \cdots + x_{i_{k_1}})$, $\cdots$, $(k_\nu, x_{i_1} + \cdots + x_{i_{k_\nu}})$, (n, s_n), where $0 < k_1 < \cdots < k_\nu < n$. Then we define

$$(2.9) \qquad \tau_x(\sigma) = (i_1, \cdots, i_{k_1})(i_{k_1}, \cdots, i_{k_2}) \cdots (i_{k_\nu+1}, \cdots, i_n).$$

It is geometrically obvious that $\tau_x(\sigma)$ is left unchanged by the transformations used in (2.6′) and (2.7′) to define $\sigma_x(\tau)$, so that $\sigma_x[\tau_x(\sigma)] = \sigma$, establishing the desired one to one correspondence.

Finally $S(\sigma x)$ is the maximum ordinate among those of the vertices of the sum polygon, but this maximum is clearly also attained at a vertex of the lowest convex majorant corresponding to σ. It follows from (2.9) that

$$S(\sigma x) = (x_{i_1} + \cdots + x_{i_{k_1}})^+ + \cdots + (x_{i_{k_\nu}+1} + \cdots + x_{i_n})^+ = T(\tau_x(\sigma)x),$$

or

$$T(\tau x) = S(\sigma_x(\tau)x),$$

which completes the proof.

In §5, we shall give a simple proof of a very surprising theorem of E. Sparre Andersen, which is stated in Equation (5.3). That theorem will be seen to follow from a combinatorial fact rather similar to Theorem 2.1.

Let $\theta(a)=1$ if $a>0$ and 0 otherwise. Let $x=(x_1, \cdots, x_n)$ be a given n-tuple as before, and let $\sigma x=(x_{\sigma_1}, \cdots, x_{\sigma_n})$ be a permutation of x. Let $\tau=(\alpha_1)(\alpha_2)\cdots(\alpha_{n(\tau)})$ be a permutation decomposed into cycles, and define

$$A(\sigma x) = \sum_{k=1}^{n} \theta(x_{\sigma_1} + \cdots + x_{\sigma_k}),$$

$$B(\tau x) = \sum_{i=1}^{n(\tau)} \left(\sum_{j \in \alpha_i} 1 \right) \theta\left[\sum_{j \in \alpha_i} x_j \right].$$

Then we have

Theorem 2.3. *The sets* $\{A(\sigma x)\}$ *and* $\{B(\tau x)\}$, *which are obtained by letting* σ *and* τ *run through all permutations on* n *objects, are identical sets.*

The proof is omitted, since it has recently been shown by H. F. Bohnenblust that far more general theorems than those considered here are easier to prove than the elegant but somewhat too special theorems considered here.

3. **The distribution of** max $[0, S_1, S_2, \cdots, S_n]$. Before proceeding to the applications of Theorem 2.2, it should be mentioned that even the simple Theorem 2.1 is not without probabilistic interest. When applied to equation (1.1) it immediately yields the following result of E. Sparre Andersen (Theorem 3 in [2]):

"*Let* $X_1, \cdots, X_{n+1}$ *be symmetrically dependent random variables and let* C *be an event which is symmetric with respect to* $X_1, \cdots, X_{n+1}$. *Let* N_n^* *be the number of points* (j, S_j), $j=1, \cdots, n$, *which lie above the straight line from* $(0, 0)$ *to* $(n+1, S_{n+1})$. *Then for* $\Pr[C]>0$,

$$\Pr[N_n^* = m \mid C] = (n+1)^{-1}, \qquad m = 0, 1, \cdots, n,$$

if and only if

$$\Pr\{[i^{-1}S_i = (n+1)^{-1}S_{n+1}] \cap C\} = 0, \qquad i = 1, 2, \cdots, n."$$

From now on $X_1, X_2, \cdots$ will be an infinite sequence of indentically distributed independent random variables, with $S_k=X_1+\cdots+X_k$ and $S_k^+=\max[0, S_k]$. If τ is the permutation

$$\tau = (\alpha_1)(\alpha_2)\cdots(\alpha_{n(\tau)}),$$

then

$$T(\tau X) = \sum_{i=1}^{n(\tau)}\left(\sum_{k\in\alpha_i} X_k\right)^+,$$

and if

$$\sigma = \begin{pmatrix}1 & 2 & \cdots & n\\ \sigma_1 & \sigma_2 & \cdots & \sigma_n\end{pmatrix},$$

then

$$S(\sigma X) = \max_{1\leqq k\leqq n}\left(\sum_{i=1}^{k} X_{\sigma_i}\right)^+.$$

For every complex-valued $f(x)$ it follows from equation (1.1) and from Theorem 2.2 that

$$Ef\left[\max_{1\leqq k\leqq n} S_k^+\right] = \frac{1}{n!}\sum_{\sigma} Ef[S(\sigma X)] = \frac{1}{n!}\sum_{\tau} Ef[T(\tau X)]. \tag{3.1}$$

Equation (3.1) must be interpreted in the sense that each member is finite and equal to the other two, provided that one of them is finite.

Everything said so far is valid for symmetrically dependent random variables. To take advantage of the independence of the X_i it is convenient to take

$$f(x) = \exp(i\lambda x), \qquad \mathrm{Im}(\lambda) \geqq 0,$$

$$\phi_n(\lambda) = E\exp\left(i\lambda \max_{1\leqq k\leqq n} S_k^+\right), \qquad \phi_0(\lambda) = 1, \tag{3.2}$$

$$\psi_k(\lambda) = E\exp(i\lambda S_k^+). \tag{3.3}$$

Consider the last member of equation (3.1). If a permutation τ consists of k_ν cycles of length ν, $\nu=1, 2, \cdots, n$, with $k_1+2k_2+\cdots+nk_n=n$, then

$$Ef[T(\tau X)] = \prod_{\nu=1}^{n}[\psi_\nu(\lambda)]^{k_\nu}.$$

The number of permutations on n objects, which, when decomposed into disjoint cycles, exhibit the above structure is exactly

$$n!\prod_{\nu=1}^{n}\nu^{-k_\nu}(k_\nu!)^{-1}.$$

Hence

$$(3.4) \qquad \phi_n(\lambda) = \sum{}^* \prod_{\nu=1}^{n} \left(\frac{\psi_\nu(\lambda)}{\nu}\right)^{k_\nu} \frac{1}{k_\nu!}, \qquad n \geqq 0,$$

where the summation $\sum^*$ extends over all n-tuples $(k_1, k_2, \cdots, k_n)$ of non-negative integers with the property $k_1+2k_2+\cdots+nk_n=n$. It is easy to verify that the identity between generating functions in equation (1.3) is equivalent to equation (3.4). Hence we have proved

THEOREM 3.1. *For* $|t|<1$, Im $(\lambda)\geqq 0$,

$$(3.5) \qquad \sum_{n=0}^{\infty} \phi_n(\lambda)t^n = \exp\left[\sum_{k=1}^{\infty} \frac{\psi_k(\lambda)}{k} t^k\right].$$

In §6, Theorem 3.1 will be found to be a special case of Theorem 6.1.

The results of M. Kac, G. A. Hunt, and E. Sparre Andersen are simple consequences of Theorem 3.1.

COROLLARY 1 (*Theorem* 4.1 *in* [7]).

$$(3.6) \qquad E \max_{1\leqq k\leqq n} S_k^+ = \sum_{k=1}^{n} \frac{1}{k} ES_k^+.$$

This formula is obtained by differentiating (3.5) with respect to λ and setting $\lambda=0$. Alternatively, it can be proved directly from equation (3.1) with the choice of $f(x)=x$.

COROLLARY 2.

$$(3.7) \qquad \sum_{n=0}^{\infty} \Pr\left\{\bigcap_{k=1}^{n} [S_k \geqq 0]\right\} t^n = \exp\left\{\sum_{k=1}^{\infty} \frac{t^k}{k} \Pr[S_k \geqq 0]\right\}.$$

This result is obtained by applying equation (3.5) to the random variables $-X_i$, setting $\lambda=iu$, and letting $u\to\infty$. Actually (3.7) remains correct if both inequalities in (3.7) are modified to be strict inequalities. In that form (3.7) was discovered by E. Sparre Andersen. It is equation (3.6) of Theorem 1 in [3]. To obtain it directly, it is easiest to use the following weak form of Theorem 2.3. The probability that the first n partial sums $S_1, S_2, \cdots, S_n$ are all positive is the same as the probability of success in the following experiment. One selects a permutation at random (i.e. with equal probability) from the $n!$ permutations of size n, and observes the lengths $\nu_1, \nu_2, \cdots, \nu_r$ of its successive disjoint cycles. One then observes r independent random variables with the same distribution as $S_{\nu_1}, S_{\nu_2}, \cdots, S_{\nu_r}$, and success is defined as the event that they are all positive. This fact then leads to a direct proof of equation (3.7) with strict inequalities quite similar to the proof of (3.5).

The analogue of Theorem 3.1 for continuous parameter separable stochastic processes with stationary independent increments will be given elsewhere.

4. The limit behavior of max $[0, S_1, \cdots, S_n]$.

THEOREM 4.1. *Let* $a_k = \Pr[S_k > 0]$, $\psi_k(\lambda) = E \exp(i\lambda S_k^+)$.

(a) *If* $\sum_1^\infty a_k/k < \infty$, *then*

$$\max_{1 \leq k \leq n} S_k^+ \to \sup_{k \geq 1} S_k^+ = \max_{k \geq 1} S_k^+ < \infty \tag{4.1}$$

with probability one;

$$\limsup_{n \to \infty} S_n = -\infty \text{ with probability one,} \tag{4.2}$$

except in the trivial case when $\Pr[X_i = 0] = 1$. $\max_{k \geq 1} S_k^+$ *has the infinitely divisible characteristic function*

$$E \exp(i\lambda \max_{k \geq 1} S_k^+) = \prod_{k=1}^{\infty} \exp\left[\frac{\psi_k(\lambda) - 1}{k}\right]. \tag{4.3}$$

(b) *If* $\sum_1^\infty a_k/k = \infty$, *then*

$$\max_{1 \leq k \leq n} S_k^+ \to \sup_{k \geq 1} S_k^+ = \limsup_{n \to \infty} S_n = \infty \tag{4.4}$$

with probability one.

(c) *If* $E|X_i| < \infty$ *and* $\Pr[X_i = 0] < 1$, *then case* (a) *corresponds to* $EX_i < 0$, *while case* (b) *corresponds to* $EX_i \geq 0$.

Proof. To prove (4.1) and (4.2) it is sufficient to show that in case (a)

$$\Pr[S_n > x \text{ i.o.}] = 0, \qquad -\infty < x < \infty. \tag{4.5}$$

Let $q_n = \Pr[\max_{1 \leq k \leq n} S_k \leq 0]$, $q_0 = 1$. In Theorem 3.1 set $\lambda = iu$ and let $u \to \infty$. Then

$$\sum_0^\infty q_n t^n = \exp\left[\sum_1^\infty \left(\frac{1 - a_k}{k}\right) t^k\right], \qquad |t| < 1,$$

$$(1 - t) \sum_0^\infty q_n t^n = \exp\left[-\sum_1^\infty \frac{a_k}{k} t^k\right], \qquad |t| < 1. \tag{4.6}$$

Since the coefficients q_n are monotone nonincreasing, a simple Tauberian argument gives

$$\lim_{n \to \infty} q_n = \exp\left[-\sum_1^\infty \frac{a_k}{k}\right] > 0, \tag{4.7}$$

so that

$$\Pr[S_n > 0 \text{ i.o.}] \leq 1 - \lim_{n \to \infty} q_n < 1. \tag{4.8}$$

But it was shown by P. Lévy [8, p. 147], that when $\Pr[X_i=0]<1$,

$$\Pr[S_n > x \text{ i.o.}] = 0 \quad \text{or} \quad 1, \qquad -\infty < x < \infty, \tag{4.9}$$

and that the probability in (4.8) is the same for every x. Equations (4.8) and (4.9) now imply (4.5), and hence (4.1) and (4.2) are proved.

Equation (4.1) implies that the sequence

$$\phi_n(\lambda) = E\exp\left[i\lambda \max_{1\leqq k\leqq n} S_k^+\right]$$

converges to a characteristic function. Therefore

$$\begin{aligned} E\exp\left[i\lambda \max_{k\geqq 1} S_k^+\right] &= \lim_{n\to\infty}\phi_n(\lambda) = \lim_{t\to 1}(1-t)\sum_1^\infty \phi_n(\lambda)t^n \\ &= \lim_{t\to 1}(1-t)\exp\left[\sum_1^\infty \frac{\psi_k(\lambda)}{k}t^k\right] \\ &= \exp\left[\sum_1^\infty \frac{\psi_k(\lambda)-1}{k}\right]. \end{aligned}$$

This characteristic function is infinitely divisible since its factors

$$\exp\left(\frac{\psi_k(\lambda)-1}{k}\right) = \exp\left[\frac{1}{k}\left\{\int_0^\infty (e^{i\lambda x}-1)d\Pr[S_k\leqq x]\right\}\right]$$

are infinitely divisible.

To prove equation (4.4), suppose that $\sum_1^\infty a_k/k=\infty$. Then it follows from (4.6) that

$$\lim_{n\to\infty} q_n = \Pr\left[\sup_{k\geqq 1} S_k \leqq 0\right] = 0.$$

Now

$$\begin{aligned} 1-\Pr[S_k > 0 \text{ i.o.}] &= \Pr\left[\sup_{k\geqq 1} S_k \leqq 0\right] + \sum_{k=1}^\infty \Pr\left[S_k>0;\ \sup_{m\geqq 1} S_{k+m}\leqq 0\right] \\ &\leqq \lim_{n\to\infty}\sum_{k=1}^n \Pr\left[\sup_{m\geqq 1} S_m \leqq 0\right] = 0. \end{aligned}$$

Hence, by (4.9)

$$\Pr[S_k > 0 \text{ i.o.}] = \Pr[S_k > x \text{ i.o.}] = 1, \qquad -\infty < x < \infty,$$

which implies equation (4.4).

It remains to prove part (c). We assume that $E|X_i|<\infty$. Then, if $EX_i<0$, the strong law of large numbers implies that $\Pr[S_n>0 \text{ i.o.}]=0$, so that $\sum_1^\infty a_k/k<\infty$. Conversely, if $EX_i>0$, we get $\Pr[S_n<0 \text{ i.o.}]=0$, so that

$\Pr[S_n>0 \text{ i.o.}]=1$, and $\sum_1^\infty a_k/k=\infty$. Finally, if $EX_i=0$, then the sequence of partial sums S_n is recurrent in the sense of Chung and Fuchs [4], so that $\Pr[S_n>0 \text{ i.o.}]=1$ if $\Pr[X_i=0]<1$, and $\sum_1^\infty a_k/k=\infty$.

Theorem 4.1 has a curious corollary which constitutes a novel form of the strong law of large numbers for identically distributed random variables. It gains in interest by comparison to the following result of P. Erdös [6].

$EX_i=m$ and $EX_i^2<\infty$, if and only if

$$\sum_1^\infty \Pr\left[\left|\frac{S_k}{k}-m\right|>\epsilon\right]<\infty \text{ for every } \epsilon>0.$$

We shall prove

THEOREM 4.2. *$EX_i=m$ if and only if*

$$\sum_1^\infty \frac{1}{k}\Pr\left[\left|\frac{S_k}{k}-m\right|>\epsilon\right]<\infty \textit{ for every } \epsilon>0. \tag{4.10}$$

Proof. It will suffice to consider the case when $m=0$. Suppose therefore that $EX_i=0$. Then $E(X_i-\epsilon)<0$, if $\epsilon>0$, and by Theorem 4.1 (c)

$$\sum_1^\infty \frac{1}{k}\Pr[S_k-k\epsilon>0]<\infty. \tag{4.11}$$

By the same argument

$$\sum_1^\infty \frac{1}{k}\Pr[S_k+k\epsilon<0]<\infty. \tag{4.12}$$

Adding (4.11) and (4.12), we have (4.10). Conversely, assume that (4.10) holds with $m=0$. Then (4.11) and (4.12) must hold for every $\epsilon>0$, which means that

$$\Pr\left[\frac{S_n}{n}\to 0\right]=1.$$

Now it follows from the converse of the strong law of large numbers that $EX_i=0$.

5. **The number of positive partial sums.** We summarize earlier definitions and make some new ones.

$$a_k=\Pr[S_k>0],$$

$$p_n=\Pr\left[\min_{1\leq k\leq n} S_k>0\right], \qquad p_0=1,$$

$$q_n=\Pr\left[\max_{1\leq k\leq n} S_k\leq 0\right], \qquad q_0=1.$$

(5.1) N_n = the number of $S_k > 0$, $k = 1, 2, \cdots, n$.
(5.2) T_n = the index (time) k at which $\max_{1 \leq k \leq n} S_k$ is first attained, with the provision that $T_n = 0$ if $\max_{1 \leq k \leq n} S_k \leq 0$.

The results of this section will depend on the following important theorem of E. Sparre Andersen (Theorem 1 in [1]).

"*If* $p_k = \Pr[N_k = k]$, $q_k = \Pr[N_k = 0]$ (as is the case in our notation), *then*

$$\Pr[N_n = k] = p_k q_{n-k} = \Pr[T_n = k], \quad k = 0, 1, \cdots, n." \tag{5.3}$$

The second half of equation (5.3) is a simple consequence of the markovian nature of the process of successive partial sums S_n. To prove the first half, we let

$$\rho_n(\lambda) = E[e^{-\lambda n \theta(S_n)}] = a_n e^{-\lambda n} + 1 - a_n, \qquad \lambda \geq 0. \tag{5.4}$$

$$\chi_n(\lambda) = E[e^{-\lambda N_n}], \qquad \lambda \geq 0. \tag{5.5}$$

One sees that in view of Theorem 2.3 the random variable N_n has the same distribution as the random variable

$$\nu_1\theta(S_{\nu_1}) + \nu_2\theta(S_{\nu_2}) + \cdots + \nu_r\theta(S_{\nu_r}),$$

where $\nu_1, \cdots, \nu_r$ are random variables indicating the length of the disjoint cycles of a permutation chosen at random from among the permutations on n objects. The partial sums $S_{\nu_1}, \cdots, S_{\nu_r}$ are taken as independent random variables with the distribution indicated by their subscript. Hence the method used to prove Theorem 3.1 applies with only the change of $\phi_n(\lambda)$ into $\chi_n(\lambda)$, and $\psi_n(\lambda)$ into $\rho_n(\lambda)$, and yields the following result analogous to equation (3.5):

$$\sum_{n=0}^{\infty} \chi_n(\lambda) t^n = \exp\left[\sum_{k=1}^{\infty} \frac{\rho_k(\lambda)}{k} t^k\right], \qquad \lambda \geq 0. \tag{5.6}$$

Using (5.4) this becomes

$$\sum_0^{\infty} \chi_n(\lambda) t^n = \exp\left[\sum_1^{\infty} \frac{a_k}{k} (te^{-\lambda})^k\right] \exp\left[-\sum_1^{\infty} \frac{a_k}{k} t^k\right],$$

and in view of equation (3.7) and the discussion following it,

$$\sum_0^{\infty} \chi_n(\lambda) t^n = \left(\sum_0^{\infty} e^{-\lambda k} t^k p_k\right)\left(\sum_0^{\infty} t^k q_k\right),$$

so that

$$\chi_n(\lambda) = \sum_{k=0}^{n} e^{-\lambda k} \Pr[N_n = k] = \sum_{k=0}^{n} e^{-\lambda k} p_k q_{n-k},$$

from which the first half of (5.3) follows.

Now we consider the limiting behavior of N_n and T_n.

THEOREM 5.2. (a) *If* $\sum_1^\infty a_k/k < \infty$, *then* $N_n \to N < \infty$ *and* $T_n \to T < \infty$, *with probability one, and*

$$(5.7) \qquad Et^N = Et^T = \exp\left[\sum_1^\infty \frac{a_k}{k}(t^k - 1)\right], \qquad |t| \leqq 1.$$

(b) *If* $\sum_1^\infty a_k/k = \infty$, *then* $N_n \to \infty$ *and* $T_n \to \infty$ *with probability one.*

Proof. Let $\sum_1^\infty a_k/k < \infty$. Then equations (4.1) and (4.2) of Theorem 4.1 imply that

$$\Pr[N_{n+1} \neq N_n \text{ i.o.}] = \Pr[T_{n+1} \neq T_n \text{ i.o.}] = 0.$$

On the other hand, if $\sum_1^\infty a_k/k = \infty$, then it follows from equation (4.4) that

$$\Pr[N_{n+1} = N_n + 1 \text{ i.o.}] = \Pr[T_n = n \text{ i.o.}] = 1.$$

This proves the probability one statements in parts (a) and (b) of the present theorem.

To prove equation (5.7) it suffices to consider the random variables N_n. In case (a), when they converge to a random variable N with probability one, we have

$$Et^N = \sum_{k=0}^\infty t^k \lim_{n\to\infty} \Pr[N_n = k].$$

By (5.3)

$$\lim_{n\to\infty} \Pr[N_n = k] = p_k \lim_{n\to\infty} q_n = p_k q.$$

In fact, (4.7) states that

$$q = \exp\left[-\sum_1^\infty \frac{a_k}{k}\right],$$

and (3.7) that

$$\sum_0^\infty p_k t^k = \exp\left[\sum_1^\infty \frac{a_k}{k} t^k\right],$$

so that

$$Et^N = Et^T = \exp\left[\sum_1^\infty \frac{a_k}{k}(t^k - 1)\right], \qquad |t| \leqq 1.$$

6. **The joint distribution of S_n and** max $[0, S_1, \cdots, S_n]$. The results and methods of the preceding sections enable us to find the joint characteristic function of S_n and max $[0, S_1, \cdots, S_n]$. For reasons of symmetry we shall concentrate on the bivariate characteristic function

$$(6.1)\quad \phi_n(\alpha, \beta) = E \exp\left[i\alpha \max_{1\leq k\leq n} S_k^+ + i\beta\left(\max_{1\leq k\leq n} S_k^+ - S_n\right)\right], \qquad \phi_0(\alpha, \beta) = 1.$$

The joint characteristic function of S_n and max $[0, S_1, \cdots, S_n]$ will then be

$$(6.2)\qquad E \exp\left[i\alpha S_n + i\beta \max_{1\leq k\leq n} S_k^+\right] = \phi_n(\alpha + \beta, -\alpha).$$

It will be seen that $\phi_n(\alpha, \beta)$ depends only on the characteristic functions

$$(6.3)\qquad u_k(\alpha) = E \exp\,[i\alpha S_k^+], \qquad (2S_k^+ = |S_k| + S_k),$$

$$(6.4)\qquad v_k(\beta) = E \exp\,[i\beta S_k^-], \qquad (2S_k^- = |S_k| - S_k),$$

for $k=1, 2, \cdots, n$.

THEOREM 6.1. *For* $|t|<1$, $\text{Im}\,(\alpha)\geqq 0$, $\text{Im}\,(\beta)\geqq 0$,

$$(6.5)\qquad \sum_{n=0}^{\infty} \phi_n(\alpha, \beta)t^n = \exp\left[\sum_{k=1}^{\infty} \frac{1}{k}(u_k(\alpha) + v_k(\beta) - 1)t^k\right].$$

The proof depends on equation (5.3)

$$\Pr\,[T_n = k] = p_k q_{n-k}.$$

If max $[0, S_1, \cdots, S_n]$ is assumed at $T_n = k$, then the random variables $S_i - S_k$, $i\leqq k$, and $S_j - S_k$, $j\geqq k$, are independent. By a simple argument one obtains

$$(6.6)\qquad \phi_n(\alpha, \beta) = \sum_{k=0}^{n} p_k q_{n-k} \frac{\int_{A_k} e^{i\alpha S_k} dP}{\Pr\,[A_k]}\cdot\frac{\int_{B_{n-k}} e^{-i\beta S_{n-k}} dP}{\Pr\,[B_{n-k}]},$$

where

$$A_k = \bigcap_{i=1}^{k} [S_i > 0],\ B_k = \bigcap_{i=1}^{k} [S_i \leqq 0],$$

so that $\Pr\,[A_k] = p_k$, $\Pr\,[B_k] = q_k$. To evaluate the integrals in (6.6) one must go back to Theorem 2.2 and apply the method of equation (1.1) just as was done in the proof of Theorem 3.1. One obtains

$$(6.7)\qquad \sum_{n=0}^{\infty} t^n \int_{A_n} e^{i\alpha S_n} dP = \exp\left[\sum_{k=1}^{\infty} \frac{1}{k}(u_k(\alpha) - 1 + a_k)t^k\right],$$

$$(6.8)\qquad \sum_{n=0}^{\infty} t^n \int_{B_n} e^{i\beta S_n} dP = \exp\left[\sum_{k=1}^{\infty} \frac{1}{k}(v_k(\beta) - a_k)t^k\right],$$

where $a_k = \Pr [S_k > 0]$. But now it follows from equation (6.6) that the series on the left in (6.5) is the product of the generating function in (6.7) and (6.8). This product gives the right-hand side of (6.5) as was to be proved.

7. **The generalized arc-sine law.** By the generalized arc-sine laws we mean the following one parameter family of distributions $F_\alpha(x)$.

$$F_0(x) = 0 \text{ if } x < 0, \qquad 1 \text{ if } x \geqq 0,$$

$$F_\alpha(x) = \frac{\sin \pi\alpha}{\pi} \int_0^x s^{\alpha-1}(1-s)^{-\alpha} ds, \qquad 0 < \alpha < 1,$$

$$F_1(x) = 0 \text{ if } x < 1, \qquad 1 \text{ if } x \geqq 1.$$

E. Sparre Andersen was the first to prove that the sequence of random variables N_n/n, i.e. the fraction of times that $S_n > 0$, converges in distribution to the law $F_\alpha(x)$, if the sequence $a_k = \Pr [S_k > 0]$ converges to α (Theorem 3 in [3]). We generalize his result in the following way.

THEOREM 7.1. *If* $(a_1 + a_2 + \cdots + a_n)/n \to \alpha$, *then*

$$\Pr\left[\frac{N_n}{n} \leqq x\right] \to F_\alpha(x). \tag{7.1}$$

If $(a_1 + \cdots + a_n)/n$ *does not tend to a limit, then neither does* $\Pr [N_n/n \leqq x]$.

The second half of the theorem is trivial, since

$$\frac{a_1 + \cdots + a_n}{n} = E\frac{N_n}{n},$$

and a sequence of uniformly bounded random variables cannot converge in distribution unless their first moments converge. Moreover, this half of the theorem may also be vacuous, since it is not known whether there exists a sequence of identically distributed independent random variables with the property that the a_k fail to have a $(C, 1)$ limit.

However, the so called universal laws of Doeblin [5] show that the ordinary limit of a_k need not exist. Hence Theorem 7.1 is a generalization of E. Sparre Andersen's.

The proof of (7.1) will be based on the following Abelian theorem.

LEMMA 7.2. *If the sequence* $\{a_k\}$, $k = 1, 2, \cdots$, *is* $(C, 1)$ *summable to* α, *and* $\lambda \geqq 0$, *then as* $s \to 1$ *through real* $s < 1$,

$$\lim_{s\to 1} \sum_{k=1}^{\infty} \frac{a_k}{k} s^k [1 - e^{-\lambda k(1-s)}] = \alpha \log (1 + \lambda). \tag{7.2}$$

Proof. Let $A_n = (a_1 + a_2 + \cdots + a_n)/n$. Then after summation by parts, equation (7.2) becomes

$$\lim_{s\to 1} \sum_{k=1}^{\infty} A_k c_k(s) = \alpha \log (1+\lambda),$$

where $c_k(s)=s^k[1-e^{-\lambda k(1-s)}]-(ks^{k+1}/(k+1))[1-e^{-\lambda(k+1)(1-s)}]$. By Toeplitz, Theorem we have to prove that

(a) $$\lim_{s\to 1} c_k(s) = 0,$$

(b) $$\lim_{s\to 1} \sum_{k=1}^{\infty} c_k(s) = \log (1+\lambda),$$

(c) $$\sum_{k=1}^{\infty} |c_k(s)| < K \text{ in some interval } 1-\delta \leqq s \leqq 1.$$

It is easily deduced that (a) and (b) hold. To verify (c), note that

$$c_k(s) = \lambda s^k \int_0^{k(1-s)} e^{-x\lambda}dx - \frac{\lambda k s^{k+1}}{k+1}\int_0^{(k+1)(1-s)} e^{-x\lambda}dx \geqq -\lambda s^k \int_{k(1-s)}^{(k+1)(1-s)} e^{-x\lambda}dx$$
$$\geqq -\lambda s^k(1-s)e^{-\lambda k(1-s)}.$$

Hence $|c_k(s)| \leqq c_k(s)+2\lambda s^k(1-s)e^{-\lambda k(1-s)}$, so that

$$\limsup_{s\to 1} \sum_{k=1}^{\infty} |c_k(s)| \leqq \log (1+\lambda) + 2\lambda,$$

so that (c) is satisfied for some $\delta>0$ with $K=4\lambda$.

We proceed to prove Theorem 7.1. As in §5 let

$$\chi_n(\lambda) = E[e^{-\lambda N_n}],$$

and equation (5.6) can be written in the form

$$(7.3)\qquad (1-s)\sum_{n=0}^{\infty} \chi_n[\lambda(1-s)]s^n = \exp\left\{\sum_{n=1}^{\infty} \frac{a_n}{n} s^n[1-e^{-\lambda n(1-s)}]\right\},$$

when $|s|<1$ and $\lambda\geqq 0$. For $|s|<1$ both members of (7.3) can be written as convergent power series in λ and we may equate the coefficients $c_k(s)$ of λ^k, $k=0, 1, \cdots$. Lemma 7.2 simply states that

$$\lim_{s\to 1} \sum_{k=0}^{\infty} c_k(s)\lambda^k = (1+\lambda)^{-\alpha},$$

so that

$$(7.4)\qquad \lim_{s\to 1} c_k(s) = \binom{-\alpha}{k}, \qquad k = 0, 1, \cdots.$$

Since $\chi_n(\lambda)$ is a moment generating function for N_n we can use equation (7.3) to express the coefficients $c_k(s)$ in terms of the moments of N_n/n. Let

$$\mu_k^{(n)} = E\left(\frac{N_n}{n}\right)^k.$$

Then

$$c_k(s) = (1-s)^{k+1}\frac{(-1)^k}{k!}\sum_{n=0}^{\infty} n^k\mu_k^{(n)}s^n, \qquad |s| < 1,$$

and equation (7.4) implies that

$$(7.5) \qquad (1-s)\sum_{n=0}^{\infty} n^k\mu_k^{(n)}s^n \sim (-1)^k\binom{-\alpha}{k}\frac{k!}{(1-s)^k}, \qquad s \to 1.$$

The left-hand member as a power series in s has the coefficients

$$(n+1)^k\mu_k^{(n+1)} - n^k\mu_k^{(n)} = E(N_{n+1}^k - N_n^k),$$

which are non-negative. Karamata's Tauberian theorem therefore applies, and yields

$$\lim_{n\to\infty}\mu_k^{(n)} = (-1)^k\binom{-\alpha}{k}, \qquad k = 0, 1, \cdots.$$

But it is easy to verify that

$$(-1)^k\binom{-\alpha}{k} = \int_{-\infty}^{\infty} x^k dF_\alpha(x),$$

and because the moment problem in this case has a unique solution, the proof of Theorem 7.1 is complete.

Bibliography

1. E. Sparre Andersen, *On sums of symmetrically dependent random variables*, Skandinavisk Aktuarietidskrift vol. 36 (1953) pp. 123–138.

2. ———, *On the fluctuations of sums of random variables*, Math. Scand. vol. 1 (1953) pp. 263–285.

3. ———, *On the fluctuations of sums of random variables* II, Math. Scand. vol. 2 (1954) pp. 195–223.

4. K. L. Chung and W. H. J. Fuchs, *On the distribution of values of sums of random variables*, Memoirs of the American Mathematical Society vol. 6 (1951) pp. 1–12.

5. W. Doeblin, *Sur l'ensemble de puissances d'une loi de probabilité*, Studia Mathematica vol. 9 (1941) pp. 71–96.

6. P. Erdös, *Remark on my paper "On a theorem of Hsu and Robbins,"* Ann. Math. Statist. vol. 21 (1950) p. 138.

7. M. Kac, *Toeplitz matrices, translation kernels and a related problem in probability theory*, Duke Math. J. vol. 21 (1954) pp. 501–509.

8. P. Lévy, *Théorie de l'addition des variables aléatoires*, Paris, 1937.

California Institute of Technology,
Pasadena, Calif.