NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.
TECHNICAL FINAL REPORT

INVESTIGATIONS ON FLUCTUATIONS OF SUMS OF RANDOM VARIABLES.

E. Sparre Andersen

Mathematical Institute
Aarhus University
Aarhus, Denmark

The research report in this document has been sponsored by
THE DIRECTORATE MATHEMATICAL SCIENCES, AFOSR, OAR
through the European Office, Aerospace Research, United States
Air Force.
TECHNICAL FINAL REPORT

INVESTIGATIONS ON FLUCTUATIONS OF SUMS OF RANDOM VARIABLES.

E. Sparre Andersen

Mathematical Institute
Aarhus University
Aarhus, Denmark

The research report in this document has been sponsored by
THE DIRECTORATE MATHEMATICAL SCIENCES, AFSOR, OAR
through the European Office, Aerospace Research, United States Air Force.
Investigations on Fluctuations of Sums of Random variables.

Abstract.

We have investigated the fluctuations of sums of random variables $X_1, X_2, \ldots$. We have generalized previous results on the random variables $H_n$ connected with this sequence, obtained new results on the validity of the Arc-sine Law for independent not identically distributed random variables, obtained a generalization of Spitzer's identity, and obtained a generalization of the equivalence principle. Furthermore Toeplitz matrices of Laurent polynomials have been studied. For the growth of the maximal order statistics of a sequence of independent, identically distributed random variables a "Law of the iterated logarithm" has been obtained.
Investigations on Fluctuations of Sums of Random Variables.

1. The main subject for work under the Contract has been the investigation of the fluctuations of sums of random variables. We have worked on different problems within this theory.

2. Let \( X_1, X_2, \ldots \) be a sequence of random variables and let \( S_0 = 0, S_k = X_1 + \ldots + X_k \) for \( k = 1, 2, \ldots \). We have investigated the behaviour of the largest convex minorant sequence \( Z_{n,k} \), \( k = 0, 1, \ldots, n \) to the sequence \( S_0, S_1, \ldots, S_n \). The number of equalities \( S_k = Z_{n,k}, k = 1, \ldots, n \) is, to some extent, a measure of the fluctuations of the sequence \( S_0, \ldots, S_n \). We denote this number by \( H_n \). (We shall define \( H_0 \) as \( 0 \)). It is clear that \( H_n \) is a random variable, the distribution of which does depend only on the distribution of \( X_1, \ldots, X_n \). The meaning of \( H_n \) may be clarified by considering figure 1., where the points have coordinates \((k, S_k), k = 0, \ldots, 10\).

In the figure \( H_{10} \) has the value 6. It follows from the figure that another natural statistics similar to \( H_n \) is the number of straight segments in the convex polygon. We denote this number by \( K_n \); evidently \( H_n \geq K_n \). If the random variables have a continuous distribution, then \( P(H_n=K_n) = 1 \).
It was shown by Andersen, \[3\], that if the random variables \(X_1, X_2, \ldots\) are independent, and have the same continuous distribution, then

\[
H(s,t) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} P(H_n=m) s^n t^m = (1-s)^{-t}, |s|<1, |t|<1,
\]

holds for the double-generating function for \(P(H_n=m)\). (Actually in \[3\] the formula looks somewhat different, due to a trivial change in the definition of \(H_n\)).

The main result obtained for \(H_n\) under the Contract is:

(cf. Technical Note No. 1.)

**Theorem 1.** Let \(X_1, X_2, \ldots\) be independent, identically distributed random variables. Let \(x_1, x_2, \ldots\) be the set of real numbers \(x\) for which \(P(S_k = k \cdot x) > 0\) for some \(k > 0\), and let

\[
\begin{align*}
\delta_j(s) &= \sum_{k=1}^{\infty} P(S_k = k \cdot x_j) s^k, & j = 1, 2, \ldots \\
\delta_0(s) &= \sum_{k=1}^{\infty} P(S_k = k \cdot x_j \text{ for } j = 1, 2, \ldots) s^k \\
&= s(1-s)^{-1} - \sum_{j=1}^{\infty} \delta_j(s).
\end{align*}
\]

Then

\[
H(s,t) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} P(H_n=m) s^n t^m
\]

\[
= \exp(t \int_0^s x^{-1} \delta_0(x) dx) \cdot \prod_{j=1}^{\infty} (1-t^{1+} \exp(-\int_0^s x^{-1} \delta_j(x) dx))^{-1}.
\]

Unfortunately this formula is much too complicated to be useful. Even in the case where the random variables \(X_1, X_2, \ldots\) have only two possible values, say \(-1\) and \(-1\), each with probability \(1/2\), it has not been possible to obtain a simple result. It was shown in TN1 that a simpler formula may be obtained if the common distribution of \(X_1, X_2, \ldots\) have only one discontinuity-point. For this case also the double-generating function of \(P(K_n=m)\) was found.
A fundamental result in the theory of fluctuations of sums of random variables is the so-called Arc-sine Law. Let $N_n$ denote the number of positive sums among $S_1, \ldots, S_n$. We say that the Arc-sine Law holds for the sequence $X_1, X_2, \ldots$ of random variables, if

$$\lim_{n \to \infty} P(N_n/n \leq x) = \frac{2}{\pi} \arcsin x^{1/2}, \quad 0 \leq x \leq 1.$$  

If for an $a \in (0,1)$ we have

$$\lim_{n \to \infty} P(N_n/n \leq x) = \frac{\sin \pi a}{\pi} \int_0^x y^{a-1}(1-y)^{a-1}dy, \quad 0 \leq x \leq 1,$$

then we say that the generalized Arc-sine Law holds for the random variables $X_1, X_2, \ldots$. For $a = 1/2$ the right hand side of (6) reduces to the right hand side of (5).

Erdős and Kac,[[13]], proved that the Arc-sine Law holds with $a = 1/2$ if the random variables $X_1, X_2, \ldots$ are independent, have mean 0 and variance 1 and obey the Central Limit Theorem. Andersen proved[[3]] that (6) holds if the random variables $X_1, X_2, \ldots$ are independent, identically distributed and if

$$\lim_{n \to \infty} P(S_n > 0) = a.$$  

It was proved by Spitzer[[24]] that (7) can be replaced by the weaker condition

$$\lim_{n \to \infty} \frac{1}{n} \left( P(S_n > 0) + \ldots + P(S_n > 0) \right) = a.$$  

It is not known whether Spitzer's result is really stronger than Andersen's result, since it is doubtful that there exist sequences of independent, identically distributed random variables for which the limit in (7) does not exist, while (8) is satisfied.
The result of Erdös and Kac apparently connects the Arc-sine Law with the Central Limit Theorem, whereas the result of Andersen shows that there is no such connection, as it is possible to choose for the common distribution of the random variables $X_1, X_2, \ldots$ any symmetric, non-degenerate distribution and thereby obtain the Arc-sine Law.

It is fairly easy to see by heuristic arguments that the number of positive sums $N_n$ ought to follow the Arc-sine Law, if the random variables $X_1, X_2, \ldots$ are in some sense "asymptotically identically distributed". We have investigated this problem under the Contract without having obtained a general result. We have, however, obtained certain results which throw some light on the problem.

**Theorem 2.** Let $X_1, X_2, \ldots$ be a sequence of independent random variables. If this sequence is periodic in the sense that there exist two positive integers $p$ and $q$ and a set of distribution functions $P_1(x), \ldots, P_p(x)$, not all degenerate, such that $P(X_{np+r} \leq x) = P_r(x), 1 \leq r < p$, if $\frac{n}{p} \approx q$, then the Arc-sine Law holds for $X_1, X_2, \ldots$ if $\lim_{n \to \infty} P(S_n > 0) = 1/2$.

**Theorem 3.** Let $X_1, X_2, \ldots$ be a sequence of independent, normally distributed random variables with mean 0. Then

(i) if $E(S_n^2) = \log n$, $n = 1, 2, \ldots$, we have

$$\lim_{n \to \infty} P(N_n/n^{1/2}x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{\pi} & \text{for } 0 \leq x < 1 \\ 1 & \text{for } 1 \leq x \end{cases}$$

(ii) if $E(S_n^2) = n^{b}\lambda(n), n = 1, 2, \ldots$, where $b > 0$ and $\lambda(x)$ is defined, positive and monotone in $0 < x < \infty$ and for every positive constant $c$ satisfies

$$\lambda(cx)/\lambda(x) \to 1 \quad \text{for } x \to \infty,$$

we have

$$\lim_{n \to \infty} P(N_n/n^{1/2}x) = P\left( \int_0^1 \frac{1}{2} \text{sign} X(t) dt / \sqrt{x} \right), 0 \leq x \leq 1,$$

where $X(t), 0 \leq t \leq 1$, is the Wiener Process.
(iii) if \( E(s_n^2) = b^n, b > 1, n = 1, 2, \ldots \), we have

\[
\lim_{n \to \infty} P(N_n/n^{\delta}x) = \begin{cases} 
0 & \text{for } x < \frac{1}{b} \\
1 & \text{for } \frac{1}{b} < x.
\end{cases}
\]

A proof of Theorem 2 is given in TN2. Theorem 3 may be proved by a modification of the invariance principle used by Erdős and Kac in [13]. Case (i) of Theorem 3 is due to G. Maruyama [22], together with (iii) it represents extreme cases where the variance of \( S_n \) increases so slowly that practically all sums have the same sign, or so fast that for large \( n \) approximately 50% of the sums are positive. Case (iii) for \( b = 1 \) gives the usual Arc-sine Law, for other values of \( b \) it has not been possible to evaluate

\[
P \left( \int_0^1 \frac{1}{2} \text{sign} \frac{X(t)}{b} \, dt \right) = \lim_{n \to \infty} P(\mu_x > 0)
\]

The most interesting case of Theorem 3, therefore, is case (ii) with \( b = 1 \). It shows that the Arc-sine Law may hold for a sequence of independent random variables \( X_1, X_2, \ldots \), even if the variance of \( X_n \) goes to infinity like \((\log n)^c, c > 0\) for \( n \to \infty \). In this case the random variables \( X_1, X_2, \ldots \) cannot be said to be "asymptotically identically distributed" in the usual sense of the words. From this, and also from Theorem 2, it follows that the Arc-sine Law must hold for a sequence of independent random variables under rather weak assumptions about the distribution functions. Several unsuccessful attempts have been made to prove that the Arc-sine Law holds under conditions which are more general than the conditions in Theorem 2, or case (ii) of Theorem 3. Our investigations have led us to the conjecture that the generalized Arc-sine Law holds for the independent random variables \( X_1, X_2, \ldots \) with distribution functions \( F_1(x), F_2(x), \ldots \), if

\[
\lim_{n \to \infty} P(S_n > 0) = a
\]
and if there exists a sequence of distribution functions \( G_1(x), G_2(x), \ldots \) such that to any \( \varepsilon > 0 \) we can find an \( N(\varepsilon) \), tending to infinity when \( \varepsilon \to 0 \), and a \( \delta(\varepsilon) > 0 \), tending to 0 when \( \varepsilon \to 0 \) such that

\[
\sup_{-\infty < x < \infty} | P(S_k + n - S_k - x) - G_n(x) | < \varepsilon
\]

whenever \( k \geq 0, k + n \leq N(\varepsilon) \) and \( n \geq \delta(\varepsilon)N(\varepsilon) \). The last condition is the way in which we think it natural in this connection to express that the variables \( X_1, X_2, \ldots \) are "asymptotically identically distributed". We have not been able to verify this conjecture.

4. A well-known result in the theory of fluctuations of sums of random variables is Spitzer's identity,\(^{[24]}\),

\[
\sum_{n=0}^{\infty} \varphi(T_n, o, S_n)s^n = \exp \sum_{n=1}^{\infty} \frac{1}{n} \varphi(s^+_n, S_n)s^n, \quad |s| < 1,
\]

where \( \varphi(X,Y) \) denotes the joint characteristic function of the random variables \( X \) and \( Y \), while \( T_n, o = \max(S_0, S_1, \ldots, S_n) \), and \( S^+_n = \max(0, S_n) \). Spitzer's identity is valid if the random variables \( X_1, X_2, \ldots \) are independent and identically distributed. Several proofs are known, some combinatorial and others relying mostly on analysis or functional analysis.

We have treated a generalisation of Spitzer's formula to symmetrically dependent random variables. The random variables \( X_1, \ldots, X_n \) are said to be symmetrically dependent, if the joint distribution function \( F_n(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n) \) is a symmetric function of \( X_1, \ldots, X_n \). The random variables \( X_1, X_2, \ldots \) are said to be symmetrically dependent if, for each positive \( n \), the random variables \( X_1, \ldots, X_n \) are symmetrically dependent. We remark that if the random variables \( X_1, X_2, \ldots \) are symmetrically dependent, then they are equivalent according to the terminology used by de Finetti,\(^{[16]}\).
It was shown by de Finetti that it is possible to deduce results for equivalent random variables from results for independent random variables. We do, however, prefer to work directly on symmetrically dependent random variables for two reasons. One is that the random variables \( X_1, ..., X_n \) may be symmetrically dependent although it is not possible to extend the finite sequence \( X_1, ..., X_n \) to an infinite sequence \( X_1, ..., X_n, X_{n+1}, ... \) such that the infinite sequence of random variables are symmetrically dependent. We therefore in some cases get more general results when we work directly on symmetrically dependent random variables. The other is that the combinatorial methods, we use, seem to be naturally adopted to the concept of symmetrically dependent random variables.

For independent random variables multiplication of characteristic functions corresponds to addition of random variables. The same is not true for dependent random variables. If, however, a set of symmetrically dependent random variables is given, then it is possible to introduce a symbolic multiplication of characteristic functions of random variables which are functions of the given set of symmetrically dependent random variables. This symbolic multiplication coincides with the ordinary multiplication, in case that the given symmetrically dependent random variables are independent and identically distributed. Having introduced this symbolic multiplication of characteristic functions the definition of the right-hand-side of (9) follows by application of the exponential series. We have proved by combinatorial methods that (9) holds for symmetrically dependent random variables \( X_1, X_2, ... \) when the multiplication of characteristic functions is the symbolic multiplication introduced above.

2. The equivalence principle in the theory of fluctuations of sums of random variables states that the number \( N_n \) of positive sums among \( S_1, ..., S_n \) has the same distribution as the index \( L_n \) of the first maximum, if the random variables \( X_1, ..., X_n \) are symmetrically dependent. We have, in [4], obtained a generalization of this result. The random variables \( N_{n,k} \) and \( L_{n,k} \) are defined by
\[ N_{n,k} = \text{number of indices } j \in \{0, \ldots, k-1, k+1, \ldots, n\} \]
for which \( S_j \geq S_k \) if \( j < k \)
\[ \text{or } \quad S_j > S_k \text{ if } j > k, \]

\[ L_{n,k} = \text{index } v \in \{0, \ldots, n\} \text{ of the sum } S_v \]
for which exactly \( k \) sums
\[ S_j \in \{j=0, \ldots, v-1, v+1, \ldots, n\} \]
satisfy \( S_j \geq S_v \) if \( j < v \)
\[ \text{or } \quad S_j > S_v \text{ if } j > v. \]

We remark that \( N_{n,0} \) is the number of positive sums usually denoted by \( N_n \) and that \( L_{n,0} \) is the index of the first maximum.

**Theorem 4.** Let \( X_1, \ldots, X_n \) be symmetrically dependent random variables and let \( C \) be an event which is symmetric with respect to \( X_1, \ldots, X_n \), then
\[
P(N_{n,k}=j, C) = P(L_{n,k}=j, C) \quad j = 0, 1, \ldots, n.
\]

The proof we have given of Theorem 4 in [4] is based on a one-to-one measure-preserving of the \( n \)-dimensional sample space onto itself such that the event \([ N_{n,k}=j, N_{k,v}=C] \) is mapped onto the event \([ L_{n,k}=j, L_{j,v}=C] \). For each point in the sample space the mapping is a permutation of the coordinates.

5. In connection with the work on fluctuations of sums of random variables we have investigated the Toeplitz matrices of Laurent polynomials. The result obtained are given in TN4. It was conjectured that this study would be valuable for the work on extension of the Arc sine Law to the case where the random variables \( X_1, X_2, \ldots \) form a stationary Markov chain. We have, however, not been able to obtain results in this direction.
For a sequence of independent, identically distributed random variables the growth of the maximum of the first $n$ variables have been investigated and a "Law of the iterated logarithm" has been proved in Th5.

**Theorem 5.** Let $X_1, X_2, \ldots$ be independent, identically distributed random variables, with distribution function $P(x)$. Let $\lambda_1, \lambda_2, \ldots$ be a nondecreasing sequence of real numbers then

$$P(\lim \sup \max(X_1, \ldots, X_n) \leq \lambda_n) = 0$$

if

$$\sum_{n=3}^{\infty} \frac{(P(\lambda_n))^n \log n}{n} < \infty$$

furthermore

$$P(\lim \sup \max(X_1, \ldots, X_n) > \lambda_n) = 1$$

if the sequence $(P(\lambda_n))^n, n = 1, 2, \ldots$ is non-increasing and

$$\sum_{n=3}^{\infty} \frac{(P(\lambda_n))^n \log n}{n} = \infty.$$

The proof of Theorem 5 is based on a generalization, obtained in Th5, of the convergence part of the Borel-Cantelli lemmas.
Bibliography.

5. Glen Baxter, Notes for a seminar in stochastic processes, University of Minnesota, 1957.
6. H.T.L. Bizley, Derivation of a new formula for the number of minimal lattice paths from (0,0) to (km, kn) having just t contacts with the line my = nx and having no points above this line; and a proof of Grossman's formula for the number of paths which may touch, but do not rise above this line, *Journal of the Institute of Actuaries*, 80 (1954), 55-62.


Technical Notes and Reports under the Contract.

TN1: E. Sparre Andersen, On the distribution of the random variable \( H_n \) (1959).


TN3: cancelled.


TSR: E. Sparre Andersen, On the distribution of the random variable \( H_n \) (1959).


INVESTIGATIONS ON FLUCTUATIONS OF SUMS OF RANDOM VARIABLES

E. SPARRE ANDERSEN

ABSTRACT: Distribution of the r.v. \( H \). Validity of the Arc-sine Law for not identically distributed r.v.'s. Generalization of Spitzer's formula and of the equivalence principle. Toeplitz matrices of Laurent polynomials. Law of the iterated logarithm for the maximal order statistics.

INVESTIGATIONS ON FLUCTUATIONS OF SUMS OF RANDOM VARIABLES

E. SPARRE ANDERSEN

ABSTRACT: Distribution of the r.v. \( H \). Validity of the Arc-sine Law for not identically distributed r.v.'s. Generalization of Spitzer's formula and of the equivalence principle. Toeplitz matrices of Laurent polynomials. Law of the iterated logarithm for the maximal order statistics.