On certain representations of real numbers and on sequences of equivalent events

By ALFRED RÉNYI in Budapest

Dedicated to Professor L. Kalmár at the occasion of his 60th birthday

Introduction

In § 1 of this paper we shall deal with certain representations of real numbers. Let $a_n$ ($n = 0, 1, \ldots$) be an absolutely monotonic sequence of numbers, i.e. such that

(1) \[ \Delta^k a_n > 0 \quad (k = 0, 1, \ldots; \ n \geq k), \]

where $\Delta^0 a_n = a_n$, $\Delta^1 a_n = \Delta a_n = a_{n-1} - a_n$ ($n \geq 1$), and

\[ \Delta^k a_n = \Delta (\Delta^{k-1} a_n) \quad (k \geq 1, n \geq k). \]

Let us also suppose that the sequence is normed, i.e.

(2) \[ a_0 = 1, \]

and regular, i.e.

(3) \[ a_n \to 0 \quad \text{and} \quad \Delta^n a_n \to 0 \quad (n \to \infty). \]

Then every real $x \in (0, 1]$ admits a uniquely determined representation of the form

(4) \[ x = \sum_{k=0}^{\infty} \Delta^k a_n, \]

where the sequence of integers $1 \leq n_0 < n_1 < n_2 < \ldots$ depends on $x$.

This representation can also be written in the form

(5) \[ x = \sum_{n=1}^{\infty} e_n \Delta^{e_1 + \ldots + e_n-1} a_n, \]

where $e_n = e_n(x)$ equals 0 or 1; clearly $e_n = 1$ if the number $n$ occurs in the sequence $n_k$, and $e_n = 0$ if not.

In § 2 we deal with the probability distribution of $n_k = n_k(x)$ provided that $x$ is chosen at random with uniform distribution in $(0, 1]$. We shall show that the sequence of random variables $n_k$ is then a Markov chain. In § 3 we deal with the joint probability distribution of the random variables $e_n$ $(n = 1, 2, \ldots)$ defined above. We prove that if $A_n$ denotes the random event that $e_n(x) = 1$ then the events $A_n$
\( (n = 1, 2, \ldots) \) form a sequence of equivalent (symmetrically dependent) events, such that
\[
P(A_{m_1} A_{m_2} \ldots A_{m_k}) = \Delta^k a_k
\]
for \( 1 \leq m_1 < m_2 < \ldots < m_k \). (Here and in what follows \( P(A) \) stands for the probability of the event \( A \).

In § 4 we show that the strong law of large numbers for equivalent events implies that for almost all \( x \) the limit
\[
\lim_{k \to \infty} \frac{k}{n_k(x)}
\]
exists. (Previously in § 2 we obtain the weaker result that the distribution of \( k/n_k \) tends to a limit distribution.) On the other hand the above mentioned connection between equivalent events and the representation (4) or (5) leads to an effective construction of any sequence of equivalent events. A consequence of this is discussed in § 5. In § 6 we construct the corresponding measure preserving transformation to each sequence \( a_n \), while in § 7 we discuss an example.

## § 1. Representation of real numbers by series of successive differences

We start with the following

Theorem 1. Let \( a_n \ (n = 0, 1, \ldots) \) be a normed, regular, absolutely monotonic sequence of real numbers. Then any real number \( x \in (0, 1] \) can be represented in the form
\[
x = \sum_{k=0}^{\infty} \Delta^k a_{n_k}
\]
where the increasing sequence of natural numbers \( n_k \) is uniquely determined by \( x \).

Proof. Let \( n_0 \) be the first natural number such that
\[
a_{n_0} < x;
\]
such a number exists because \( a_0 = 1 \) and \( a_n \to 0 \). Let \( n_1 \) be the first natural number such that
\[
a_{n_0} + \Delta a_{n_1} < x;
\]
such a number \( n_1 \) exists because \( \Delta a_n \to 0 \). Moreover, by the definition of \( n_0 \) we have \( a_{n_0} < x \leq a_{n_0 - 1} \), i.e. \( x - a_{n_0} \leq \Delta a_{n_0} \), hence it follows \( n_1 > n_0 \). Similarly if \( n_0, n_1, \ldots, n_r \) are already determined so that
\[
\sum_{k=0}^{r} \Delta^k a_{n_k} < x \leq \sum_{k=0}^{r-1} \Delta^k a_{n_k} + \Delta^r a_{n_r-1},
\]
let \( n_{r+1} \) be the least natural number such that
\[
\sum_{k=0}^{r+1} \Delta^k a_{n_k} < x.
\]
It follows from (1.4) that

\[(1.6) \quad 0 < x - \sum_{k=0}^{r} A^k a_{n_k} < A^{r+1} a_{n_r}\]

which implies \(a_n\) as by supposition \(A^{r+1} a_n\) is decreasing in \(n\) that \(n_{r+1} > n_r\).

Thus \(n_{r+1} > r+1\). Therefore \(a_n\) using again the monotonicity of \(A^{r+1} a_n\) it follows from (1.6) that

\[(1.7) \quad 0 < x - \sum_{k=0}^{r} A^k a_{n_k} < A^{r+1} a_{n_r+1} \]

In view of the condition \(A^n a_n \to 0\) it follows that if the numbers \(n_k\) are determined by the algorithm described above, then (1.1) holds. This proves Theorem 1.

Let us note that according to a well-known theorem of F. HAUSDORFF [1] every normed absolutely monotonic sequence can be represented in the form

\[(1.8) \quad a_n = \int_0^1 t^n dF(t)\]

where \(F(t)\) is non-decreasing on the closed interval \([0, 1]\), is continuous from the left in the interior, and such that \(F(0) = 0\) and \(F(1) = 1\). Evidently,

\[\lim_{n \to \infty} a_n = F(1) - F(1 - 0),\]

thus condition \(\lim a_n = 0\) implies that \(F(t)\) is continuous at \(x = 1\). We have further

\[(1.9) \quad A^k a_n = \int_0^1 (1 - t)^k t^{n-k} dF(t)\]

for \(k = 0, 1, \ldots\) and \(n \geq k\); thus in particular

\[(1.10) \quad A^k a_k = \int_0^1 (1 - t)^k dF(t).\]

Hence

\[\lim_{k \to \infty} A^k a_k = F(1) = 0.\]

Thus the condition of regularity \(\lim A^n a_n = 0\) implies that \(F(t)\) is continuous at \(t = 0\). Thus every normed, regular, absolutely monotonic sequence \(a_n\) can be represented in the form (1.8) where \(F(t)\) is the distribution function of a probability distribution in the open interval \((0, 1)\).

In view of formula (1.9) the representation (1.1) can be written in the form

\[(1.11) \quad x = \int_0^1 \left( \sum_{k=0}^{\infty} (1 - t)^k t^{n_k - k} \right) dF(t).\]
Thus it follows that for every \( x \in (0, 1) \) there is exactly one function \( g(t) \) of the form

\[
g(t) = \sum_{k=0}^{\infty} (1-t)^k \ell_{m_k-k}
\]

with \( 1 \leq n_0 < n_1 < n_2 < \ldots \), such that

\[
x = \int_{0}^{1} g(t) \, d\ell(t).
\]

\[
\secref{2} \quad \text{Statistical theory of the difference-series representation of real numbers}
\]

Let \( x \) be a random variable, uniformly distributed in the interval \((0, 1)\). Let us consider the representation of \( x \) in the form

\[
(2.1) \quad x = \sum_{k=0}^{\infty} A_k a_n_k
\]

where \( a_n \) is a given normed, regular, absolutely monotonic sequence. According to Theorem 1 the natural numbers \( n_k = n_k(x) \) are uniquely determined by \( x \); thus they are well defined random variables. We shall study now the probability laws governing the behaviour of these random variables. It is easy to see that if \( n_1, \ldots, n_k \) are fixed, then \( x \) belongs to an interval of length \( A_k a_n_k \). It follows that denoting by \( P(A|B) \) the conditional probability of the event \( A \) under condition \( B \), we have

\[
(2.2a) \quad P(n_k = n | n_0 = m_0, n_1 = m_1, \ldots, n_{k-1} = m_{k-1}) = \frac{A_k a_n}{A_k a_{n_k-1}}
\]

provided that \( 1 \leq m_0 < m_1 < \ldots < m_{k-1} < n \). Thus the conditional distribution of \( n_k \) by given \( n_0, \ldots, n_{k-1} \) depends on \( n_{k-1} \) only; that is the sequence of random variables \( n_k (k = 0, 1, \ldots) \) is a Markov chain with the transition probabilities

\[
(2.2b) \quad P(n_k = n | n_{k-1} = m) = \frac{A_k a_n}{A_k a_{m}}.
\]

As the probability on the right-hand side of (2.2b) depends in general on \( k \) too, the Markov chain \( n_k \) is in general inhomogeneous. It is easy to see that the Markov chain is homogeneous if and only if \( a_n = (1-p)^n (n = 0, 1, \ldots) \) where \( 0 < p < 1 \). In this particular case

\[
(2.3) \quad P(n_k = n | n_{k-1} = m) = p(1-p)^{n-m-1}.
\]

This particular case corresponds to the representation of the real number \( x \) in the form

\[
(2.4) \quad x = \sum_{k=0}^{\infty} p^k (1-p)^{n_k-k}.
\]

In this case if \( A_n \) denotes the event that \( n \) is contained in the sequence \( n_k \) then the
Certain representations of real numbers

67

events $A_n (n=1, 2, \ldots)$ are independent and each has the probability $P(A_n) = p$. Especially if $p = \frac{1}{2}$ the representation (2.4) reduces to

$$x = \sum_{k=0}^{\infty} \frac{1}{2^n_k} \quad \text{where} \quad 1 \leq n_0 < n, < \ldots,$$

by other words to the representation of $x$ in the binary number system.

Let us return to the random variables $n_k$ in the general case. The unconditional distribution of $n_k$ can be determined as follows: As mentioned above if $n_0, n_1, \ldots, n_k$ are fixed, then $x$ belongs to an interval of length $A^{k+1}a_n$. Now if only $n_k$ is fixed, $n_k = n$, then the values of $n_0, n_1, \ldots, n_k-1$ can be chosen in $\binom{n-1}{k}$ different ways; thus we have

$$P(n_k = n) = \binom{n-1}{k} A^{k+1}a_n.$$

Especially in the case when $a_n = (1-p)^n$, we have

$$P(n_k = n) = \binom{n-1}{k} p^{k+1}(1-p)^{n-k-1}$$

i.e. $n_k - k - 1$ has a negative binomial distribution of order $k+1$.

In the general case it follows from (2.6) and (1.9) that

$$P(n_k = n) = \binom{n-1}{k} \int_0^1 (1-t)^{k+1} t^{n-k-1} dF(t)$$

for $n \leq k+1$.

The distribution (2.8) may be called a mixed negative binomial distribution of order $k+1$. The characteristic function of $\frac{n_k}{k+1}$ is

$$M \left( e^{it} \right) = e^{iun} \int_0^1 \left( \frac{1-t}{1-te^{it}} \right)^{k+1} dF(t).$$

(Here and in what follows $M$ stands for “expectation”.

We obtain by passing to the limit

$$\lim_{k \to \infty} M \left( e^{it} \right) = \int_0^1 e^{iu} dF(t).$$

It follows that the probability distribution of $\frac{k+1}{n_k}$ tends to the distribution having the distribution function $1 - F(1 - z)$

$$\lim_{k \to \infty} P \left( \frac{k}{n_k} \leq z \right) = 1 - F(1 - z).$$
In the special case \( a_n = (1 - p)^n \) we have

\[
F(t) = \begin{cases} 
0 & \text{if } t \leq 1 - p \\
1 & \text{if } t > 1 - p,
\end{cases}
\]

thus (2.11) implies that in this case \( k/n_k \) tends in probability to \( p \). This is of course well known, because

\[
\frac{k}{n_k} = \frac{e_1 + e_2 + \ldots + e_{n_k}}{n_k},
\]

and the (weak) law of large numbers applies to the independent random variables \( e_n \), each having the expectation \( p \).

We shall show in the next paragraph that much more is true than (2.11): not only does the distribution of \( k/n_k \) tend for \( k \to \infty \) to the limit distribution \( 1 - F(1 - z) \), but the random variables \( k/n_k \) themselves tend for \( k \to \infty \) with probability 1 to a random variable \( \kappa \) having the distribution function \( 1 - F(1 - z) \).

Using the formula (2.8) we can of course compute all the moments of \( n_k \). Especially we have

\[
M(n_k) = (k + 1) \int_0^1 \frac{dF(t)}{1 - t}.
\]

Thus the expectation of \( \frac{n_k}{k + 1} \) does not depend on \( k \); it is finite if and only if the integral on the right of (2.13) is convergent, otherwise it is equal to \( +\infty \).

§ 3. Connection with the theory of equivalent events

Let \( A_n \) denote the event that the natural number \( n \) is contained in the sequence \( n_k(x) \) where \( x \) is a random variable, uniformly distributed in the interval \((0, 1)\). We have evidently

\[
P(A_n) = \sum_{k=0}^{n-1} P(n_k = n).
\]

It follows from (2.8) that

\[
P(A_n) = \int_0^1 \left( \sum_{k=0}^{n-1} \frac{n-1}{k} (1-t)^{k+1} t^{n-k-1} \right) dF(t) = \int_0^1 (1-t) dF(t)
\]

for \( n = 1, 2, \ldots \). Before proceeding further we have to compute the \( r \)-step transition probabilities of the Markov-chain \( n_k \). Clearly we have for \( r \geq 2 \) and \( n \geq m + r \)

\[
P(n_{k+r} = n | n_k = m) = \frac{A_{k+r+1} a_n}{A_{k+1} a_m} \sum_{m < n_1 < \ldots < n_{r-1} < n} 1,
\]

thus

\[
P(n_{k+r} = n | n_k = m) = \left( \frac{n - m - 1}{r - 1} \right) \frac{A_{k+r+1} a_n}{A_{k+1} a_m}.
\]

All rights reserved © Bolyai Institute, University of Szeged
Certain representations of real numbers 69

It follows that for $m\preceq n$

\[(3.5) \quad P(n_{k+r} = n, n_k = m) = \binom{n-m-1}{r-1} \binom{m-1}{k} A^{k+r+1} a_n.\]

Thus we have

\[(3.6) \quad P(A^m A_n) = \sum_{k=0}^{m-1} \sum_{l=k+1}^{m-n-1} \binom{n-m-1}{l-k-1} \binom{m-1}{k} A^{l+1} a_n.\]

Taking (1.9) into account we obtain for $1 \preceq m < n$

\[(3.7) \quad P(A^m A_n) = \int_0^1 (1-t)^2 dF(t).\]

We shall show now that for any $r \preceq 1$ and for $1 \preceq m_1 < m_2 < \ldots < m_r$ we have

\[(3.8) \quad P(A_{m_1} A_{m_2} \ldots A_{m_r}) = \int_0^1 (1-t)^r dF(t).\]

The proof is essentially the same as for $r=2$. We obtain in the same way as (3.5) was shown — using that $n_k$ is a Markov chain — that for $k_1 < k_2 < \ldots < k_r$, $m_1 < m_2 < \ldots < m_r$

\[(3.9) \quad P(n_{k_1} = m_1, \ldots, n_{k_r} = m_r) = \binom{m_1-1}{k_1} \prod_{j=1}^{r-1} \binom{m_{j+1} - m_j - 1}{k_{j+1} - k_j - 1} A^{k_r+1} a_{m_r}.\]

Of course the probability (3.9) is positive only if $m_1 \preceq k_1 + 1$ and $m_{j+1} - m_j \preceq k_{j+1} - k_j$ ($j = 1, 2, \ldots, r-1$). From (3.9) one obtains (3.8) by means of the identity

\[(3.10) \quad P(A_{m_1} A_{m_2} \ldots A_{m_r}) = \sum_{k_1 < k_2 < \ldots < k_r} P(n_{k_1} = m_1, \ldots, n_{k_r} = m_r).\]

As clearly

\[(3.11) \quad \int_0^1 (1-t)^r dF(t) = A^r a_r,\]

we have proved the following

**Theorem 2.** Let $A_n$ denote the event that the natural number $n$ is contained in the sequence $\{n_\xi(x)\}$ defined by Theorem 1, where $x$ is a random variable uniformly distributed in the interval $(0, 1)$. Then the events $A_n$ ($n=1, 2, \ldots$) are equivalent, and one has for $1 \preceq m_1 < m_2 < \ldots < m_r$ ($r=1, 2, \ldots$)

\[(3.12) \quad P(A_{m_1} A_{m_2} \ldots A_{m_r}) = A^r a_r.\]

**Remark.** Note that the sequence $\bar{w}_r = A^r a_r$ is absolutely monotonic too, because setting

\[(3.13) \quad G(t) = 1 - F(1-t+0)\]

we have

\[(3.14) \quad \bar{w}_r = \int_0^1 t^r dG(t).\]
It is easy to see also that

\[(3.15)\quad A^h w_r = A^{r-h} a_r.\]

Conversely let us be given a sequence of equivalent events \(B_n\) \((n = 1, 2, \ldots)\) in a probability space \([\Omega, \mathcal{A}, P]\) where \(\Omega\) is a non-empty set, the generic element of which will be denoted by \(\omega\), \(\mathcal{A}\) is a \(\sigma\)-algebra of subsets of \(\Omega\) and \(P\) a probability measure on \(\mathcal{A}\). It is known (see [2], [3]) that there exists an \(\mathcal{A}\)-measurable function \(\beta = \beta(\omega)\) on \(\Omega\) — called the density of the sequence of events \(B_n\) — such that \(0 \leq \beta \leq 1\) and for \(r = 1, 2, \ldots\) and \(m_1 < m_2 < \ldots < m_r\) one has

\[(3.16)\quad P(B_{m_1}B_{m_2} \ldots B_{m_r}) = \int_{\Omega} \beta^r \, dP.\]

Let us consider first the case when \(\beta = 1\) on a set \(B\) of positive probability. Let \(\beta_n\) denote the indicator of the set \(B_n\). It was shown in [3] that if \(n_1 < n_2 < \ldots < n_k\) \(m_1 < m_2 < \ldots < m_l\) and \(n_j \neq m_j\), then

\[(3.17)\quad P(B_{n_1}B_{n_2} \ldots B_{n_k}B_{m_1}B_{m_2} \ldots B_{m_l}) = \int_{\Omega} \beta_k \beta_{m_1} \beta_{m_2} \ldots \beta_{m_l} \, dP.\]

It follows that

\[(3.18)\quad P\left(\prod_{n=r}^{\infty} B_n\right) = \int_{B} \prod_{r \leq j < s} \beta_j \, dP.\]

As (3.18) holds for \(s = r\) too (the empty product is equal to 1), we have

\[P\left(\prod_{n=r}^{\infty} B_n\right) = P(B).\]

Thus we obtain, putting \(\prod_{n=1}^{\infty} B_n = B^*\),

\[P(B) = P(B^*) = P(BB^*).\]

This implies that the sets \(B\) and \(B^*\) are identical up to a set of \(P\)-measure 0. Let us denote now by \(\bar{B}\) the complementary event of \(B\), i.e. \(\bar{B} = \Omega - B\). It follows that the events \(A_n = \bar{B}B_n\) also are equivalent, and have the density \(\alpha\) defined as follows:

\[\alpha(\omega) = \begin{cases} 
\beta(\omega) & \text{if } \omega \in \bar{B}, \\
0 & \text{if } \omega \in B.
\end{cases}\]

As a matter of fact we have

\[P(A_{m_1}A_{m_2} \ldots A_{m_l}) = P(B_{m_1}B_{m_2} \ldots B_{m_l}) - P(B) = \int_{\Omega} \alpha^k \, dP.\]

As \(P(\alpha = 1) = 0\), we have shown that without restriction of generality one can suppose that \(P(\beta = 1) = 0\).

Similarly one can suppose without restricting the generality that \(P(\beta = 0) = 0\). As a matter of fact if \(C\) denotes the set on which \(\beta = 0\) and \(0 < P(C) < 1\) then the set \(C\) is disjoint to all the sets \(B_n\) (up to a set of probability 0) and thus instead of the probability space \([\Omega, \mathcal{A}, P]\) we may consider the space \([\Omega, \mathcal{A}, P^*]\) where
\[ p^k(A) = \frac{P(A \cap C)}{P(C)} \] and the events \( B_n \) will be equivalent with respect to this probability space too, with the same density \( \beta \). Thus the case of an arbitrary sequence of equivalent events can be reduced to a sequence of equivalent events the density \( \beta \) of which is such that \( P(\beta = 0) = P(\beta = 1) = 0 \). Let us call such a sequence a \textit{regular} sequence of equivalent events. If \( A_n \) is a regular sequence of equivalent events with density \( \beta \) and if we put

\[ w_k = P(A_{n_1} A_{n_2} \ldots A_{n_k}) = \int \beta^k dP \]

then clearly we have

\[ \lim_{k \to \infty} w_k = 0 \quad \text{and} \quad \lim_{k \to \infty} \Delta^k w_k = 0. \]

Putting \( a_k = \Delta^k w_k \), clearly \( w_k = \Delta^k a_k \) and the sequence \( a_k \) is a normed regular absolutely monotonic sequence. Thus the events \( \{ A_n \} \) can be realized as the events connected with the representation of the random real number \( x \) uniformly distributed in \((0, 1)\) in the form \((1, 1)\), so that the event \( A_n \) is identified with the event that \( n \) is contained in the sequence \( n_k \).

\[ \text{§ 4. The strong law of large numbers for the Markov chain } n_k \]

We first give — to make this paper self-contained — a short proof of the following known result:

\textbf{Theorem 3.} \( A_n \) be an arbitrary sequence of equivalent events; let \( \alpha_n \) denote the indicator of \( A_n \) and \( \alpha \) the density of the sequence \( A_n \). Then we have

\[ P \left( \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \alpha_k = \alpha \right) = 1. \]

\textbf{Proof.} Let us consider the random variables

\[ \delta_k = \alpha_k - \alpha \]

and let us put for \( k_1 < k_2 < \ldots < k_r \), \( r = 1, 2, \ldots \)

\[ (4.1) \]

\[ P(A_{k_1} A_{k_2} \ldots A_{k_r}) = w_r . \]

It follows from (3.17) that

\[ (4.2) \]

\[ M(\delta_{k_1}, \delta_{k_2}, \delta_{k_3}, \delta_{k_4}) = \begin{cases} A & \text{if } k_1 = k_2 = k_3 = k_4 , \\
B & \text{if } k_1 = k_2 \text{ and } k_3 = k_4 \neq k_1 , \\
or if & k_1 = k_3 \text{ and } k_2 = k_4 \neq k_1 , \\
or if & k_1 = k_4 \text{ and } k_2 = k_3 \neq k_1 , \\
0 & \text{otherwise,} \end{cases} \]

\[ 2) \text{ Theorem 3 can also be deduced from Birkhoff’s ergodic theorem.} \]
where
\begin{align}
A &= w_1 - 4w_2 + 6w_3 - 3w_4 \\
B &= w_2 - 2w_3 + w_4.
\end{align}
and
\begin{align}
\sum_{n=1}^{\infty} \left( \frac{\delta_1 + \ldots + \delta_n}{n} \right)^4 = 0 \left( \frac{1}{n^2} \right).
\end{align}
This implies
\begin{align}
M \left( \left( \frac{\delta_1 + \ldots + \delta_n}{n} \right)^4 \right) = 0 \left( \frac{1}{n^2} \right).
\end{align}
Thus the series
\begin{align}
\sum_{n=1}^{\infty} \left( \frac{\delta_1 + \ldots + \delta_n}{n} \right)^4
\end{align}
is convergent with probability 1 and therefore (4.1) holds.

In view of (2.11) and the results of § 3 this implies that the following theorem holds:

**Theorem 4.** If the sequence \( n_k(x) \) is defined according to Theorem 1 then the limit
\begin{align}
\lim_{k \to \infty} \frac{k}{n_k(x)} = \kappa(x)
\end{align}
exists for almost all \( x \) in \((0, 1)\); denoting by \( \mu(A) \) the Lebesgue measure of the set \( A \) one has
\begin{align}
\mu(\kappa(x) \equiv y) = 1 - F(1 - y) \quad \text{for} \quad 0 \leq y \leq 1.
\end{align}

§ 5. Consequences for equivalent events

In the preceding § we applied the theory of equivalent sequences of events to prove the existence almost everywhere of the limit \( \lim_{k \to \infty} \frac{k}{n_k(x)} \). Conversely, our results lead to the proof of a property of equivalent events which seems not to be noticed up to now. This is expressed by

**Theorem 5.** Let \( A_n \) \((n=1, 2, \ldots)\) be a regular sequence of equivalent events. Let us set
\begin{align}
P(A_{n_1} A_{n_2} \ldots A_{n_k}) = w_k \quad (n_1 < n_2 < \ldots < n_k; \ k = 1, 2, \ldots).
\end{align}
Denote by \( \alpha_n \) the indicator of the event \( A_n \) and define the random variables \( v_k \) as follows: \( v_k \) is the least value of \( n \) such that \( \alpha_1 + \alpha_2 + \ldots + \alpha_n = k \). By other words, \( v_k \) denotes the index of the \( k \)-th event in the sequence of events \( A_n \) \((n=1, 2, \ldots)\) which takes place. Then the random variables \( v_k \) form a Markov chain with the transition probabilities
\begin{align}
P(v_{k+1} = n | v_k = m) = \frac{A^{n-k-1} w_n}{A^{m-k} w_m}.
\end{align}
§ 6. The measure preserving transformation corresponding to a series
of successive differences

To every representation (7) — i.e. to every normed, regular, absolutely monotonous sequence \( \{a_n\} \) — there corresponds a measure preserving transformation \( T \)
of the interval \((0, 1)\) defined as follows: If

\[
\chi = \sum_{k=0}^{\infty} A^k a_{n_k}^k(x)
\]

then

\[
n_k(T\chi) = n_{k+n_1}(x) = 1 \quad (k = 0, 1, \ldots).
\]

Clearly \( 1 \leq n_0(T\chi) \), because if \( n_0(x) = 1 \) then \( e_1(x) = 1 \) and thus \( n_0(T\chi) = n_1(x) - 1 \approx 1 \) and if \( n_0(x) \approx 2 \) then \( n_0(T\chi) = n_0(x) - 1 \approx 1 \); the inequality \( n_{k+1}(T\chi) > n_k(T\chi) \) is evident. The inverse transformation \( T^{-1}y \) can be defined as follows: \( T^{-1}y \) is two-valued, namely if

\[
y = \sum_{k=0}^{\infty} A^k a_{n_k}
\]

then \( T^{-1}y \) has the two values \( x_1 \) and \( x_2 \) where

\[
x_1 = \sum_{k=0}^{\infty} A^k a_{n_k+1}, \quad x_2 = a_1 + \sum_{k=1}^{\infty} A^k a_{n_k-1+1}.
\]

Clearly if \( y \) belongs to the interval \( I_r \) defined by fixing the values of \( n_0, n_1, \ldots, n_r \)
in (6.3) \((1 \leq n_0 < n_1 < \ldots < n_r)\) and having the length \( A^{r+1} a_{n_r+1} \), then \( x_1 \) belongs to an interval \( I'_r \) of length \( A^{r+1} a_{n_r+1} \) and \( x_2 \) to an interval \( I''_r \) of length \( A^{r+2} a_{n_r+1} \). As

\[
A^{r+1} a_{n_r+1} + A^{r+2} a_{n_r+1} = A^{r+1} a_{n_r}
\]

it follows that denoting by \( \mu(A) \) the Lebesgue measure of the set \( A \) one has

\[
\mu(T^{-1}I_r) = \mu(I'_r) + \mu(I''_r) = \mu(I_r).
\]

It follows from (6.6) that \( T\chi \) is measure preserving.
The transformation \( T\chi \) can of course also be defined by

\[
v_k(T\chi) = e_{k+1}(x) \quad (k = 0, 1, \ldots).
\]

Thus \( T \) is equivalent to the shift transformation in the sequence-space \((e_1, e_2, \ldots, e_n, \ldots)\).

It is easy to see that the transformation \( T \) is ergodic if and only if \( a_n = q^n \) with
\( 0 < q < 1 \), because it follows from (6.7) and Theorem 4 that

\[
\kappa(T\chi) = \kappa(x)
\]

and thus each level set of \( \kappa \) is an invariant set of \( T \); thus \( T \) is ergodic if and only if
\( \kappa \) is constant almost everywhere, i.e. if \( a_n = q^n \). Especially in the case \( a_n = 2^{-n} \), \( T \)
is the well known transformation \( T\chi = 2\chi \) where \((Z)\) denotes the fractional part of \( Z \).
§ 7. An example

As an example let us consider the sequence \( a_n = \frac{1}{n+1} \) \((n = 0, 1, 2, \ldots)\). Evidently,

\[
\Delta^k a_n = \frac{1}{(n+1) \binom{n}{k}},
\]

hence

\[
\Delta^n a_n = a_n = \frac{1}{n+1}.
\]

Thus \( a_n \) is a normed, regular, absolutely monotonic sequence. Theorem 1 asserts for this case that every real number \( x \) with \( 0 < x \leq 1 \) has a unique representation of the form

\[
x = \sum_{k=0}^{\infty} \frac{1}{(n_k+1) \binom{n_k}{k}}
\]

where the \( n_k \) are integers, \( 1 \leq n_1 < n_2 < \ldots \). The function \( F(t) \) figuring in (1.8) is in this example equal to \( t \) \((0 \leq t \leq 1)\). The transition probabilities (2.2b) are in this example

\[
P(n_k = n | n_{k-1} = m) = \frac{(m+1) \binom{m}{k}}{(n+1) \binom{n}{k+1}}
\]

and the distribution of \( n_k \) is given by

\[
P(n_k = n) = \frac{k+1}{n(n+1)} \quad \text{for} \quad n \geq k+1.
\]

Thus the random variables \( n_k \) have an infinite expectation. The equivalent events \( A_n \) can in this case be interpreted as the events of the following Pólya urn model: Let us consider an urn containing one white and one red ball. Let us draw one of the balls at random (each having the probability \( \frac{1}{2} \) to be drawn) and put it back into the urn together with another ball of the same colour, then draw another ball from theurn which now contains 3 balls, each ball having the same probability to be drawn, put it back together with another ball of the same colour and continue this procedure indefinitely. Let \( A_n \) denote the event that at the \( n \)-th occasion a red ball has been drawn from the urn. Clearly in this interpretation a red ball is drawn the \( k+1 \)-st time at the \( n_k \)-th drawing; the limit \( x \) of \( k/n_k \) is in this case of course uniformly distributed in the interval \((0, 1)\).

References


(Received January 15, 1965)