On tables of random numbers

A.N. Kolmogorov

Academy of Sciences. USSR

1. Introduction

The set theoretic axioms of the calculus of probability, in formulating which I had
the opportunity of playing some part (Kolmogorov, 1950), had solved the majority of
formal difficulties in the construction of a mathematical apparatus which is useful for
a very large number of applications of probabilistic methods, so successfully that the
problem of finding the basis of real applications of the results of the mathematical
theory of probability became rather secondary to many investigators.

I have already expressed the view [see Kolmogorov (1950), Chapter I] that the
basis for the applicability of the results of the mathematical theory of probability to
real 'random phenomena' must depend on some form of the frequency concept of
probability, the unavoidable nature of which has been established by von Mises in a
spirited manner. However, for a long time I had the following views:

(1) The frequency concept based on the notion of limiting frequency as the number
of trials increases to infinity, does not contribute anything to substantiate the applic-
ability of the results of probability theory to real practical problems where we
have always to deal with a finite number of trials.

(2) The frequency concept applied to a large but finite number of trials does not admit
a rigorous formal exposition within the framework of pure mathematics.

Accordingly I have sometimes put forward the frequency concept which involves
the conscious use of certain not rigorously formal ideas about 'practical reliability',
'approximate stability of the frequency in a long series of trials', without the precise
definition of the series which are 'sufficiently large' etc. [see Foundations of the Tho-
ory of Probability, Chapter I and for more details Great Soviet Encyclopaedia (section
on Probability) and Mathematika iou metod i Znachenye (Chapter on Probability
Theory)].

I still maintain the first of the two theses mentioned above. As regards the second,
however, I have come to realise that the concept of random distribution of a property
in a large finite population can have a strict formal mathematical exposition. In fact,
we can show that in sufficiently large populations the distribution of the property may
be such that the frequency of its occurrence will be almost the same for all sufficiently
large sub-populations, when the law of choosing these is sufficiently simple. Such a conception in its full development requires the introduction of a measure of the complexity of the algorithm. I propose to discuss this question in another article. In the present article, however, I shall use the fact that there cannot be a very large number of simple algorithms.

For definiteness we shall consider the table

$$T = (t_1, t_2, \ldots, t_N)$$

of $N$ zeros and ones: $t_k = 0$ or $1$.

Such a table will be called random, if, while choosing the subset $A$ of sufficiently large size from $1, N$ by different methods there is a stability in the frequency

$$\pi(A) = \frac{1}{n} \sum_{k \in A} t_k$$

of appearance of ones in $A$. One can, for example, choose $A$ as

(a) the set of first $n$ even integers 2, 4, 6, \ldots, $2n$,

(b) the set of first $n$ prime numbers $p_1, p_2, \ldots, p_n$ and so on.

The ordinary notion of 'randomness' of a table $T$ does not consist merely of the stability of the frequencies while choosing $A$ by methods entirely independent of the composition of the table $T$. One can for example, choose the set $A$ as

(c) the set of first $n$ values $k \geq 2$ for which $t_{k-1} = 0$,

(d) the set of first $n$ values $k > s$ for which $t_{k-1} = a_1, t_{k-2} = a_2, \ldots, t_{k-s} = a_s$,

(e) the set of the first $n$ even numbers $k = 2i$ for which $t_i = 1$,

(f) the set of numbers $k_1, k_2, \ldots, k_n \ldots$ chosen according to the law

$$k_1 = 1,$$

$$k_{i+1} = k_i + 1 + t_k p_i$$

and so on.

The precise formulation of the concept of 'admissible algorithm' of choosing the set $A$ will be given in Section 2.

If while using a table of sufficiently large size $N$ at least one single test of randomness of this type with sufficiently large size of the sample $n$ leads to a 'significant' departure from the principle of frequency stability then we immediately reject the hypothesis of 'pure random' origin of the given table.

2. Admissible algorithms of selection and $(n, \varepsilon)$-random tables

An admissible algorithm of choosing the set

$$A = R(T) \subset \overline{1, N}$$
according to the table $T$ of size $N$ is defined by the functions\textsuperscript{2} 

\begin{align*}
F_0, G_0, H_0 \\
F_1(\xi_1, \tau_1), G_1(\xi_1, \tau_1), H_1(\xi_1, \tau_1) \\
F_2(\xi_1, \tau_1; \xi_2, \tau_2), G_2(\xi_1, \tau_1; \xi_2, \tau_2), H_2(\xi_1, \tau_1; \xi_2, \tau_2) \\
\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\
F_{N-1}(\xi_1, \tau_1; \xi_2, \tau_2; \ldots; \xi_{N-1}, \tau_{N-1}), G_{N-1}(\xi_1, \tau_1; \xi_2, \tau_2; \ldots; \xi_{N-1}, \tau_{N-1}), \\
H_{N-1}(\xi_1, \tau_1; \xi_2, \tau_2; \ldots; \xi_{N-1}, \tau_{N-1})
\end{align*}

where the arguments $\tau_k$ and the functions $G_k$ and $H_k$ take values 0 or 1 and the arguments $\xi_k$ and functions $F_k$ take values from $1, N$. The functions $F_k$ are subject to an additional condition

$$F_k(\xi_1, \xi_1, \ldots; \xi_k, \xi_k) \neq \xi_k. \quad (2.1)$$

Defining an algorithm is equivalent to forming the sequence

\begin{align*}
x_1 &= F_0, \\
x_2 &= F_1(x_1, t_{x_1}), \\
x_3 &= F_2(x_1, t_{x_1}; x_2, t_{x_2}), \\
\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\
x_s &= F_{s-1}(x_1, t_{x_1}; \ldots; x_{s-1}, t_{x_{s-1}}) \quad (2.2)
\end{align*}

and determining those elements of the sequence which are found in $A$. The sequence terminates as soon as the value\textsuperscript{3}

$$H_s(x_1, t_{x_1}; \ldots; x_s, t_{x_s}) = 1 \quad (2.3)$$

appears. In this case the sequence terminates with the element $x_s$. If when $k < N$ we have all the time

$$H_k(x_1, t_{x_1}; \ldots; x_k, t_{x_k}) = 0,$$

the sequence is terminated by the element $x_s$ with $s = N$, i.e., by exhausting all the elements of the set $I, N$: in view of the condition (2.1) all the elements of the sequence (2.2) are distinct.

The set $A$ is formed from those $x_k$ for which

$$G_{k-1}(x_1, t_{x_1}; \ldots; x_{k-1}, t_{x_{k-1}}) = 1. \quad (2.4)$$

It seems to me that the given construction correctly reflects the basic concept of von Mises in its complete generality, preserving, however, the basic limitation that for determining whether $x \in I, N$ falls in the set $A$ the value of $t_x$ is not used.

\textsuperscript{2} The functions in the first line are constants (functions on the empty set of arguments).

\textsuperscript{3} In particular, if $H_0 = 0$ then the selection cannot begin and the set $A$ is found to be empty.
Now let the system
\[ \mathcal{R}_N = \{ R \} \]
of admissible algorithms of selection (the size \( N \) of the table being fixed) be given.

**Definition.** The table \( T \) of size \( N \) is called \((n, \varepsilon)\)-random with respect to the system \( \mathcal{R}_N \), if there exists a constant \( p, 0 < p < 1 \), such that for any
\[ A = R(T), \ R \in \mathcal{R}_N \]
with the number of elements
\[ V \geq n, \]
the frequency
\[ \pi(A) = \frac{1}{V} \sum_{k \in A} t_k \]
satisfies the inequality
\[ |\pi(A) - p| \leq \varepsilon. \]

Sometimes, it is convenient to say \((n, \varepsilon, p)\)-randomness, assuming that the constant \( p \) is fixed. Then the following theorem holds.

**Theorem 1.** If the number of elements of the system \( \mathcal{R}_N \) does not exceed
\[ \tau(n, \varepsilon) = \frac{1}{2} e^{2n\varepsilon^2} \] (2.5)
then for any \( p, 0 \leq p \leq 1 \), there exists a table \( T \) of size \( N \) that is \((n, \varepsilon, p)\)-random with respect to \( \mathcal{R}_N \).

The interpretation of the estimate, contained in the theorem, is made more transparent, if we introduce the binary logarithm
\[ \lambda(\mathcal{R}_N) = \log_2 \rho(\mathcal{R}_N) \]
of the number of elements \( \rho \) of the system \( \mathcal{R}_N \). \( \lambda(\mathcal{R}_N) \) is equal to the quantity of information, which is necessary for choosing an individual element \( R \) from \( \mathcal{R}_N \). It is clear that in the case of large \( \lambda(\mathcal{R}_N) \) the system \( \mathcal{R}_N \) must contain algorithms, the very determination (and not merely the actual realisation) of which is complicated (requires for its formulation not less than \( \lambda(\mathcal{R}_N) \) binary symbols).

In our theorem the condition of existence of tables which are \((n, \varepsilon)\)-random with respect to \( \mathcal{R}_N \) with arbitrary \( p \) is written in the form of the inequality
\[ \lambda(\mathcal{R}_N) \leq 2(\log_2 e) ne^2(1 - \varepsilon) - 1. \] (2.6)
Such a qualitative formulation of the result contained in the theorem is instructive by itself. If the ratio \( \lambda/n \) is sufficiently small then for any previously given \( \varepsilon \) and any \( N \) and \( p \) there exist tables which are \((n,\varepsilon)\)-random with respect to any system of admissible algorithms with

\[
\lambda(\mathcal{R}_n) \leq \lambda.
\]

The proof of this theorem will be given in Section 3. In Section 4, we shall examine the possibility of improving the estimates contained in the theorem. Now we make two supplementary remarks.

**Remark 1.** Since the algorithm of choosing the set \( A = R(T) \) is determined by the functions \( F_k, H_k, G_k \) it is natural to consider two algorithms to be same when and only when their corresponding functions \( F_k, H_k, G_k \) coincide. Already from this point of view the number of distinct possible algorithms of selection for a given \( N \) is finite.

It is possible to hold on to a different point of view and consider two algorithms of selection to be different only in the case when they give different sets \( A = R(T) \) at least for one table \( T \). From such a point of view the number of distinct algorithms is further reduced. But in any case it is not greater than

\[
(2^N)^N = 2^{N^2}.
\]

The question of precise estimation of the number of admissible algorithms under the second approach is not so simple. The problem is very simple only for algorithms, by which the set \( A \) is formed independently of the properties of the table \( T \). Distinct number of such algorithms is equal to \( 2^N \) according to the number of different sets \( A \in \mathcal{A}_N \).

**Remark 2.** The admissible algorithms of selection from the set of all possible natural numbers was considered by Church (1940). Now, in our definition, instead of the finite table \( T \) we consider an infinite sequence of zeroes and ones, \( \{ t_1, t_2, \ldots \} \).

We assume that the values of the arguments \( \xi_k \) and the function \( F_k \) are arbitrary natural numbers. But we reject the requirement that the selection must stop at \( s = N \) and instead assume that any (now infinite) table of functions \( F_k, G_k, H_k \) is 'computable' in the sense sufficiently well-known in all the numerous propositions for such formal definitions. Under these considerations we obtain the inessential generalisation of Church's concept. The basis of Church's result is the existence, for any \( p \), of sequences \( t_1, t_2, \ldots, t_N \ldots \) the density of which is equal to \( p \) in any infinite \(^4\) set \( A \) obtained by an admissible algorithm.

---

\(^4\) In this concept of Church substantial interest lies only in algorithms which extend infinitely. That is why, in this case, the functions \( G_k \) and all that is connected with these functions must be omitted.
3. Proof of Theorem 2

This result belongs to the Theory of Finite Algorithms and its formulation does not contain any concept borrowed from Probability Theory. If, in proving this, we make use of certain results of Probability Theory then this proof will have a formal character as it would only include a certain distribution of ‘weights’ in the set of tables $T$ of size $N$, the weight

$$P(T) = p^M (1 - p)^{N-M}$$

being assigned to the table containing $M$ ones. This method of proof does not affect the logical nature of the theorem itself, and does not hinder its use in the discussions needed for defining the domain of applicability of Probability Theory.

In another paper we shall prove the following inequality relating to the ‘Bernoulli Scheme’:

$$P \left( \sup_{k \geq n} \left| \frac{\mu_k}{n} - p \right| \geq \varepsilon \right) \leq 2e^{-2n\varepsilon^2(1-\varepsilon)} \quad (3.1)$$

Here $p$ is the probability of success in each of a sequence of independent trials; $\mu_k$ is the number of successes in the first $k$ trials. We can easily derive the following corollary from (3.1).

**Corollary.** Let $^5$

$$P(\xi_k = 1 \mid k \leq v, \xi_1, \ldots, \xi_{k-1}) = p$$

where $\xi_1, \xi_2, \ldots, \xi_{v}$ is a sequence of a random number of random quantities and $p$ is a constant. Then

$$P \left( v \geq n, \left| \frac{\mu_v}{v} - p \right| \geq \varepsilon \right) \leq 2e^{-2v\varepsilon^2(1-\varepsilon)} \quad (3.2)$$

We shall now examine the system $\mathcal{R}_N$ of admissible algorithms, $\rho$ in number.

We consider a table formed randomly with probability $p$ for $t_x = 1$ independently of the values taken by the other $t_{x'}$. If we fix $R \in \mathcal{R}_N$ and denote by

$$\xi_1, \xi_2, \ldots, \xi_{v}$$

those elements of the sequence

$$x_1, x_2, \ldots, x_{v}$$

which fall in $A = R(T)$ (numbering them as they appear in the course of the algorithm) it can easily be seen that the conditions under which (3.2) is valid are fulfilled. Hence the probability that, for any given $R \in \mathcal{R}_N$, the number of elements $v$ of the set $A$ will

---

$^5$ We are concerned here with the conditional probability that $\xi_k = 1$ when $k = v$ and $\xi_1, \xi_2, \xi_{k-1}$ are given.
not be less than $n$ and the inequality $|\pi(A) - p| \geq \varepsilon$ will also be satisfied, will be less than $2 e^{-2n\varepsilon^2(1-\varepsilon)}$.

If

$$\rho \leq \frac{1}{2} e^{2n\varepsilon^2(1-\varepsilon)}$$

then the sum of the probabilities of failure of the inequality

$$|\pi(A) - p| \leq \varepsilon$$

for those algorithms which lead to the sets with not less than $n$ elements will be less than unity. Hence with positive probability the table $T$ will be found to be $(n, \varepsilon, p)$-random in the sense of the definition of Section 2. Hence follows the existence of tables which are $(n, \varepsilon, p)$-random with respect to $\mathcal{R}_N$ (indeed independently of the probabilistic assumptions on the distribution of $P(T)$ in the space of tables).

4. On the possibilities of improving the estimate by the theorem of Section 2

If we fix $n, \varepsilon, N, p$, then, for an integral non-negative $\rho$ one of the two situations is possible:

(a) whatever be the system $\mathcal{R}_N$ of $\rho$ admissible algorithms of selection, there exists a table $T$ of size $N$ which is $(n, \varepsilon, p)$-random with respect to $\mathcal{R}_N$;

(b) there exists a system $\mathcal{R}_N$ of $\rho$ admissible algorithms of selection relative to which there are no $(n, \varepsilon, p)$-random tables $T$ of size $N$.

We can easily find that the existence of the situation (a) for some $\rho$ follows from the existence of the same situation for $\rho' < \rho$. It is clear that for $\rho = 0$ the situation (a) will always be true. Hence, there exists an upper bound

$$\tau(n, \varepsilon, N, p) = \sup_{\rho \leq \rho'} \rho$$

of those $\rho$ for which the case (a) holds. For all $\rho$ greater than $\tau(n, \varepsilon, N, p)$ the case (b) holds.

If we put

$$\tau(n, \varepsilon) = \inf_{p, N} \tau(n, \varepsilon, N, p)$$

then the substance of the theorem of Section 2 can be expressed in the form of the inequality:

$$\tau(n, \varepsilon) \geq \frac{1}{2} e^{2n\varepsilon^2(1-\varepsilon)}.$$  \hfill (4.1)

Now taking logarithms

$$l(n, \varepsilon, N, p) = \log_2 \tau(n, \varepsilon, N, p), \quad l(n, \varepsilon) = \log_2 \tau(n, \varepsilon),$$
we can write (2.6) in the form

$$I(n, \varepsilon) \geq 2n\varepsilon^2 (1 - \varepsilon) - 1.$$  \hfill (4.2)

In fact, the main interest lies in the asymptotically precise estimation of $I(n, \varepsilon)$ when $\varepsilon$ is small and $n$ and $I(n, \varepsilon)$ are large. When

$$\varepsilon > 0, \ n\varepsilon^2 > \infty$$

we get from (4.2)

$$I(n, \varepsilon) \geq 2n\varepsilon^2 + o(n\varepsilon^2).$$  \hfill (4.3)

We shall find later, on the other hand, that when

$$\varepsilon \to 0, \ n\varepsilon \to \infty$$

the relation

$$I(n, \varepsilon) \leq 4n\varepsilon + o(n\varepsilon)$$  \hfill (4.4)

will hold. Unfortunately, I cannot remove the discrepancy between the power of $\varepsilon$ in (4.3) and (4.4).

The estimate (4.4) is a simple consequence of the following theorem the formulation of which is unfortunately somewhat complex and will become clear through the method of proof chosen by us.

**Theorem 2.** If $k \leq (1 - 2\varepsilon)/4\varepsilon$, $n \leq (k - 1)m$, $N \geq km$ then

$$\tau \left( n, \varepsilon, N, \frac{1}{2} \right) \leq k \cdot 2^m. \hfill (4.5)$$

For proving the theorem it is enough to construct, under the condition

$$k \leq \frac{1 - 2\varepsilon}{4\varepsilon}, \ n = (k - 1)m, \ N = km,$$

a system $\mathcal{R}_N$ out of

$$\rho = k \cdot 2^m + 1$$

admissible algorithms, for which there does not exist an $(n, \varepsilon, \frac{1}{2})$-random table $T$.

We partition $\overline{1, N}$ into $k$ sets $\Delta_i$, $i = 1, \ldots, k$, with $m$ elements in each. Every $\Delta_i$ contains $2^m$ subsets. We form the set

$$\Delta_{is}, \ i = 1, 2, \ldots, k; \ s = 1, 2, \ldots, 2^m$$

by taking the union of all $\Delta_i$, $j \neq i$ and the $\rho$-th subset of $\Delta_i$. We form the system $\mathcal{R}_N$ from

(a) $k \cdot 2^m$ algorithms $R_{is}$ for selecting the sets $A_{is}$;
(b) one algorithm $R$ for selecting $A = 1, N$.

We prove that there does not exist a table $T$ which is $(n, \varepsilon, \frac{1}{2})$-random with respect to $\mathcal{R}_N$.

Let us take an arbitrary table $T$ and assume that it is $(n, \varepsilon, \frac{1}{2})$-random with respect to $\mathcal{R}_N$. Then it must contain at least $\left(\frac{1}{2} - \varepsilon\right)N$ zeroes and $\left(\frac{1}{2} - \varepsilon\right)N$ ones. Hence we can find $i$ and $j$ such that $A_i$ contains $\alpha \geq \left(\frac{1}{2} - \varepsilon\right)m$ zeroes and $A_j$ contains $\beta \geq \left(\frac{1}{2} - \varepsilon\right)m$ ones.

Let

\[ \gamma = \min(\alpha, \beta) \geq \left(\frac{1}{2} - \varepsilon\right)m. \]

There exists an algorithm $R' \in R_N$ ($R'' \in R_N$) for selecting the set $A'$ ($A''$) which consists of the entire $1, N$ except $\gamma$ elements in $A_i$ ($A_j$) which correspond to zero (one) in the table $T$. It is easy to see that the corresponding frequencies are equal to

\[ \pi(A') = \frac{M}{N - \gamma}, \quad \pi(A'') = \frac{M - \gamma}{N - \gamma}, \]

where $M$ is the total number of ones in the table $T$. Let us estimate the difference between these frequencies:

\[ \lambda(A') - \lambda(A'') = \frac{\gamma}{N - \gamma} \geq \frac{\left(\frac{1}{2} - \varepsilon\right)m}{\gamma} \geq 2\varepsilon. \]

This estimate contradicts the set of inequalities

\[ \left| \pi(A') - \frac{1}{2} \right| \leq \varepsilon, \quad \left| \pi(A'') - \frac{1}{2} \right| \leq \varepsilon, \]

which follow from the hypothesis of $(n, \varepsilon, \frac{1}{2})$-randomness of the table $T$. This contradiction proves the theorem.

References