Little Mathematics Library

I. M. SOBOL

THE MONTE CARLO METHOD

Mir Publishers · Moscow


ИЗДАТЕЛЬСТВО
МИР

ПОПУЛЯРНЫЕ ЛЕКЦИИ ПО МАТЕМАТИКЕ

И.М. Соболь

МЕТОД МОНТЕ-КАРЛО

Издательство «Наука»

LITTLE MATHEMATICS LIBRARY

I. M. Sobol

THE MONTE CARLO METHOD

Translated from the Russian
by
V.I. Kisin, Cand. Sc. (Math.)

MIR PUBLISHERS
MOSCOW

First Published 1975
Revised from the 1972 Russian edition

This translation has been read and approved
by the author I. M. Sobol, D. Sc. (Math.)

На английском языке

CONTENTS

Preface 7

Introduction 9
Sec. 1. General 9

Chapter I. Simulation of Random Variables 14
Sec. 2. Random Variables 14
Sec. 3. Generation of Random Variables by Electronic Computers 25
Sec. 4. Transformations of Random Variables 30

Chapter II. Examples of Application of the Monte Carlo Method 39
Sec. 5. Simulation of a Mass Servicing System 39
Sec. 6. Computation of Quality and Reliability of Complex Devices 44
Sec. 7. Computation of Neutron Transmission Through a Plate 51
Sec. 8. Calculation of a Definite Integral 58

Appendix 64
Sec. 9. Proofs of Selected Statements 64
Sec. 10. On Pseudo-Random Numbers 70

Random Number Table 74

Bibliography 77

PREFACE

Several years ago it was suggested that I should deliver two lectures on the Monte Carlo method to the students of the Computer Technology Faculty of the Public University. I agreed. On the eve of the first lecture I found out to my dismay that most of the audience were not familiar with the probability theory. It was too late to retreat. I had no other choice but to present in the course of the lecture a complementary section which surveyed the fundamentals of the probability theory.

These lectures were repeated several years running. Gradually their content has "set". The complimentary paragraph (Sec. 2) is retained in this edition, and I feel that it merits a few words on my part.

Everybody had at some moment used the words "probability" and "random variable". The intuitive idea of the probability (considered as frequency) corresponds more or less to the true meaning of this concept. But as a rule the intuitive idea of the random variable differs quite considerably from the mathematical definition. Thus, the notion of the probability is assumed known in Sec. 2, and only the more complicated notion of the random variable is clarified. This section cannot replace a course in the probability theory: the presentation is simplified and proofs are omitted. But it still presents certain concept of the random variables sufficient for understanding of Monte Carlo techniques.

The basic aim of this book is to prompt the specialists in various branches of knowledge to the fact that there are problems in their fields that can be solved by the Monte Carlo method.

The problems considered in these lectures are sufficiently simple and diverse. Of course, they could not encompass all the problems this method may be applied to. One example will be sufficient to illustrate this statement. Not a word

in this book is devoted to medicine. But the methods described in Sec. 7 make it possible to calculate irradiation doses in radio-therapy. If we have a program for computing radiation absorbed in various body tissues we can select both the dose and direction of irradiation most efficiently and with no damage done to normal tissues.

Everything that was presented in the lectures is included into this text, the examples being given in some more detail. An additional section (Sec. 9) is provided.

I. Sobol

Moscow, 1967

The second edition is practically identical to the first. Only negligible editors' corrections were introduced.

I. Sobol

Moscow, 1971

Introduction

Sec. 1. General

The Monte Carlo method is a numerical method of solving mathematical problems by means of random sampling.

1.1. The Origin of the Monte Carlo Method. The method is conventionally assumed to be born in 1949 when the paper "The Monte Carlo method" was published*. American mathematicians John von Neumann and Stanislav Ulam are thought to have developed this method. First papers on the Monte Carlo method in the USSR were published in 1955-1956**.

Incidentally, the theoretical foundation of the method has been known already for a long time. Furthermore, certain problems in statistics were sometimes calculated by means of random sampling, i.e. in fact by means of the Monte Carlo method. Until the advent of electronic computers, however, this method could not be used on any significant scale because manual simulation of random variables is a very labour-consuming procedure. Thus, the Monte Carlo method as an extremely versatile numerical technique became feasible only owing to the computers.

As for the term "Monte Carlo", it is due to the name of the city of Monte Carlo in the Principality of Monaco famous for its casino. The point is that the roulette is one of the simplest mechanical devices for ... generation of random numbers. We will discuss it in Sec. 3. But it appears worthwhile to answer here one frequently posed question: "Does the Monte Carlo method help to win in roulette?" The answer is "no". It is not even concerned with it.

---

* Metropolis N., Ulam S. The Monte Carlo method, *J. Amer. statistical assoc.*, 1949, 44, No. 247, 335-341.

** Those were the papers by V. V. Chavchanidze, Yu.A. Shreider and V.S. Vladimirov.

1.2. Example. To help the reader get a clear idea of what it is all about we will take a fairly simple example. Let us assume that we have to calculate the surface area of a plane figure S. It may be a completely arbitrary figure with curvilinear boundary whether it be connected or consisting

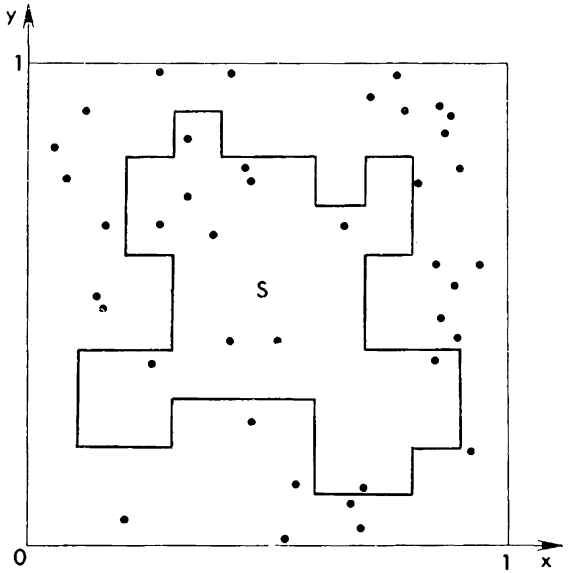

FIG. 1

of several sections, and specified graphically or analytically. Let it be a figure drawn in Fig. 1 and let us assume that it is enclosed within a unit square.

Let us select N random points inside the square. Let N' designate the number of points that happened to fall within S. It is geometrically obvious that the area S is appoximately equal to the ratio N'/N. The greater N the higher the accuracy of the estimate.

The number of points selected in Fig. 1 is $N = 40$. Of these points $N' = 12$ are inside S. The ratio $N'/N = 12/40 = 0.30$ while the true area S is 0.35*.

---

* In practice the Monte Carlo method is not applied to calculate the area of the plane figure because other methods are available which, though more complicated, provide much higher accuracy.

However, the Monte Carlo method, as illustrated in the example

1.3. Two Distinctive Features of the Monte Carlo Method. The first feature is the simple structure of the computation algorithm. As a rule, a program is prepared to perform only one random trial (in Example 1.2 we have to select a random point within the limits of the square and check whether it lies within S). This trial is repeated N times, each trial being independent of all others, and the results of all trials are averaged.

Thus, the Monte Carlo method is sometimes called the *method of statistical trials.*

The second feature of the method is as follows: the error of calculations is, as a rule, proportional to $\sqrt{D/N}$ where D is a constant and N is the number of trials. This formula shows that 100-fold increase of N (and thus of the work) is necessary to reduce the error by a factor of 10 (in other words, to obtain one more accurate decimal digit in the result).

It is clear that high accuracy cannot be attained with such an approach. Thus it is usually said that the Monte Carlo method is especially efficient in solving those problems which do not require a high degree of accuracy (e.g. 5 to 10 per cent).

However, a given problem can be solved through various versions of the Monte Carlo method*, each having a different value of D. Accuracy can be considerably improved for many problems by a proper choice of the computation technique with an appreciably smaller value of D.

1.4. Problems Solved by the Monte Carlo Method. First of all, the Monte Carlo method allows to simulate any process whose development is affected by random factors. Secondly, many mathematical problems not affected by any random influences can be connected with an artificially constructed probabilistic model (even more than one) making possible the solution of these problems. This in fact has been done in Example 1.2.

---

above, permits to calculate just as simply the "multidimensional volume" of a body in a multidimensional space. In this case the Monte Carlo method is often the only numerical method enabling us to solve the problem.

* It becomes more widespread nowadays in publications abroad to speak of Monte Carlo methods (in plural), meaning that one and the same problem can be computed by means of simulating various random variables.

Thus we can consider the Monte Carlo technique as a universal method of solving mathematical problems.

Of particular interest is the fact that in certain cases it is expedient to use an artificial model instead of simulating a real random process. Such a problem is discussed in Sec. 7.

1.5. More about the Example. Let us go back to Example 1.2. The computation required the selection of random points within the unit square. How should we actually realize it?

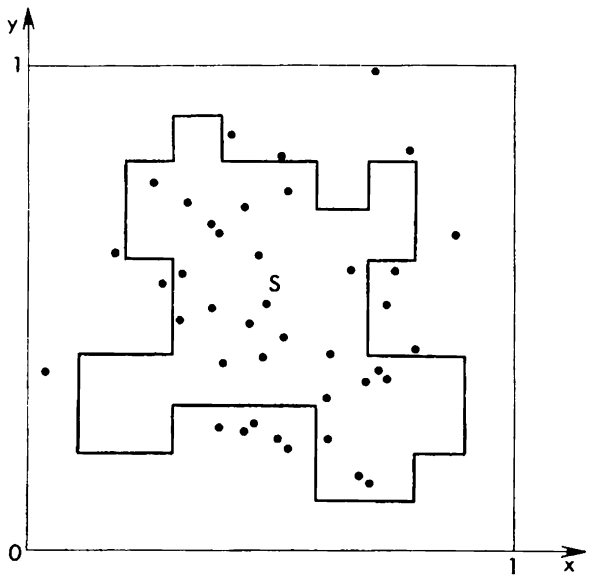

FIG. 2

Let us imagine the following experiment. Fig. 1 scaled up with the figure S and the square is fixed on the wall as a target. A marksman, placed some distance from the wall, shoots N times aiming at the centre of the square. Of course, the bullets will not all strike the centre of the target: they will pierce the target in N random points*. Can we estimate area S from these points?

* We assume that the marksman is not the world champion and is removed at a sufficiently long distance from the target.

The result of such an experiment is shown in Fig. 2. In this experiment $N = 40$, $N' = 24$ and the ratio $N'/N = 0.60$ which is almost twice the true value of the area (0.35). It is quite obvious however that very high qualification of the marksman will lead to a very bad result of the experiment, because almost all punctures will lie close to the centre, i.e. inside S*.

It is easily seen that the above method of calculating the area will be valid only if the random points are not "just random" but also "uniformly scattered" over the whole square. In order to attach a precise meaning to these terms we have to introduce the definition of random variables and to cite some of their properties. This information is given without proof in Sec. 2. Readers familiar with any course in the probability theory can skip the whole of Sec. 2 with the exception of Item 2.5.

* The manner in which random points were selected in Figs. 1 and 2 will be elucidated in Sec. 4.5.

CHAPTER I

Simulation of Random Variables

Sec. 2. Random Variables

We shall assume that the reader is more or less familiar with the concept of *probability* and shall turn directly to the concept of *random variable.*

The term "random variable", in its commonplace sense is used to emphasize that one does not know what specific value this variable will assume. Sometimes these words merely hide the fact that one simply does not know the value of this variable.

Mathematicians, on the contrary, use the same expression "random variable" in quite a definite sense. They say that though we indeed do not know what value will this variable take on in a given case, we do know what values it can assume and what are the respective probabilities. The result of a single trial representing this random variable cannot be accurately predicted from these data but we can predict very reliably the result of a large number of trials. The greater the number of trials (or, as it is normally expressed, the larger the statistics) the more accurate will our predictions be.

Thus, to present a random variable we must indicate what values it can assume and what are the probabilities of these values.

2.1. Discrete Random Variables. If a random variable ξ takes on values from a discrete set $x_1, x_2, \ldots, x_n$ it is called a *discrete* random variable.*

The discrete random variable ξ is specified by the table

$$\xi = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ p_1 & p_2 & \ldots & p_n \end{pmatrix}, \qquad \text{(T)}$$

* The probability theory also considers discrete random variables which can assume infinite number of values.

where $x_1, x_2, \ldots, x_n$ are the possible values of the variable ξ and $p_1, p_2, \ldots, p_n$ are the corresponding probabilities. To be precise, the probability that the random variable ξ will be equal to x_i (we will designate it as $\mathbf{P}\{\xi = x_i\}$) is equal to p_i:

$$\mathbf{P}\{\xi = x_i\} = p_i.$$

The table (T) is called the *distribution of the random variable.*

Generally speaking, the values $x_1, x_2, \ldots, x_n$ can be arbitrary. But the probabilities $p_1, p_2, \ldots, p_n$ must satisfy two conditions:

(a) all p_i are positive:

$$p_i > 0; \tag{1}$$

(b) the sum of all p_i is equal to 1:

$$p_1 + p_2 + \ldots + p_n = 1. \tag{2}$$

The last condition means that in each trial ξ necessarily takes on one of the values $x_1, x_2, \ldots, x_n$.

The quantity

$$\mathbf{M}\xi = \sum_{i=1}^{n} x_i p_i \tag{3}$$

is called the *mathematical expectation*, or the *expected value* of the random variable ξ.

To elucidate the physical meaning of this value we will rewrite it in the following form:

$$\mathbf{M}\xi = \frac{\sum_{i=1}^{n} x_i p_i}{\sum_{i=1}^{n} p_i}.$$

As follows from this relation, $\mathbf{M}\xi$ is the *mean value* of the variable ξ and the values of x_i associated with greater probabilities are included into the sum with larger weights*.

---

* Weight-averaging is encountered very often and in most different spheres of science. In mechanics, for example, if masses $m_1, m_2, \ldots, m_n$ are located at the points $x_1, x_2, \ldots, x_n$ (on the Ox-axis), then the abscissa of the centre of gravity of this system is calculated by means

Let us discuss the basic properties of the expected value. If c is an arbitrary non-random number, then

$$\mathbf{M}(\xi + c) = \mathbf{M}\xi + c, \tag{4}$$

$$\mathbf{M}(c\xi) = c\mathbf{M}\xi; \tag{5}$$

if ξ and η are two arbitrary random variables, then

$$\mathbf{M}(\xi + \eta) = \mathbf{M}\xi + \mathbf{M}\eta. \tag{6}$$

The quantity

$$\mathbf{D}\xi = \mathbf{M}[(\xi - \mathbf{M}\xi)^2] \tag{7}$$

is called the *variance* of the random variable ξ. Thus the variance $\mathbf{D}\xi$ is the expected value of the square of the deviation of the random variable ξ from its mean value $\mathbf{M}\xi$. It is obvious that in all cases $\mathbf{D}\xi > 0$.

Expected value and variance are the two most important numerical characteristics of the random variable ξ. What is their practical significance?

If in the course of many observations of the variable ξ we obtain a set of values $\xi_1, \xi_2, \ldots, \xi_N$ (each of these will be equal to one of the numbers $x_1, x_2, \ldots, x_n$), then the arithmetic mean of these numbers will be close to $\mathbf{M}\xi$

$$\frac{1}{N}(\xi_1 + \xi_2 + \ldots + \xi_N) \approx \mathbf{M}\xi. \tag{8}$$

The variance $\mathbf{D}\xi$ characterizes the spreading of these quantities around the mean value $\mathbf{M}\xi$.

Formula (7) for the variance can be transformed by means of Eqs. (4)-(6):

$$\mathbf{D}\xi = \mathbf{M}[\xi^2 - 2\mathbf{M}\xi \cdot \xi + (\mathbf{M}\xi)^2] =$$
$$= \mathbf{M}(\xi^2) - 2\mathbf{M}\xi \cdot \mathbf{M}\xi + (\mathbf{M}\xi)^2,$$

whence

$$\mathbf{D}\xi = \mathbf{M}(\xi^2) - (\mathbf{M}\xi)^2. \tag{9}$$

---

of the following formula:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i m_i}{\sum_{i=1}^{n} m_i}$$

Obviously, in this case the sum of masses should not necessarily equal unity.

Usually the calculation of variance via Eq. (9) is simpler than by Eq. (7).

Let us mention the basic properties of variance: if c is an arbitrary non-random number, then

$$\mathbf{D}(\xi + c) = \mathbf{D}\xi, \tag{10}$$

$$\mathbf{D}(c\xi) = c^2\mathbf{D}\xi. \tag{11}$$

The concept of *independence* of random variables plays an important role in the probability theory. This is in fact a rather complicated concept, though it may be quite clear in the simplest cases*.

The following relations are valid for independent random variables ξ and η

$$\mathbf{M}(\xi\eta) = \mathbf{M}\xi \cdot \mathbf{M}\eta, \tag{12}$$

$$\mathbf{D}(\xi + \eta) = \mathbf{D}\xi + \mathbf{D}\eta. \tag{13}$$

Example. Let us consider a random variable $\varkappa$ with the distribution specified by the table

$$\varkappa = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}.$$

It is obvious that the number of points obtained in one throw of a die can be considered a realization of this variable: any of the values is equally probable. Let us calculate the expected value and the variance of $\varkappa$. According to (3)

$$\mathbf{M}\varkappa = 1 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} = 3.5.$$

According to (9)

$$\mathbf{D}\varkappa = \mathbf{M}(\varkappa^2) - (\mathbf{M}\varkappa)^2 = 1^2 \cdot \frac{1}{6} + \ldots + 6^2 \cdot \frac{1}{6} - (3.5)^2 = 2.917.$$

Example. Let us consider a random variable θ with the distribution

$$\theta = \begin{pmatrix} 3 & 4 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

The game of tossing a coin, or "heads or tails", with the condition that a toss of heads brings 3 points and that of tails, 4 points, can be

---

* Let us assume simultaneous observation of random variables ξ and η. If the distribution of the variable ξ does not change after we have found the value assumed by the variable η, then it is natural to consider ξ independent of η.

considered a realization of this variable. In this case

$$\mathbf{M}(\theta) = 0.5 \cdot 3 + 0.5 \cdot 4 = 3.5;$$

$$\mathbf{D}(\theta) = 0.5\,(3^2 + 4^2) - (3.5)^2 = 0.25.$$

We see that $\mathbf{M}\theta = \mathbf{M}\varkappa$, but $\mathbf{D}\theta < \mathbf{D}\varkappa$. The inequality could be easily predicted because the maximum value of deviation of 0 from 3.5 is ± 0.5 while the spreading of $\varkappa$ can reach ± 2.5.

2.2. Continuous Random Variables. Let us assume that a certain amount of radium is placed on a plane at the origin of coordinates. An α-particle is emitted in each decay of a radium atom. The direction of motion of this particle

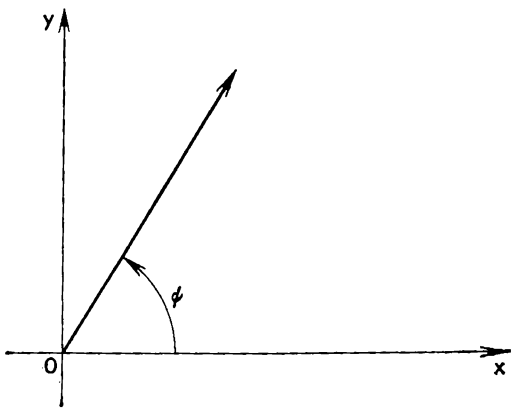

FIG. 3

will be characterized by an angle ψ (Fig. 3). Since theoretically and practically any direction of emission is possible, this random variable can assume any value from 0 to 2π.

We shall call a random variable ξ *continuous* if it takes on any value out of a certain interval (a, b).

A continuous random variable ξ is defined by specifying the interval (a, b) of its variation, and the function $p(x)$, called the *probability density* of the random variable ξ (or the *distribution density* of ξ).

The physical meaning of $p(x)$ is the following: let (a', b') be an arbitrary interval contained within (a, b) (i.e. $a \leqslant a'$, $b' \leqslant b$). Then the probability that ξ falls inside

(a', b') is given by the integral

$$\mathbf{P}\{a' < \xi < b'\} = \int_{a'}^{b'} p(x)\,dx. \tag{14}$$

This integral (14) is equal to the hatched area in Fig. 4.

The set of values of ξ can be any interval. The cases when $a = -\infty$ as well as $b = \infty$ are also possible. However the density $p(x)$ must satisfy two conditions similar to conditions (1) and (2) for discrete variables:

(a) the density $p(x)$ is positive:

$$p(x) > 0; \tag{15}$$

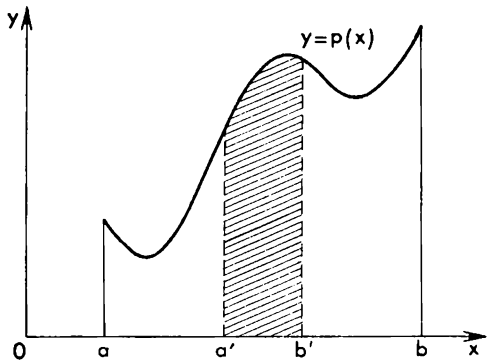

FIG. 4

(b) the integral of the density $p(x)$ over the whole interval (a, b) is equal to 1:

$$\int_a^b p(x)\,dx = 1. \tag{16}$$

The value

$$\mathbf{M}\xi = \int_a^b xp(x)\,dx \tag{17}$$

is called the *expected value* of the continuous random variable.

This characteristic has the same meaning as in the case of a discrete random variable. In fact, it directly follows

from

$$\mathbf{M}\xi = \frac{\int_a^b xp(x)\,dx}{\int_a^b p(x)\,dx},$$

that this is the mean value of ξ: indeed, any value of x from the interval (a, b), entering the integral with the weight $p(x)$ can be the value of ξ*.

All the results of Sec. 2.1, starting with Eq. (4) and including Eq. (13) are valid for continuous random variables. The definition of variance (7) and the computation formula (9) as well as all the statements with respect to properties of $\mathbf{M}\xi$ and $\mathbf{D}\xi$ are also valid. We shall not repeat them.

We will mention only one more expression for the expected value of a function of ξ. Let us assume again that a random variable ξ is characterized by the probability density $p(x)$. Let us take an arbitrary continuous function $f(x)$ and introduce a random variable $\eta = f(\xi)$. It can be shown that

$$\mathbf{M}f(\xi) = \int_a^b f(x)\,p(x)\,dx. \tag{18}$$

We must emphasize that in general $\mathbf{M}f(\xi) \neq f(\mathbf{M}\xi)$.

A random variable γ, defined on the interval (0, 1) with the density $p(x) = 1$ is said to be *uniformly distributed* in (0, 1) (Fig. 5).

Indeed, for any interval (a', b') within (0, 1) the probability that γ will take on the value within (a', b') is equal to

$$\int_{a'}^{b'} p(x)\,dx = b' - a',$$

---

* In this case it is equally possible to indicate a similar formula from mechanics: if the linear density of a rod $a \leqslant x \leqslant b$ is $\rho(x)$, then the abscissa of the rod gravity centre is calculated by means of the expression

$$\bar{x} = \frac{\int_a^b x\rho(x)\,dx}{\int_a^b \rho(x)\,dx}.$$

i.e. to the length of this interval. If, in particular, we divide the (0, 1) interval into an arbitrary number of intervals of equal length, the probability of γ falling within any one of these intervals will be the same.

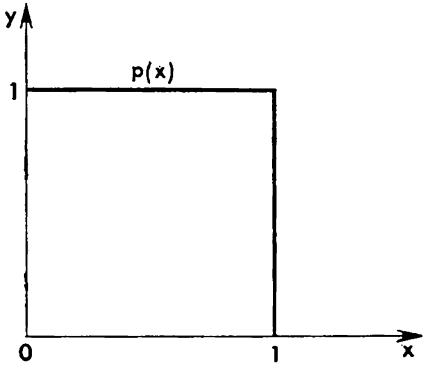

FIG. 5

It can be easily shown that

$$\mathbf{M}\gamma = \int_0^1 xp(x)\,dx = \int_0^1 x\,dx = \frac{1}{2};$$

$$\mathbf{D}\gamma = \int_0^1 x^2 p(x)\,dx - (\mathbf{M}\gamma)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

This random variable γ will be more than once encountered in further discussion.

2.3. Normal Random Variables. A random variable ζ, defined on the entire axis $(-\infty, \infty)$ and characterized by the density

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, \qquad (19)$$

where a and $\sigma > 0$ are numerical parameters*, is said to be *normal* (or *Gaussian*) random variable.

The parameter a does not affect the shape of the $p(x)$ curve; its variation only shifts the curve as a whole along

* σ is a number and not a random variable. The Greek letter is used here as the generally accepted one.

the x-axis. On the contrary, variation of σ changes the shape of the curve. It is indeed easily seen that

$$\max p(x) = p(a) = \frac{1}{\sigma\sqrt{2\pi}}.$$

A decrease of σ is accompanied by an increase of $\max p(x)$. But according to the condition (16) the entire area below the $p(x)$ curve is equal to unity. Therefore the curve will stretch upward in the vicinity of $x = a$ but will decrease

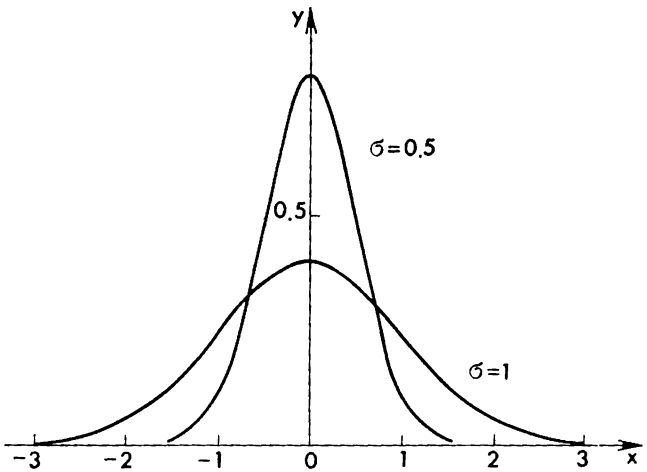

FIG. 6

at all sufficiently large values of x. Two normal densities corresponding to $a = 0$, $\sigma = 1$ and $a = 0$, $\sigma = 0.5$ are plotted in Fig. 6. (Another normal density is plotted in Fig. 21, see below, p. 49.)

It can be proved that

$$\mathbf{M}\zeta = a, \quad \mathbf{D}\zeta = \sigma^2.$$

Normal random variables are frequently encountered in investigation of most diverse problems. The reason will be discussed somewhat later. For example, the experimental error δ is usually a normal random variable. If there is no systematic error (bias) then $a = \mathbf{M}\delta = 0$. The value $\sigma = \sqrt{\mathbf{D}\delta}$, called the *standard deviation*, characterizes the error of the method of measurement.

The "three sigma" rule. It is easily calculated that regardless of the values of a and σ in (19)

$$\int_{a-3\sigma}^{a+3\sigma} p(x)dx = 0.997.$$

As follows from (14),

$$\mathbf{P}\{a - 3\sigma < \zeta < a + 3\sigma\} = 0.997. \tag{20}$$

The probability 0.997 is so close to 1 that sometimes the last formula is given the following interpretation: *it is practically impossible to obtain in a single trial a value of* ζ *deviating from* $\mathbf{M}\zeta$ *by more than* 3σ.

2.4. Central Limit Theorem of the Probability Theory. The statement of this spectacular theorem has first been given by Laplace. A number of outstanding mathematicians, P. L. Chebyshev, A. A. Markov and A. M. Lyapunov among them, investigated the problems of generalizing this theorem. Its proof is rather complicated.

Let us consider N identical independent random variables $\xi_1, \xi_2, \ldots, \xi_N$ such that their probability distributions coincide. Consequently, both their expected values and variances coincide as well.

Let us designate

$$\mathbf{M}\xi_1 = \mathbf{M}\xi_2 = \ldots = \mathbf{M}\xi_N = m,$$
$$\mathbf{D}\xi_1 = \mathbf{D}\xi_2 = \ldots = \mathbf{D}\xi_N = b^2,$$

and designate the sum of all these values as ρ_N

$$\rho_N = \xi_1 + \xi_2 + \ldots + \xi_N.$$

It follows from Eqs. (6) and (13) that

$$\mathbf{M}\rho_N = \mathbf{M}(\xi_1 + \xi_2 + \ldots + \xi_N) = Nm,$$
$$\mathbf{D}\rho_N = \mathbf{D}(\xi_1 + \xi_2 + \ldots + \xi_N) = Nb^2.$$

Let us consider now a normal random variable ζ_N with the same parameters: $a = Nm$, $\sigma^2 = Nb^2$.

The central limit theorem states that *for any interval* (a', b') *for large* N

$$\mathbf{P}\{a' < \rho_N < b'\} \approx \int_{a'}^{b'} p_{\zeta_N}(x)\,dx.$$

The physical meaning of this theorem is obvious: the sum ρ_N of a large number of identical random variables is approximately normal ($p_{\rho_N}(x) \approx p_{\xi_N}(x)$).

Actually this theorem is valid for much more general conditions: all the addends $\xi_1, \xi_2, \ldots, \xi_N$ are not required to be identical and independent, it is only essential that none of them play too great a role in this sum.

It is this theorem that explains why normal random variables are met so often in nature. And indeed, each time we come across an aggregate effect of a large number of negligible random factors we find that the resulting random variable is normal.

Deviation of an artillery shell off the target, for example, is almost always a normal random variable because it depends on meteorological conditions in various sections of the trajectory as well as on many other factors.

2.5. The General Scheme of the Monte Carlo Method. Let us assume that we need to calculate some unknown value m. We shall try to find a random variable ξ such that $\mathbf{M}\xi = m$. Let us assume therewith that $\mathbf{D}\xi = b^2$.

Let us consider N random independent variables $\xi_1, \xi_2, \ldots, \xi_N$ with distributions identical to that of ξ. If N is sufficiently large then it follows from 2.4 that the distribution of the sum $\rho_N = \xi_1 + \xi_2 + \ldots + \xi_N$ will be approximately normal and will have the parameters $a = Nm$, $\sigma^2 = Nb^2$. Formula (20) yields

$$\mathbf{P}\{Nm - 3b\sqrt{N} < \rho_N < Nm + 3b\sqrt{N}\} \approx 0.997.$$

If we divide the inequality in braces by N we will arrive at an equivalent inequality and the probability will not change:

$$\mathbf{P}\left\{m - \frac{3b}{\sqrt{N}} < \frac{\rho_N}{N} < m + \frac{3b}{\sqrt{N}}\right\} \approx 0.997.$$

We will rewrite the last expression in a somewhat different form:

$$\mathbf{P}\left\{\left|\frac{1}{N}\sum_{j=1}^{N}\xi_j - m\right| < \frac{3b}{\sqrt{N}}\right\} \approx 0.997. \qquad (21)$$

This formula is very important to the Monte Carlo method. It gives us both the method of calculating m and the estimate of the error.

Let us indeed sample N values of the random variable ξ^*. The formula (21) indicates that the arithmetic mean of these values will be approximately equal to m. The error of this approximation will most probably not exceed the value $3b/\sqrt{N}$. This error evidently approaches zero as N increases.

Sec. 3. Generation of Random Variables by Electronic Computers

Sometimes the statement "generation of random variables" by computers is the cause of bewilderment: everything performed by a computer must be programmed in advance, so where does the chance come in?

This statement is indeed fraught with certain difficulties but these mostly concern philosophy and we shall not consider them here.

We will only note that random variables dealt with in Sec. 2 represent an ideal mathematical concept. Whether some natural phenomenon can be actually described by means of these variables can only be found out experimentally. Such a description is always approximate. Moreover, a random variable which describes quite satisfactorily a certain physical quantity in one type of phenomena may prove quite an unsatisfactory characteristic of the same quantity when other phenomena are being investigated.

The situation is similar to that of a road which can be considered straight on the map of the country (ideal mathematical straight line with "zero width"), but appears as a twisted band on a large-scale map of a settlement...

The methods of generating random variables are usually divided into three types: tables of random numbers, random number generators and the method of pseudo-random numbers.

3.1. Random Number Tables. Let us run the following experiment. Let us write down the digits 0, 1, 2, . . ., 9 on ten identical sheets of paper, put them into a hat, mix them and then extract one sheet at a time, each time returning it to the hat and again mixing the sheets. We shall list

---

* Determining a single value of each of the variables $\xi_1, \xi_2, \ldots, \xi_N$ is equivalent to determining N values of a single variable ξ, because all these random variables are identical (have identical distributions).

the numbers obtained in this way in a table similar to Table A at the end of this book on page 74. (For the sake of convenience the numbers in Table A are arranged in groups of 5 digits.)

Such a table is called the *Random Number Table*, though "The Table of Random Digits" would be a better title. This

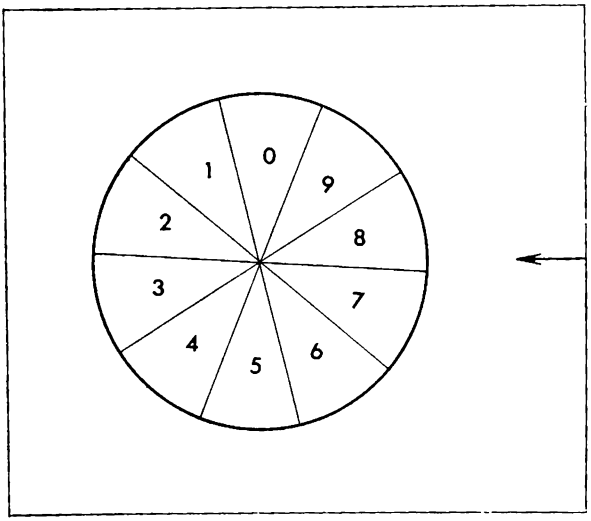

FIG. 7

table can be introduced into the memory of a computer. When later in the course of calculations we shall need values of the random variable with the distribution

$$\begin{pmatrix} 0 & 1 & 2 & \dots & 9 \\ 0.1 & 0.1 & 0.1 & \dots & 0.1 \end{pmatrix}, \tag{22}$$

we shall retrieve consecutive numbers from this table.

The most comprehensive of all published tables of random numbers contains 1 000 000 numbers*. Of course, this table has been compiled by means of a device more advanced than a hat: a special roulette with electric devices was constructed. The simplest design of such a roulette is shown in Fig. 7 (a rotating disk is abruptly stopped and the number indicated by a fixed pointer is recorded).

* RAND Corporation, *A Million Random Digits with 100 000 Normal Deviates*, The Free Press, 1955.

It should be noted that a good table of random numbers is not so easy to compile as one might expect. Any real physical instrument generates random numbers with distribution that to a certain extent differs from the ideal distribution (22). Besides, errors are possible during the experiment (for instance, one of the paper sheets may stick for some time to the lining of the hat). For this reason the compiled tables are thoroughly checked by means of special statistical tests: whether some of the properties of the group of numbers contradict the hypothesis that these numbers are the values of the random variable (22).

Consider one of the simplest tests. Let us take a table containing N numbers, in which ν_0 is the number of zeros, ν_1 is the number of unities, ν_2—that of two's, and so on. Let us calculate the sum

$$\sum_{i=0}^{9} (\nu_i - 0.1N)^2.$$

The probability theory enables us to predict in what range this sum should be found: its value should not be too large (since the expected value of each ν_i is equal to $0.1N$) but neither should it be too small (because it would mean "too regular" distribution of numbers).

Random number tables are used in Monte Carlo calculations only if these are performed manually. The thing is, all computers have a comparatively small internal store, and a large table cannot be stored in it; storage of the table in the external data store and constantly retrieving them considerably slows down the computation.

It is not excluded though that time will bring a sharp increase of memory capacity in computers; this will make the use of random number tables quite practicable.

3.2. Generators of Random Numbers. It might seem that a roulette mentioned in Sec. 3.1 could be coupled to a computer and would generate random numbers as soon as they are required. However, any mechanical device will be too slow for a computer. Therefore, noise signals in electron tubes are most often used as generators of random variables: if the noise level has exceeded a fixed threshold an even number of times during a certain fixed time interval Δt, zero is recorded, and if this number is odd, unity is recorded.*

* There also exist more advanced devices.

At a first glance this appears to be a very convenient method. Let m such generators operate simultaneously and send random zeros and unities into all binary digit places of a special cell. Each cycle produces one m-digit number. At any phase of computations we can refer to this cell and retrieve a value of the random variable γ distributed uniformly in the interval (0, 1). This value is, of course, approximate and is in the form of an m-digit binary fraction $0, \alpha_{(1)}\alpha_{(2)} \ldots \alpha_{(m)}$ in which each α_i simulates a random variable with the distribution

$$\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

But neither this method is free of shortcomings. Firstly, it is difficult to check the "quality" of generated numbers. Checks are necessarily made periodically since the so-called "distribution drift" may appear due to some faults (that is, zeros and unities appear in some bits at nonidentical frequencies). Secondly, all calculations in computers are usually performed twice to exclude the possibility of occasional failure. However, it is impossible to reproduce the same random numbers if they are not stored in the course of computations. But in the case of storing them we again face the situation typical for tables.

Generators of this type will undoubtedly prove useful when specialized computers will be designed for solving problems by means of the Monte Carlo method. But it would not be economical to install and operate a special unit in multipurpose computers in which computations, involving random numbers, are performed only occasionally. It is more expedient to use the so-called *pseudo-random numbers.*

3.3. Pseudo-Random Numbers. Since the "quality" of random numbers used for computations is checked by means of special tests it is of no importance how they were generated if only they satisfy the accepted set of tests. We may even calculate them by means of a prescribed formula. But such a formula must, of course, be a very ingenious one.

The numbers calculated by means of some formula and simulating the values of the random variable γ are called *pseudo-random numbers.* The term "simulating" means that

these numbers satisfy a set of tests as if they represented the values of this random variable.

The first algorithm for generating random variables was suggested by J. Neumann. It is referred to as the *mid-square method.* Let us elucidate it by means of an example.

Suppose we have a four-digit number $\gamma_0 = 0.9876$. Let us raise it to a second power. We get an 8-digit number $\gamma_0^2 = 0.97535376$. Let us select four middle digits of this number and assume $\gamma_1 = 0.5353$.

Now we raise γ_1 to a second power ($\gamma_1^2 = 0.28654609$) and again extract four middle figures. We get $\gamma_2 = 0.6546$. Further on we obtain $\gamma_2^2 = 0.42850116$, $\gamma_3 = 0.8501$; $\gamma_3^2 = 0.72267001$, $\gamma_4 = 0.2670$; $\gamma_4^2 = 0.07128900$, $\gamma_5 = 0.1289$, and so on.*

However, this algorithm did not prove acceptable: the fraction of smaller values is higher than is necessary. Thus other algorithms were developed by a number of investigators. Some of them are oriented at specific features of particular computers. As an example we will consider one of such algorithms operating in the *Strela* computer.

Example.** *Strela* is a three-address computer with a floating decimal point. The cell in which a number x is stored consists of 43 binary digits (Fig. 8). The computer operates on normalized binary

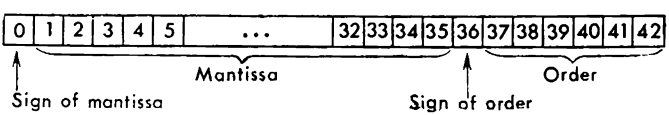

FIG. 8

numbers $x = \pm q \cdot 2^{\pm p}$ where p is the order of the number and q is its mantissa. The j-th digit of the cell may contain a zero or a unity; we designate this value as ε_j. Then

$$q = \frac{\varepsilon_1}{2^1} + \frac{\varepsilon_2}{2^2} + \ldots + \frac{\varepsilon_{35}}{2^{35}},$$

$$p = \varepsilon_{37}2^5 + \varepsilon_{38}2^4 + \ldots + \varepsilon_{42}2^0.$$

---

* This algorithm can be written in the form $\gamma_{k+1} = F(\gamma_k)$ where F designates a set of operations that must be performed over the number γ_k to obtain γ_{k+1}. The number γ_0 is specified.

** See I.M. Sobol, Pseudo-random numbers for the *Strela* computer, *The Probability Theory and Its Applications*, 3, No. 2 (1958), 205-211.

The normalization condition $0.5 \leqslant q < 1$ means that ε_1 is invariably equal to 1. The sign "+" is represented by zero, and the sign "—", by unity.

Three operations are necessary to obtain γ_{k+1} from γ_k:

1. γ_k is multiplied by a large constant, usually 10^{17}.
2. The machine word representing $10^{17}\gamma_k$ is shifted 7 digits to the left (so that the first 7 digits of the product are lost and the digits from 36th to 42th become zeros).
3. The absolute value of the obtained number is formed (and the number is normalized); this gives γ_{k+1}.

If we start with $\gamma_0 = 1$, this process generates more than 80 000 different numbers γ_k; after that a cycle occurs and the numbers begin to repeat themselves. Various tests of the first 50 000 numbers gave quite satisfactory results. These numbers were repeatedly used in solving most diverse problems.

Advantages of the method of pseudo-random numbers are fairly obvious. Firstly, only a few simple operations are necessary to calculate each number so that the rate of generation is of the same order as the speed of the computer itself. Secondly, the program takes only a few cells of the storage. Thirdly, any of the numbers γ_k can easily be reproduced. And finally, only one check of the "quality" of this sequence is required and thereupon this sequence can be used without taking any risk in calculations of similar problems.

The only shortcoming of the method is that the "store" of pseudo-random numbers is limited. However, there are methods that make it possible to generate much greater arrays of numbers. In particular, starting numbers γ_0 can be changed.

At present an overwhelming majority of Monte Carlo calculations are carried out by using pseudo-random numbers. Additional information on random numbers is given in Sec. 10.

Sec. 4. Transformations of Random Variables

Solution of various problems requires simulation of various random variables. At an early stage of application of the Monte Carlo method some specialists tried to design a special roulette for each random variable. For instance, a roulette with the disk divided into unequal portions proportional to p_i shown in Fig. 9 and functioning identically to that of Fig. 7, can be used to generate the values of the random

variable with the distribution

$$\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ 0.5 & 0.25 & 0.125 & 0.125 \end{pmatrix}. \tag{22'}$$

However, this proved to be absolutely unnecessary: the values of any random variable can be obtained by transforming a selected random variable ("standard", so to speak).

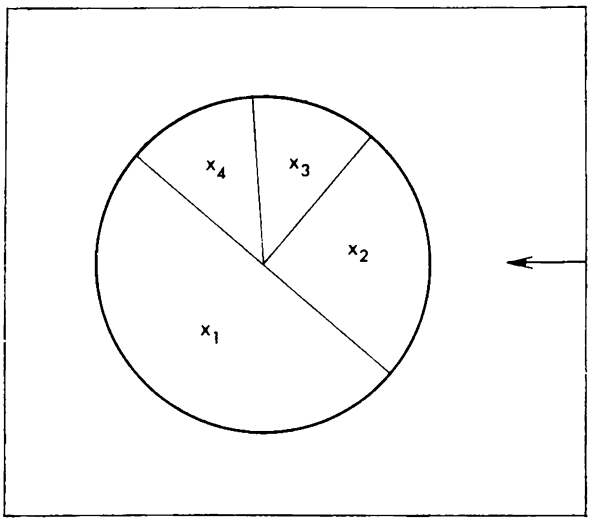

FIG. 9

Usually this role is played by the random variable γ, uniformly distributed in (0, 1). We already know how to obtain values of γ.

Let us call the process of obtaining a value of some random variable ξ by means of transforming one or several values of γ the *drawing* of the random variable ξ.

4.1. Drawing of a Discrete Random Variable. Let us suppose that we want to calculate the values of the random variable ξ with the distribution

$$\xi = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}.$$

Let us consider the interval $0 < y < 1$ and divide it into n intervals with lengths equal to $p_1, p_2, \dots, p_n$.

The coordinates of division points will apparently be $y = p_1$, $y = p_1 + p_2$, $y = p_1 + p_2 + p_3$, . . ., $y = p_1 + p_2 + \ldots + p_{n-1}$. Now we enumerate these intervals by the numbers 0, 1, . . ., n (Fig. 10). This completes the preparation procedure for drawing ξ. Each time we have to "run the experiment" and draw a value of ξ, we shall

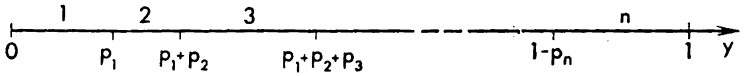

FIG. 10

select a value of γ and fix the point $y = \gamma$. If this point falls into the i-th interval we shall assume that $\xi = x_i$ (in this experiment).

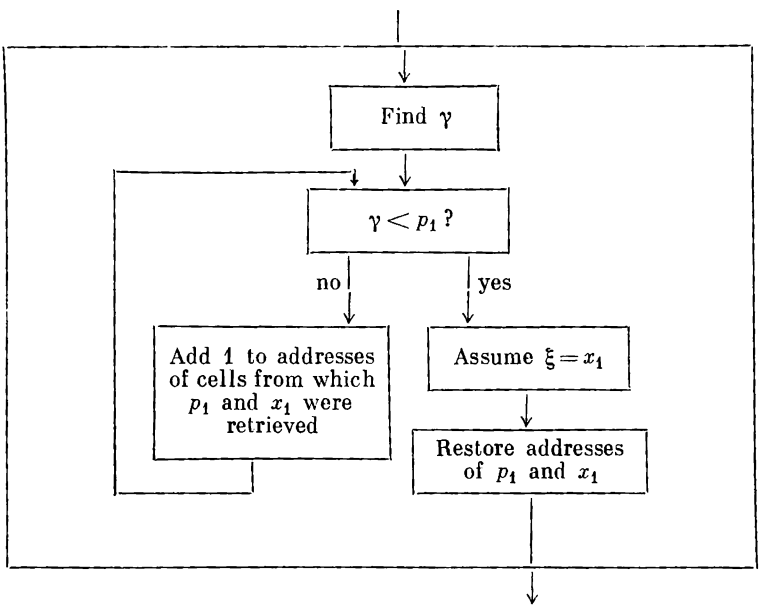

FIG. 11

Validity of this procedure is proved very easily. Indeed, since the random variable γ is distributed uniformly within (0, 1), the probability of γ lying within one of the inter-

vals is equal to the length of this interval. Hence,

$$\mathbf{P}\{0 < \gamma < p_1\} = p_1,$$
$$\mathbf{P}\{p_1 < \gamma < p_1 + p_2\} = p_2,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\mathbf{P}\{p_1 + p_2 + \ldots + p_{n-1} < \gamma < 1\} = p_n.$$

According to our procedure $\xi = x_i$ if

$$p_1 + p_2 + \ldots + p_{i-1} < \gamma < p_1 + p_2 + \ldots + p_i,$$

and the probability of this event is equal to p_i.

Of course, Fig. 10 can be avoided when a computer is used. Let us suppose that the numbers $x_1, x_2, \ldots, x_n$ are placed in succession in storage cells and the probabilities p_1, $p_1 + p_2$, $p_1 + p_2 + p_3, \ldots, 1$ also form a sequence in the data store. The block diagram of the subroutine for drawing ξ is shown in Fig. 11.

Example. To draw 10 values of the random variable with the distribution

$$\theta = \begin{pmatrix} 3 & 4 \\ 0.58 & 0.42 \end{pmatrix}.$$

Let us select as values of γ ten pairs of numbers from the Table A on p. 74, multiplied by 0.01*. Thus, $\gamma = 0.86, 0.51, 0.59, 0.07, 0.95, 0.66, 0.15, 0.56, 0.64, 0.34$.

It is obvious that according to our scheme the value $\theta = 3$ corresponds to the values of γ smaller than 0.58 and $\theta = 4$, to the values of $\gamma \geqslant 0.58$. Hence, we obtain $\theta = 4, 3, 4, 3, 4, 4, 3, 3, 4, 3$.

Let us note that the order of enumerating the numbers $x_1, x_2, \ldots, x_n$ in the distribution of ξ can be arbitrary, but it must be fixed before the drawing.

4.2. Drawing of a Continuous Random Variable. Let us assume now that we need to generate the values of the random variable ξ distributed in the interval (a, b) with the density $p(x)$.

---

* Two decimal digits are sufficient in the values of γ since p_i in this example are specified only with two decimal digits. The case of $\gamma = 0.58$ is possible in this approximation and should be combined with the case of $\gamma > 0.58$ (because $\gamma = 0.00$ is possible while $\gamma = 1.00$ is impossible). When multivalued γ are used, the case of $\gamma = p_1$ is almost improbable and can be attributed to any of the inequalities.

Let us prove that ξ values can be found from the equation

$$\int_a^{\xi} p(x)\,dx = \gamma, \tag{23}$$

that is we must solve Eq. (23) for a selected consecutive value of γ and calculate the consecutive value of ξ.

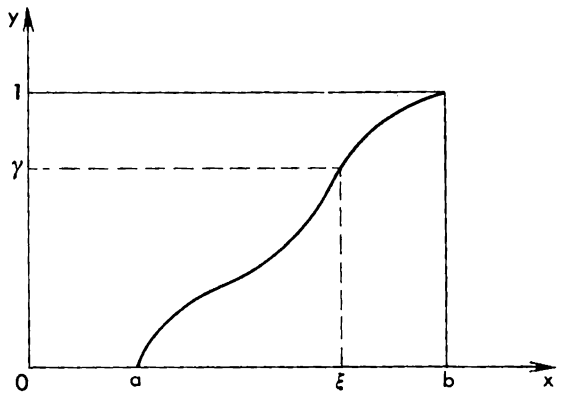

FIG. 12

In order to prove it we consider the function

$$y = \int_a^x p(x)\,dx.$$

It follows from the general properties of density, expressed by Eqs. (15) and (16), that

$$y(a) = 0, \quad y(b) = 1,$$

and the derivative

$$y'(x) = p(x) > 0.$$

It means that the function $y(x)$ monotone increases from 0 to 1 (Fig. 12). Any straight line $y = \gamma$ where $0 < \gamma < 1$, intersects the curve $y = y(x)$ at one and only one point whose abscissa is taken for the value of ξ. Thus Eq. (23) always has a unique solution.

Now we select an arbitrary interval (a', b') within (a, b). The ordinates of the curve $y = y(x)$ satisfying the inequality

$$y(a') < y < y(b')$$

correspond to the points of this interval

$$a' < x < b'.$$

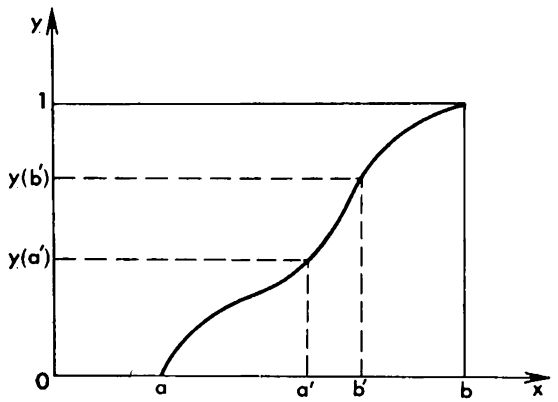

FIG. 13

Consequently if ξ belongs to the interval $a' < x < b'$, then γ belongs to the interval $y(a') < y < y(b')$, and vice versa (Fig. 13). Hence

$$\mathbf{P}\{a' < \xi < b'\} = \mathbf{P}\{y(a') < \gamma < y(b')\}.$$

Since γ is uniformly distributed in (0, 1),

$$\mathbf{P}\{y(a') < \gamma < y(b')\} = y(b') - y(a') = \int_{a'}^{b'} p(x)\,dx.$$

Thus

$$\mathbf{P}\{a' < \xi < b'\} = \int_{a'}^{b'} p(x)\,dx,$$

and this precisely means that the random variable ξ which is the root of Eq. (23) has the probability density $p(x)$.

Example. A random variable η is referred to as *uniformly distributed in the interval* (a, b) if its density in this interval is constant

$$p(x) = (b-a)^{-1} \quad \text{for } a < x < b.$$

To draw the value of η, we apply Eq. (23):

$$\int_a^{\eta} \frac{dx}{b-a} = \gamma.$$

The integral is easily calculated:

$$\frac{\eta - a}{b-a} = \gamma.$$

Hence we arrive at an explicit expression for η:

$$\eta = a + \gamma (b - a). \tag{24}$$

Other examples of application of Eq. (23) are given in paragraphs 5.2 and 8.3.

4.3. Neumann's Method of Drawing a Continuous Random Variable. It may so happen that Eq. (23) is very hard

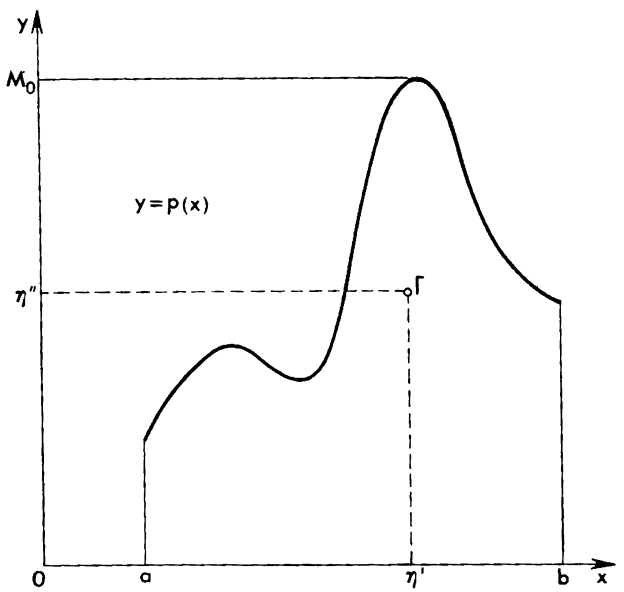

FIG. 14

to solve for ξ. For instance, when the integral of $p(x)$ cannot be expressed via elementary functions or when the density $p(x)$ is specified graphically.

Let us assume that the random variable ξ is specified on a finite interval (a, b) and its density is limited (Fig. 14):

$$p(x) \leqslant M_0.$$

The variable ξ may be drawn as follows:

1. Select two values γ' and γ'' of the random variable γ and generate a random point $\Gamma(\eta'; \eta'')$ with coordinates

$$\eta' = a + \gamma'(b-a), \quad \eta'' = \gamma'' M_0.$$

2. If the point Γ lies below the curve $y = p(x)$, assume $\xi = \eta'$; if the point Γ is above the curve $y = p(x)$, reject the pair (γ', γ'') and select a new pair (γ', γ'').

The validity of this method is substantiated in Sec. 9.1.

4.4. Drawing of Normal Variables. A large number of various techniques exist for drawing various random variables. We shall not consider them here. They are usually applied only when the methods described in Sections 4.2 and 4.3 prove ineffective.

Such is the case, in particular, for the normal random variable ζ, since the equation

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\zeta} e^{-\frac{x^2}{2}} dx = \gamma$$

cannot be solved explicitly and the interval of possible values ζ is infinite.

The Table B at the end of the book (on page 74) lists the values (already drawn) of the normal random variable ζ with mathematical expectation $\mathbf{M}\zeta = 0$ and variance $\mathbf{D}\zeta = 1$. It is not hard to prove* that the random variable

$$\zeta' = a + \sigma\zeta \tag{25}$$

will also be normal; as follows from Eqs. (10) and (11)

$$\mathbf{M}\zeta' = a, \quad \mathbf{D}\zeta' = \sigma^2.$$

Thus, formula (25) enables us to draw by using Table B the values of any normal variable.

4.5. More on the Example from Sec. 1.2. Now we can explain how the random points for Figs. 1 and 2 were sampled. Points plotted in Fig. 1 have coordinates

$$x = \gamma', \quad y = \gamma''.$$

---

* The proof is given in Sec. 9.2.

The values of γ' and γ'' were calculated from groups of five digits of Table A: $x_1 = 0.86515$; $y_1 = 0.90795$; $x_2 = 0.66155$; $y_2 = 0.66434$, etc.

It can be proved* that since the abscissae and ordinates of these points are independent the probability for one such point to fall into any region inside the square is equal to the area of this region.

The points drawn in Fig. 2 have the coordinates

$$x = 0.5 + 0.2\zeta', \quad y = 0.5 + 0.2\zeta'',$$

where the values of ζ' and ζ'' were selected from Table B in succession:

$$x_1 = 0.5 + 0.2 \cdot 0.2005, \quad y_1 = 0.5 + 0.2 \cdot 1.1922;$$
$$x_2 = 0.5 + 0.2 \cdot (-0.0077), \ \ldots$$

One of the points fell outside the square and was rejected.

As follows from Eq. (25), the abscissae and ordinates of these points represent normal random variables with mean values $a = 0.5$ and variances $\sigma^2 = 0.04$.

---

* The proof is given in Sec. 9.3.

CHAPTER II

Examples of Application of the Monte Carlo Method

Sec. 5. Simulation of a Mass Servicing System

5.1. Description of the Problem. Consider one of the simplest systems of mass servicing. This system consists of n lines (or channels, or servicing stations) each of which can "serve on the customers". The system receives requests arriving at random moments of time. Each request arrives at the $N1$ line. If the arrival time of the k-th request (let us call it T_k) finds this line free, the line starts servicing the request; this takes t_b minutes (t_b is *the holding time of the line*). If $N1$ line is busy at the moment T_k, the request is immediately transferred to the $N2$ line. And so on ...

Finally, if all n lines are busy at the moment T_k, the system rejects the request.

The problem is, what will be the (average) number of requests serviced by the system during the period T and how many rejections will be given?

It is clear that problems of this type are encountered when functioning of organizations of any type, not only of the social service sector, is investigated. In some very particular cases analytical solutions were found. But in complicated cases (we shall discuss them below) the Monte Carlo method proved to be the only one available computational method.

5.2. Simple Flow of Requests. The first question that comes on when such systems are analyzed is: what is the flow of arriving requests? This question is answered experimentally by a sufficiently long observation of the requests. Investigation of the request flows in various conditions permitted to single out certain cases which are met with sufficiently frequently.

A *simple flow* (or a *Poisson flow*) is the flow of requests in which the time interval τ between two consecutive re-

quests is a random variable distributed in the interval $(0, \infty)$ with the density

$$p(x) = ae^{-ax}. \tag{26}$$

The density (26) is also called the *exponential distribution* (see Fig. 15 where $p(x)$ is plotted for $a = 1$ and $a = 2$).

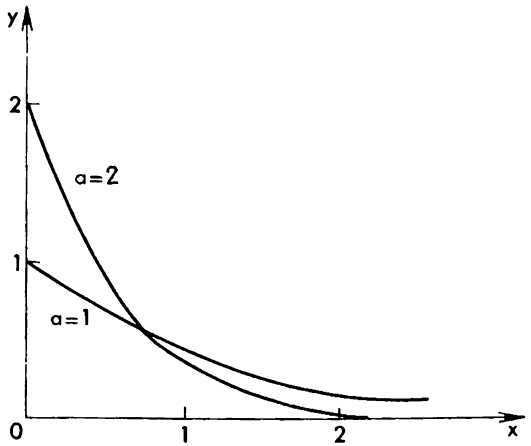

FIG. 15

Calculation of the expected value of τ is straightforward:

$$\mathbf{M}\tau = \int_0^\infty xp(x)\,dx = \int_0^\infty xae^{-ax}\,dx.$$

Integration by parts ($u = x$, $dv = ae^{-ax}\,dx$) yields

$$\mathbf{M}\tau = [-xe^{-ax}]_0^\infty + \int_0^\infty e^{-ax}\,dx = \left[-\frac{e^{-ax}}{a}\right]_0^\infty = \frac{1}{a}.$$

The parameter a is called the *request flow density.*

The formula for drawing τ is easily obtained by means of Eq. (23) which in our case takes the form

$$\int_0^\tau ae^{-ax}\,dx = \gamma.$$

By calculating the integral on the left-hand side we obtain

$$1 - e^{-a\tau} = \gamma,$$

whence

$$\tau = -\frac{1}{a} \ln (1 - \gamma).$$

However, the distribution of the variable $1 - \gamma$ is identical with that of γ so that instead of the above formula we can use the expression

$$\tau = -\frac{1}{a} \ln \gamma. \tag{27}$$

5.3. The Sequence of Computations. We shall now consider the functioning of the system of Sec. 5.1 in case of a simple flow of requests.

Let each line correspond to one cell of an internal data store of the electronic computer; the instant when this line clears will be recorded in this cell. We denote the moment of freeing the i-th line by t_i. The moment of arrival of the first request is chosen as the origin on the time axis: $T_1 = 0$. We can assume that all t_i are equal at this moment to T_1: all lines are free. Computation termination time is $T_{end} = T_1 + T$.

The first request arrives at the $N1$ line. Hence this line will be busy during t_b. Thus, we must replace t_1 by a new value $(t_1)_{new} = T_1 + t_b$, add unity to the counter of serviced requests and turn to deal with the second request.

Let us suppose that k requests have already been considered. Now we have to draw the moment of arrival of the $(k + 1)$-th request. We sample a consecutive value of γ and calculate the consecutive value $\tau = \tau_k$ by means of Eq. (27). Then we calculate the time of arrival

$$T_{k+1} = T_k + \tau_k.$$

Is the first line free at this moment? To find it out we must check the condition

$$t_1 \leqslant T_{k+1}. \tag{28}$$

If this condition is satisfied it means that at the moment T_{k+1} the line is already cleared and can service the request. We replace t_1 by $T_{k+1} + t_b$, add a unit to the counter of the serviced requests, and then switch over to the next request.

If condition (28) is not satisfied it means that at the moment T_{k+1} the first line is busy. We shall now check whether the second line is free:

$$t_2 \leqslant T_{k+1}? \tag{29}$$

If condition (29) is satisfied we replace t_2 by $T_{k+1} + t_b$, add unity to the counter of serviced requests and switch over to the next request.

If condition (29) is not satisfied, we turn to checking the condition

$$t_3 \leqslant T_{k+1}.$$

It may happen that for all i from 1 to n

$$t_i > T_{k+1},$$

i.e. all lines are busy at the moment T_{k+1}. We then add unity to the counter of rejections and switch over to processing the next request.

Each time when T_{k+1} is calculated we have to check the condition of termination of the experiment

$$T_{k+1} > T_{end}.$$

Once this condition is satisfied, the experiment is terminated. The counter of satisfied requests and the counter of rejections will contain the numbers μ_{sat} and μ_{rej}.

This experiment is repeated N times (each time with different values of γ). The results of all experiments are averaged:

$$\mathbf{M}\mu_{sat} \approx \frac{1}{N} \sum_{j=1}^{N} \mu_{sat(j)},$$

$$\mathbf{M}\mu_{rej} \approx \frac{1}{N} \sum_{j=1}^{N} \mu_{rej(j)},$$

where $\mu_{sat(j)}$ and $\mu_{rej(j)}$ are the values of μ_{sat} and μ_{rej} obtained in the j-th experiment.

Fig. 16 shows the block-diagram of a program realizing these computations (if necessary, we can retrieve the values $\mu_{sat(j)}$ and $\mu_{rej(j)}$ of individual trials in the "end of experiment" block).

5.4. More Complicated Problems. It is not hard to show that the same method permits computation for much more

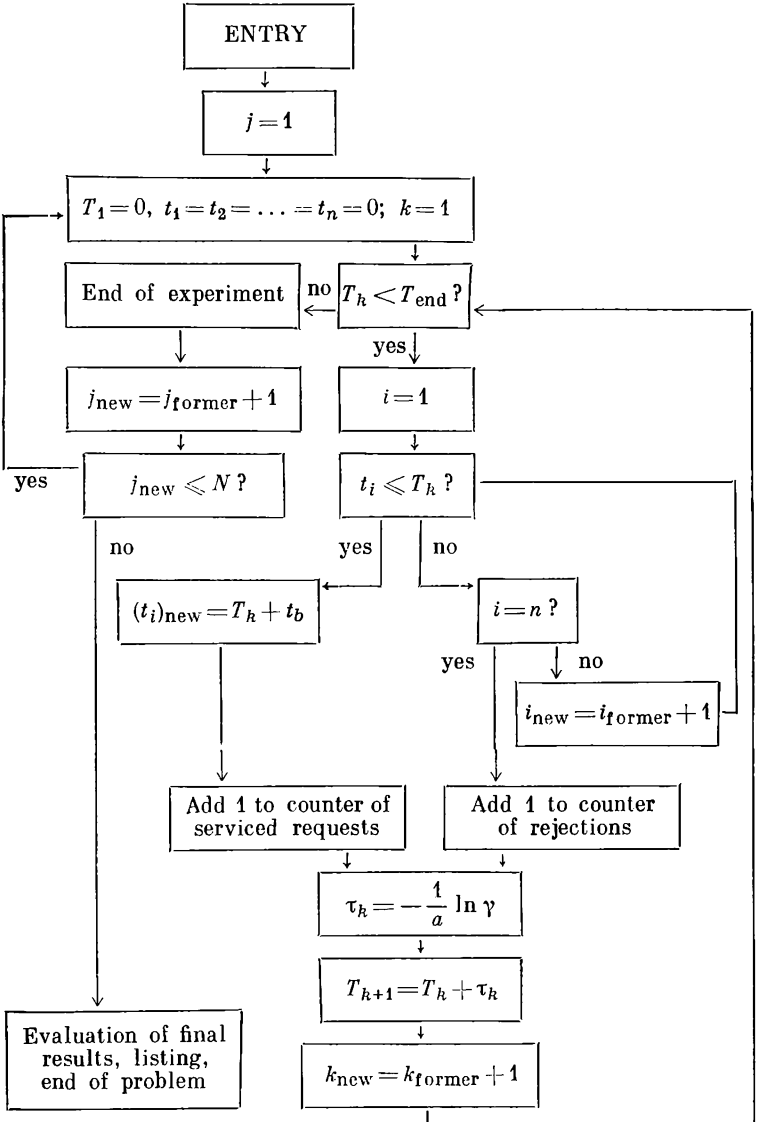
ENTRY
$j=1$
$T_1=0,\ t_1=t_2=\ldots=t_n=0;\ k=1$
End of experiment
no
$T_k<T_{\text{end}}$?
yes
$j_{\text{new}}=j_{\text{former}}+1$
$i=1$
yes
$j_{\text{new}}\leqslant N$?
$t_i\leqslant T_k$?
no
yes
no
$(t_i)_{\text{new}}=T_k+t_b$
$i=n$?
yes
no
$i_{\text{new}}=i_{\text{former}}+1$
Add 1 to counter of serviced requests
Add 1 to counter of rejections
$\tau_k=-\frac{1}{a}\ln\gamma$
$T_{k+1}=T_k+\tau_k$
$k_{\text{new}}=k_{\text{former}}+1$
Evaluation of final results, listing, end of problem

FIG. 16

complicated systems. For instance, the holding time t_b may be not constant but random and vary from line to line (this corresponds to different types of equipment and to unequal skills of the staff). The scheme of computations remains basically unchanged but the values of t_{b_i} will have to be drawn in each trial and the formula for drawing will be specific for each line.

We can also consider the so-called *systems with waiting* in which the rejection is not immediate: the request is kept for some time t_r (*time the request remains in the system*) and is serviced if one of the lines is cleared during this time.

Another type of systems, in which the next request is serviced by the first line to be cleared, can be considered. We can take into account a random failure of each individual line and a random time interval necessary to repair it. We can take into account time variations of the request flow density, as well as many additional factors.

As a matter of course, a certain price has to be paid for simulation of such systems. Practically valuable results are only obtained if a good model is selected. For this purpose we have to study thoroughly actual request flows, to carry out timing of functioning of individual equipment units, etc.

Generally speaking, we have to know the probabilistic laws governing the functioning of separate parts of the system. Then the Monte Carlo method makes it possible to calculate the probabilistic laws of the entire system operation, regardless of its complexity.

Such methods of calculation are extremely helpful in designing economic units: expensive (and sometimes even impossible) actual experiment is substituted by a simulated computer experiment in which various versions of work organization or equipment operation are simulated.

Sec. 6. Computation of Quality and Reliability of Complex Devices

6.1. The Simplest Scheme of Computing Quality. Let us consider a device S consisting of a certain (possibly large) number of elements. If, for instance, S is an electric apparatus, then its elements can be resistors ($R_{(k)}$), capacitors ($C_{(k)}$), electron tubes, etc. Let us assume that the quality

of the device is determined by the value of a single output parameter V which can be calculated from the known parameters of all the elements

$$V = f(R_{(1)}, R_{(2)}, \ldots; C_{(1)}, C_{(2)}, \ldots; \ldots). \qquad (30)$$

If, for example, V is the voltage across an operating part of an electric circuit, the equations for this circuit can be written by using Ohm's laws so that V is obtained by solving these equations.

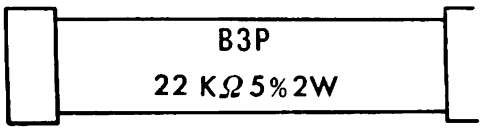

FIG. 17

In practice, however, the parameters of the elements are not exactly equal to the indicated values. For instance, the resistance of the resistor shown in Fig. 17 may vary within the range from 20.9 to 23.1 kΩ.

The question now is: how will V be affected by deviations of the parameters of all elements from their rated values?

We can try and estimate the limits of V fluctuations by choosing the "worst" values of the parameters for all elements. However, it is not always known which set of parameters is the "worst"? Moreover, if the total number of elements is large, such an estimate will be much too high: it is indeed very unlikely that all parameters will take on their worst values simultaneously.

It is, therefore, more reasonable to consider the parameters of all elements and the variable V itself random and to try and estimate the expected value $\mathbf{M}V$ and the variance $\mathbf{D}V$. The value $\mathbf{M}V$ is the mean value of V for the whole lot of items and $\mathbf{D}V$ shows what deviations of V from $\mathbf{M}V$ will be encountered in practice.

Let us recall (it was mentioned in Sec. 2.2) that

$$\mathbf{M}V \neq f(\mathbf{M}R_{(1)}, \mathbf{M}R_{(2)}, \ldots; \mathbf{M}C_{(1)}, \mathbf{M}C_{(2)}, \ldots; \ldots).$$

The distribution of V cannot be computed analytically if the function f is complicated to even the smallest extent. Sometimes we can do it experimentally by scanning a large

batch of final items of the device. However, this is not always possible; and it is never possible at the design stage.

Now let us attempt to apply the Monte Carlo method. For this we need: (a) knowledge of the probabilistic characteristics of all elements; (b) knowledge of the function f (rather, we must be able to calculate V values from all specified values $R_{(1)}, R_{(2)}, \ldots; C_{(1)}, C_{(2)}, \ldots; \ldots$).

Probabilistic distributions of parameters for each individual element can be obtained experimentally, by scanning a large batch of such elements. These distributions are very often normal. For this reason many investigators use the following approach: they consider, for example, the resistance of the element shown in Fig. 17, as a normal random variable ρ with the expected value $\mathbf{M}\rho = 22$ and $3\sigma = 1.1$ (it will be recalled that it is practically improbable that we obtain in one trial the value of ρ that deviates from $\mathbf{M}\rho$ by more than 3σ; see Eq. (20)).

The scheme for the computation is quite simple: first the value of the parameter is drawn for each element; then the V value is calculated by formula (30). Having repeated this experiment N times we obtain values $V_1, V_2, \ldots, V_N$ and can approximately assume that

$$\mathbf{M}V \approx \frac{1}{N} \sum_{j=1}^{N} V_j,$$

$$\mathbf{D}V \approx \frac{1}{N-1} \left[\sum_{j=1}^{N} (V_j)^2 - \frac{1}{N} \left(\sum_{j=1}^{N} V_j \right)^2 \right].$$

At large values of N the factor $1/(N-1)$ in the last formula can be replaced by $1/N$; then this formula will directly follow from Eqs. (8) and (9). Mathematical statistics proves that the factor $1/(N-1)$ provides better results for small N.

6.2. Examples of Calculating Reliability. Suppose we want to estimate the mean duration of failure-proof operation of the apparatus provided all characteristics of the failure-proof operation of each of the elements are known.

If we assume that the duration of failure-proof operation of each element $t_{(k)}$ is fixed, then the calculation of time t of failure-proof operation of the whole unit presents no problems. For example, failure of any one of the elements of

the unit drawn in Fig. 18 entails the failure of the whole unit, i.e.

$$t = \min(t_{(1)};\ t_{(2)};\ t_{(3)};\ t_{(4)}). \tag{31}$$

For the unit shown in Fig. 19 where one of the elements has a stand-by element

$$t = \min[t_{(1)},\ t_{(2)};\ \max(t_{(3)};\ t_{(4)};\ t_{(5)}], \tag{32}$$

since when one element, $N3$ for example, will break down, the unit will continue functioning on the remaining element $N4$.

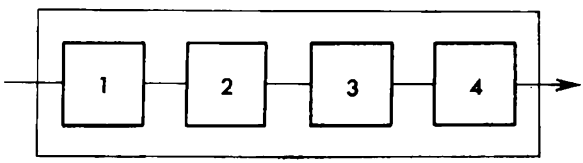

FIG. 18

In reality the duration of failure-proof operation of any single element is a random variable $\theta_{(k)}$. When we say that the service life of an electric bulb is 1000 hours it is only

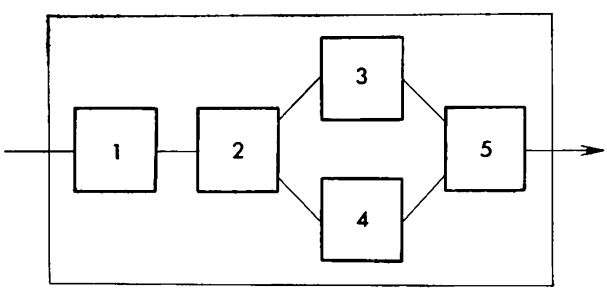

FIG. 19

the mean value $M\theta$ of the variable θ: everyone is aware that one bulb burns out sooner while another one (exactly identical with the former) lives longer.

When distribution densities of $\theta_{(k)}$ are known for each of the elements of the unit, $M\theta$ can be computed exactly as it was performed in Sec. 6.1. Indeed, $\theta_{(k)}$ can be drawn

for each element; let it be $t_{(k)}$. Then the value t of the random variable θ can be calculated by means of relevant formulas (31) or (32). Having rerun this trial a sufficient number of times (N) we can accept that

$$\mathrm{M}\theta \approx \frac{1}{N} \sum_{j=1}^{N} t_j,$$

where t_j is the value of t obtained in the j-th trial.

It should be mentioned that the problem of distribution of service life $\theta_{(k)}$ of individual elements is not so simple as one might think: running of an experiment involving maximum longevity elements is cumbersome because we need to wait until a sufficiently large number of elements break down.

6.3. Further Possibilities of the Method. The examples given above demonstrate that in principle the techniques of computing quality of the articles being designed is quite simple. One must know probabilistic characteristics of all elements of the unit and be able to calculate the value of interest as a function of the parameters of these elements. Then the random character of these parameters is taken into account by simulation.

Simulation can yield much more useful information than just the expected value and the variance of the variable of interest. Suppose, for example, that we have sampled a large number N of values $U_1, U_2, \ldots, U_N$ of a random variable U. On the basis of these numbers we can plot an approximate distribution density of U. In fact this is the domain of statistics because we touched upon processing of experimental data (produced in computer-simulated experiments). Thus we will restrict ourselves to the presentation of one specific example.

Let us assume that all in all we obtained $N = 120$ values $U_1, U_2, \ldots, U_{120}$ of the random variable U, all distributed in the range

$$1 < U_j < 6.5.$$

Let us divide the interval $1 < x < 6.5$ into 11 (or any other number, not too large and not too small) equal intervals with lengths $\Delta x = 0.5$ and count how many values of U_j fall into each of the intervals. Corresponding numbers are indicated in Fig. 20.

The frequency of falling into any one of the intervals is calculated by dividing these numbers by $N = 120$. In our case the frequencies are: 0.017; 0; 0.008; 0.12; 0.20; 0.27; 0.14; 0.16; 0.06; 0.008; 0.017.

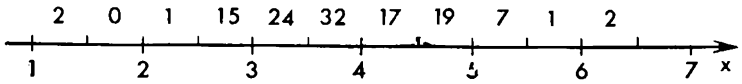

FIG. 20

Now we draw a rectangle over each of the intervals with the area equal to the frequency corresponding to this interval (Fig. 21). In other words, the height of each rectangle is equal to frequency divided by Δx. The step line obtained is called a *histogram.*

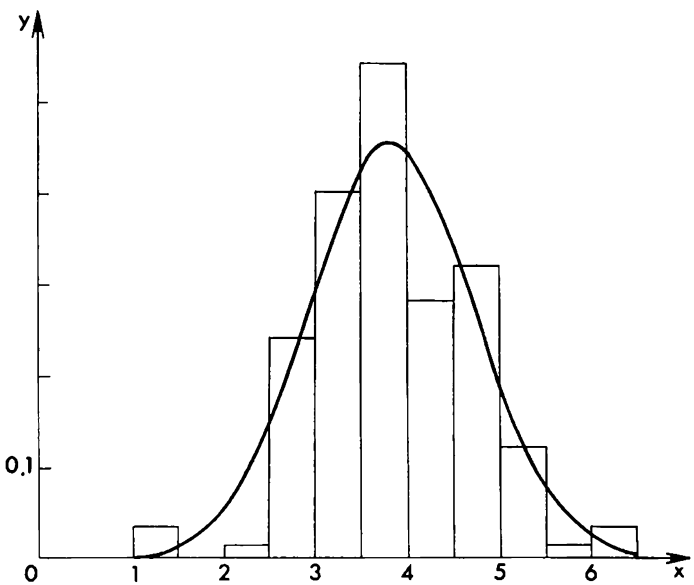

FIG. 21

The histogram is an approximation to the unknown density of the random variable U. Thus, for instance, the area of this histogram between $x = 2.5$ and $x = 5.5$ yields an

approximate value of the probability

$$\mathbf{P}\{2.5 < U < 5.5\} \approx 0.94.$$

Hence, we can assume on the basis of the above calculation that with the probability approximately equal to 0.94 the value of U lies within the interval $2.5 < U < 5.5$.

For the purpose of comparison the density of a normal random variable ζ' with parameters $a = 3.85$, $\sigma = 0.88$ is plotted in Fig. 21. If the probability of ζ' being inside the interval $2.5 < \zeta' < 5.5$ is calculated from the normal density, we obtain a fairly close value 0.91*.

6.4. Note. Unfortunately, calculations of this type are still fairly scarce. It is not easy to say what is the main reason for this. The most likely explanation is that designers simply are not aware of this possibility.

Besides, before performing computations concerning devices of any type we have to analyze the probabilistic characteristics of all elements incorporated into these devices. This means a lot of work. But once these characteristics are known, we can evaluate the quality of any device consisting of these elements. The variation of the quality caused by replacing some of the elements by other ones can be estimated as well.

---

* The course of calculations. According to Eq. (14) we have

$$\mathbf{P}\{2.5 < \zeta' < 5.5\} = \frac{1}{\sigma\sqrt{2\pi}} \int_{2.5}^{5.5} e^{-\frac{(x-a)^2}{2\sigma^2}}\, dx.$$

We change the variable in the integral $x - a = \sigma t$. This yields

$$\mathbf{P}\{2.5 < \zeta' < 5.5\} = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{t^2}{2}}\, dt,$$

where $t_1 = \dfrac{2.5 - a}{\sigma} = -1.54$, $t_2 = \dfrac{5.5 - a}{\sigma} = 1.88$. The last integral is evaluated by using one of the tables of the so-called error function which lists the values of the function

$$\Phi(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}}\, dt.$$

We obtain

$$\mathbf{P}\{2.5 < \zeta' < 5.5\} = 0.5\,[\Phi(1.88) + \Phi(1.54)] = 0.91.$$

We should hope that in the nearest future such calculations will become a routine business, and the probabilistic characteristics of the elements will invariably be supplied by their manufacturers.

Sec. 7. Computation of Neutron Transmission Through a Plate

The probabilistic laws of interaction of an individual elementary particle (neutron, photon, meson, etc.) with matter are known. Usually it is needed to find the macroscopic characteristics of processes in which the number of participating particles is enormous, such as densities, fluxes and so on. This situation is quite similar to that encountered in Secs. 5 and 6, and offers a very convenient object for the Monte Carlo calculations.

The neutron physics is probably the field where the Monte Carlo method is used more frequently than anywhere else. We shall consider the simplest version of the problem of neutron transmission through a plate.

7.1. Statement of the Problem. Let the flux of neutrons with energy E_0 be incident on a homogeneous infinite plate $0 \leqslant x \leqslant h$. The angle of incidence is 90°. Upon collision with atoms of the plate material neutrons may be either elastically scattered or captured. Let us assume for simplicity that energy of neutrons is not changed in scattering and that any direction of "recoil" of a neutron from an atom is equally probable (this is sometimes the case in neutron collisions with heavy atoms). Fig. 22 illustrates possible fates of a neutron: neutron (a) passes through the plate, neutron (b) is captured, and neutron (c) is reflected by the plate.

It is necessary to calculate the probability of neutron transmission through the plate p^+, the probability of neutron reflection by the plate p^- and the probability of neutron capture inside the plate p^0.

Interaction of neutrons with matter is characterized in the case under consideration by two constants $\sum_c$ and $\sum_s$ which are called the *capture cross-section* and the *scattering cross-section*, respectively.

The sum of these two cross-sections is called the *total cross-section*

$$\sum = \sum_c + \sum_s.$$

The physical meaning of these cross-sections is as follows: the probability of neutron capture upon collision with an atom is equal to $\sum_c^1/\sum^1$, and the probability of scattering is equal to $\sum_s^1/\sum^1$.

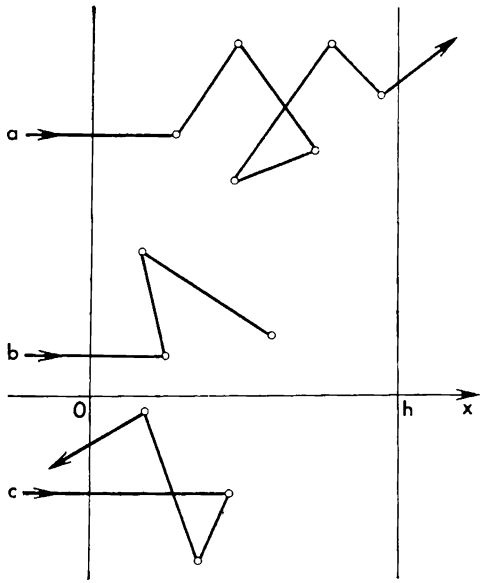

FIG. 22

The *free path length of a neutron* λ (i.e. the length of the path from one collision to another) is a random variable. It can take on any positive value with the probability density

$$p(x) = \sum^1 e^{-\Sigma x}.$$

It is easily seen that this density of λ coincides with the density (26) of the random variable τ of a simple request flow. Similarly to Sec. 5.2 we can immediately put down an expression for the mean free path length

$$\mathbf{M}\lambda = 1/\sum^1$$

and the formula for drawing λ values

$$\lambda = -\frac{1}{\sum^1} \ln \gamma.$$

We now have to find out how to select a random direction of a neutron after scattering. As the problem is symmetrical about the x-axis, the direction is completely defined by an angle φ between the neutron velocity and the Ox-axis. It can be proved* that the requirement of equal probability of all directions is equivalent in this case to the requirement that the cosine of this angle $\mu = \cos \varphi$ be uniformly distributed over the interval $(-1, 1)$. Formula (24) for $a = -1$, $b = 1$ yields the expression for drawing μ values:

$$\mu = 2\gamma - 1.$$

7.2. Computation by Means of Simulation of Real Trajectories. Let us assume that a neutron has undergone the k-th

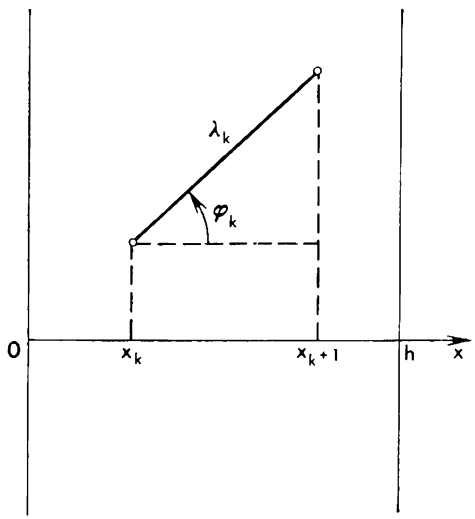

FIG. 23

scattering inside the plate at the point with the abscissa x_k and started moving after it in the direction μ_k.

We draw the free path length

$$\lambda_k = -(1/\Sigma) \ln \gamma$$

and calculate the abscissa of the next collision (Fig. 23)

$$x_{k+1} = x_k + \lambda_k \mu_k.$$

* The proof is given in Sec. 9.4.

We check the condition of neutron transmission through the plate:

$$x_{k+1} > h.$$

If this condition is satisfied, the computation of the trajectory is terminated and unity is added to the counter of transmitted particles. Otherwise we check the condition of reflection:

$$x_{k+1} < 0.$$

If this condition is satisfied, the computation of the trajectory is also terminated and unity is added to the counter of reflected particles. If this condition is not satisfied, i.e. $0 \leqslant x_{k+1} \leqslant h$, it means that the neutron has undergone the $(k + 1)$-th collision inside the plate and we have to draw its "fate" in this collision.

According to Sec. 4.1 we choose the next value of γ and check the capture condition

$$\gamma < \Sigma_c/\Sigma.$$

If this last inequality is satisfied, the trajectory is terminated and we add unity to the counter of captured particles. Otherwise we assume that the neutron had been scattered at the point with the abscissa x_{k+1}. In this case we draw a new direction of the neutron velocity

$$\mu_{k+1} = 2\gamma - 1$$

and then repeat the whole cycle (but, of course, with new values of γ).

In all of the above formulas we added no subscript to γ because it is meant that each one value of γ is used only once. Three values of γ are necessary to compute one lap of the trajectory.

The initial values for each trajectory are:

$$x_0 = 0, \quad \mu_0 = 1.$$

After N trajectories are sampled we obtain that N^+ neutrons were transmitted through the plate, N^- neutrons were reflected and N^0 neutrons captured by it. It is obvious that the probabilities we seek are appoximately equal to the ratios

$$p^+ \approx \frac{N^+}{N}, \quad p^- \approx \frac{N^-}{N}, \quad p^0 \approx \frac{N^0}{N}.$$

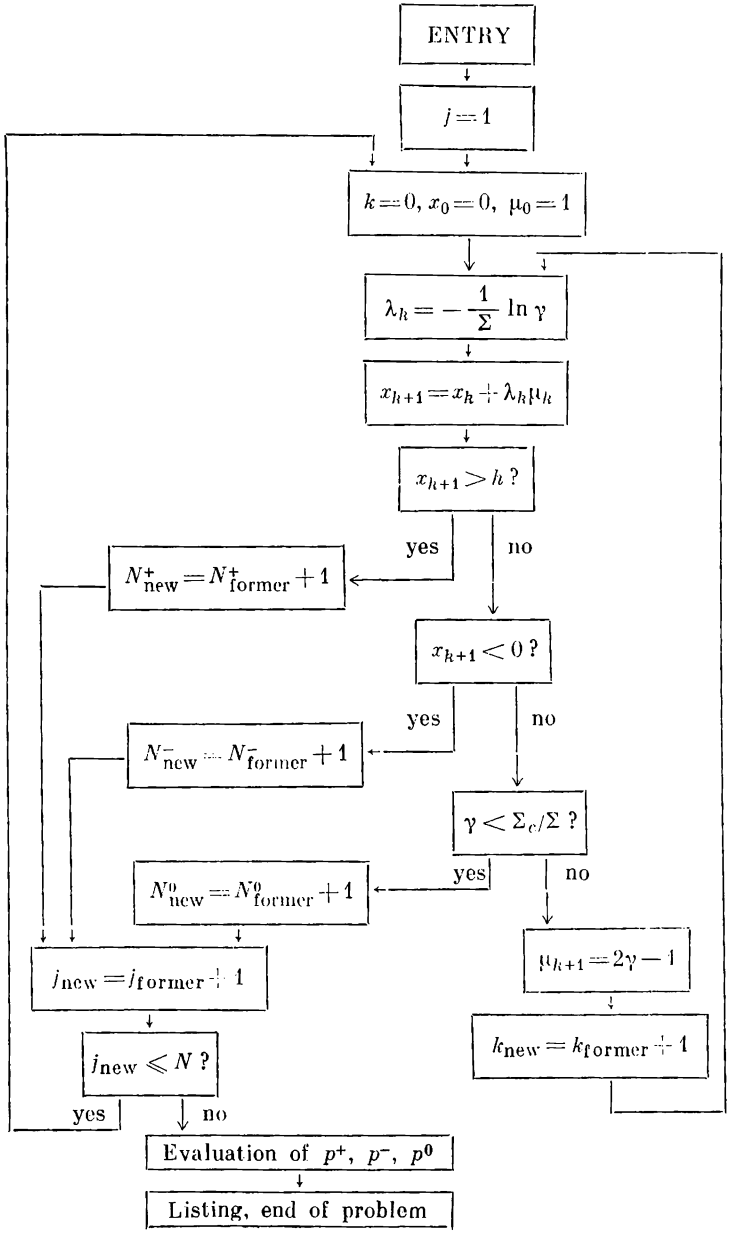
ENTRY
$j = 1$
$k = 0,\ x_0 = 0,\ \mu_0 = 1$
$\lambda_k = -\frac{1}{\Sigma} \ln \gamma$
$x_{k+1} = x_k + \lambda_k \mu_k$
$x_{k+1} > h$?
yes
no
$N^+_{\text{new}} = N^+_{\text{former}} + 1$
$x_{k+1} < 0$?
yes
no
$N^-_{\text{new}} = N^-_{\text{former}} + 1$
$\gamma < \Sigma_c / \Sigma$?
yes
no
$N^0_{\text{new}} = N^0_{\text{former}} + 1$
$\mu_{k+1} = 2\gamma - 1$
$j_{\text{new}} = j_{\text{former}} + 1$
$k_{\text{new}} = k_{\text{former}} + 1$
$j_{\text{new}} \leqslant N$?
yes
no
Evaluation of p^+, p^-, p^0
Listing, end of problem

FIG. 24

Fig. 24 shows the block diagram of the program of computations for this problem. The subscript j is the trajectory number and the subscript k is the collision number (along the trajectory).

This computation procedure, though quite natural, is far from being perfect. In particular, by applying this method we run into difficulties if the probability p^+ is very small. However, this is just the case that we have in designing a radiation-proof shielding.

There exist some more "ingenious" versions of the Monte Carlo method which make it possible to carry out calculations in this case as well. We shall now briefly discuss one of the simplest versions of computations involving the so-called "weights".

7.3. Computation Scheme Using Weights which Replace Absorption of Neutrons. Let us consider now the same problem of neutron transmission. Let us suppose that a "packet" consisting of a large number w_0 of identical neutrons moves along a single trajectory. The average number of neutrons captured from the "packet" in a collision at the point with the abscissa x_1, will be equal to $w_0\ (\Sigma_c/\Sigma)$, and the average number of scattered neutrons, to $w_0\ (\Sigma_s/\Sigma)$.

We add the quantity $w_0\ (\Sigma_c/\Sigma)$ to the counter of captured particles and will follow the scattered "packet" under the assumption that the remaining "packet" was scattered as a whole in one direction.

All computation formulas given in Sec. 7.2 remain unchanged but the number of neutrons in the "packet" upon each collision will be reduced to

$$w_{k+1} = w_k\,(\Sigma_s/\Sigma),$$

since part of the "packet", containing $w_k\ (\Sigma_c/\Sigma)$ neutrons will be captured. The trajectory cannot be terminated by capture any more.

The quantity w_k is usually referred to as the *weight* of a neutron; and rather than speak about a "packet" consisting of w_k neutrons we speak about one neutron with the weight w_k. The initial weight w_0 is usually assumed to be 1. This is not contrary to what was said about a "large packet" because it is easily seen that all w_k obtained in computations of a single trajectory contain w_0 as a common multiplier.

The block diagram of a program realizing this computation is shown in Fig. 25. This diagram is not a shade more

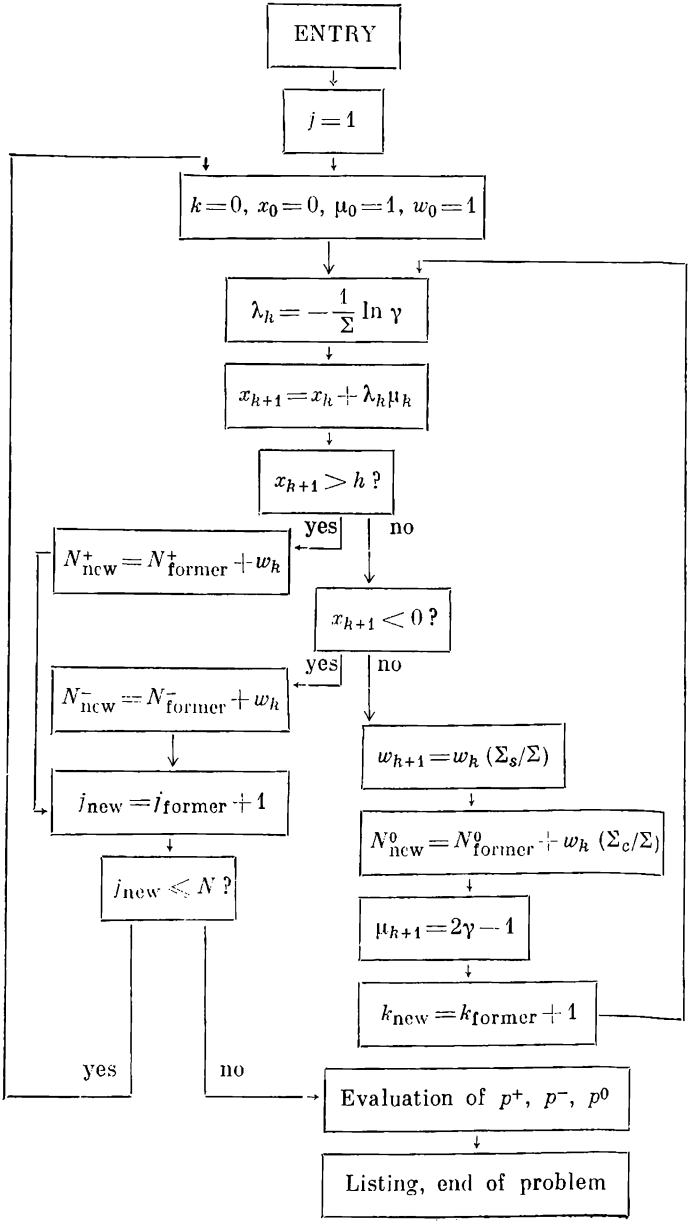
ENTRY
$j=1$
$k=0,\ x_0=0,\ \mu_0=1,\ w_0=1$
$\lambda_k=-\frac{1}{\Sigma}\ln\gamma$
$x_{k+1}=x_k+\lambda_k\mu_k$
$x_{k+1}>h$?
yes
no
$N^+_{\text{new}}=N^+_{\text{former}}+w_k$
$x_{k+1}<0$?
yes
no
$N^-_{\text{new}}=N^-_{\text{former}}+w_k$
$w_{k+1}=w_k\,(\Sigma_s/\Sigma)$
$j_{\text{new}}=j_{\text{former}}+1$
$N^0_{\text{new}}=N^0_{\text{former}}+w_k\,(\Sigma_c/\Sigma)$
$j_{\text{new}}\leqslant N$?
$\mu_{k+1}=2\gamma-1$
$k_{\text{new}}=k_{\text{former}}+1$
yes
no
Evaluation of p^+, p^-, p^0
Listing, end of problem

FIG. 25

complicated than that shown in Fig. 24. However, it can be proved* that the computation of p^+ by the method of this section is always more expedient than by the method covered in Sec. 7.2.

7.4. Note. The number of various methods of calculations using different weights is quite considerable. We cannot dwell on all of them here.

We will only emphasize that the Monte Carlo method makes it possible to solve much more complicated problems concerning elementary particles: the investigated medium may consist of different substances and be of an arbitrary geometric shape; the energy of a particle may change in each collision. Many other nuclear processes may be taken into account (for example, the possibility of atom fission upon collision with a neutron and subsequent formation of additional neutrons). Conditions at which chain reaction is initiated and maintained can also be calculated, etc.

Sec. 8. Calculation of a Definite Integral

The problems considered in Secs. 5, 6 and 7 were of probabilistic nature so that the application of Monte Carlo techniques to their solution appeared fairly natural. Here we consider a standard mathematical problem: approximate computation of a definite integral.

Since the calculation of definite integrals is equivalent to the calculation of areas, we might utilize the method presented in Sec. 1.2. As alternative to this method we shall describe another more efficient method making possible the construction of various probabilistic models to solve this problem by the Monte Carlo method, and indicate how to select a "better" model among these.

8.1. Method of Calculations. Let us consider a function $g(x)$ defined on the interval $a < x < b$. We have to approximate the integral

$$I = \int_a^b g(x)\,dx. \tag{33}$$

Let us choose an arbitrary distribution density $p_\xi(x)$ specified on the interval (a, b) (i.e. an arbitrary function $p_\xi(x)$ satisfying the conditions (15) and (16)).

* The proof is given in Sec. 9.5.

Along with the random variable ξ defined on the interval (a, b) with the density $p_\xi(x)$, we shall need another random variable

$$\eta = g(\xi)/p_\xi(\xi).$$

According to (18)

$$\mathbf{M}\eta = \int_a^b [g(x)/p_\xi(x)]\, p_\xi(x)\, dx = I.$$

Let us consider now N identical independent random variables $\eta_1, \eta_2, \ldots, \eta_N$ and apply the central limit theorem of Sec. 2.4 to their sum. In this case Eq. (21) will take the form

$$P\left\{\left|\frac{1}{N}\sum_{j=1}^{N}\eta_j - I\right| < 3\sqrt{\frac{\mathbf{D}\eta}{N}}\right\} \approx 0.997. \qquad (34)$$

The last relation means that if we sample N values $\xi_1, \xi_2, \ldots, \xi_N$, then for sufficiently large N

$$\frac{1}{N}\sum_{j=1}^{N}\frac{g(\xi_j)}{p_\xi(\xi_j)} \approx I. \qquad (35)$$

It also shows that the error of approximation (35) will not exceed, with very high probability, the value

$$3\sqrt{\mathbf{D}\eta/N}.$$

8.2. How to Choose a Computation Scheme. We have seen that any random variable ξ, defined in the interval (a, b) can be used to calculate integral (33). In all cases

$$\mathbf{M}\eta = \mathbf{M}[g(\xi)/p_\xi(\xi)] = I.$$

But the variance $\mathbf{D}\eta$, and with it the estimated error of (35) depend on what specific variable ξ is used. Indeed

$$\mathbf{D}\eta = \mathbf{M}(\eta^2) - I^2 = \int_a^b [g^2(x)/p_\xi(x)]\, dx - I^2.$$

It can be proved* that this expression takes on the minimum when $p_\xi(x)$ is proportional to $|g(x)|$.

Of course, too complex $p_\xi(x)$ should not be chosen because the procedure of drawing ξ values will become excessively

* The proof is given in Sec. 9.6.

time-consuming. But the suggestion given above is a useful guide-line in the selection of $p_\xi(x)$ (see the Example of Sec. 8.3).

In practice integrals of type (33) are not calculated by means of the Monte Carlo method: more accurate methods, namely quadrature formulas are available. But the situation changes drastically when we consider multiple integrals: quadrature formulas become extremely complicated while the Monte Carlo method remains almost unchanged.

8.3. Numerical Example. Let us approximately calculate the integral

$$I = \int_0^{\pi/2} \sin x \, dx.$$

The exact value of this integral is known

$$\int_0^{\pi/2} \sin x \, dx = [-\cos x]_0^{\pi/2} = 1.$$

We shall use two different random variables ξ for this calculation: one with constant density $p_\xi(x) \equiv 2/\pi$ (i.e. ξ is uniformly distri-

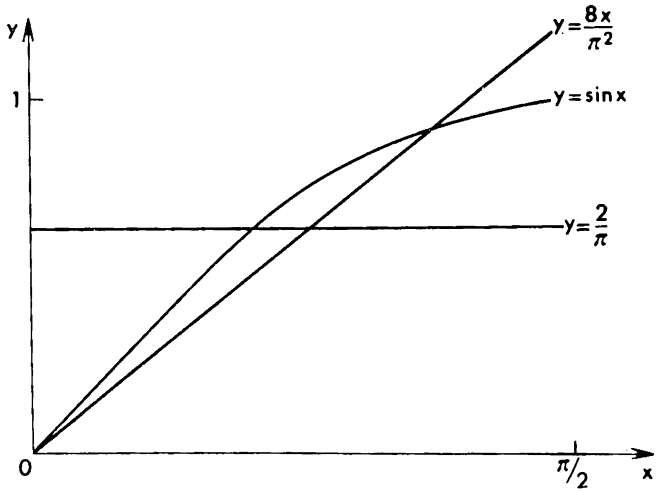

FIG. 26

buted in the interval $(0, \pi/2)$ and the other with linear density $p_\xi(x) = 8x/\pi^2$. Both these densities and the integrand $\sin x$ are plotted in Fig. 26. This figure shows that the linear density agrees better with

the suggestion of Sec. 8.2 that proportionality of $p_\xi(x)$ and $\sin x$ be desirable. Thus we can expect that the second way of computation will produce a better result.

(a) Let $p_\xi(x) \equiv 2/\pi$ in the interval $(0, \pi/2)$. The formula for drawing ξ can be obtained from Eq. (24) with $a = 0$ and $b = \pi/2$:

$$\xi = \pi\gamma/2.$$

Now formula (35) takes the form

$$I \approx \frac{\pi}{2N} \sum_{j=1}^{N} \sin \xi_j.$$

Let us set $N = 10$. For the values of γ we use triads of numbers from Table A (multiplied by 0.001). Intermediate results are listed in Table 1.

TABLE 1

j	1	2	3	4	5
γ_j	0.865	0.159	0.079	0.566	0.155
ξ_j	1.359	0.250	0.124	0.889	0.243
$\sin \xi_j$	0.978	0.247	0.124	0.776	0.241
j	6	7	8	9	10
γ_j	0.664	0.345	0.655	0.812	0.332
ξ_j	1.043	0.542	1.029	1.275	0.521
$\sin \xi_j$	0.864	0.516	0.857	0.957	0.498

The final result is

$$I \approx 0.952.$$

(b) Now we set $p_\xi(x) = 8x/\pi^2$. To draw ξ we use Eq. (23)

$$\int_0^\xi (8x/\pi^2)\, dx = \gamma,$$

whence, after simple calculations, we obtain

$$\xi = \frac{\pi}{2}\sqrt{\gamma}.$$

Eq. (35) takes now the following form

$$I \approx \frac{\pi^2}{8N} \sum_{j=1}^{N} \frac{\sin \xi_j}{\xi_j}.$$

Let $N = 10$. We choose the same values of γ as in the case (a). The intermediate results are listed in Table 2.

TABLE 2

j	1	2	3	4	5
γ_j	0.865	0.159	0.079	0.566	0.155
ξ_j	1.461	0.626	0.442	1.182	0.618
$\frac{\sin \xi_j}{\xi_j}$	0.680	0.936	0.968	0.783	0.937
j	6	7	8	9	10
γ_j	0.664	0.345	0.655	0.812	0.332
ξ_j	1.280	0.923	1.271	1.415	0.905
$\frac{\sin \xi_j}{\xi_j}$	0.748	0.863	0.751	0.698	0.868

The final result is

$$I \approx 1.016.$$

As we have expected the second way of computation yielded a more accurate result.

8.4. On the Error Evaluation. As noted in Sec. 8.2 the absolute error of computation of the integral I by means of formula (35) cannot practically exceed the value $3\sqrt{\mathbf{D}\eta/N}$. But the actual error is, as a rule, appreciably smaller than this estimate. Thus in practice another quantity is used to characterize the error, namely, the so-called probable error

$$\delta_{prob} = 0.675\sqrt{\mathbf{D}\eta/N}.$$

The actual absolute error depends on random numbers used in the computation and may prove to be 2-3 times greater or several times smaller than δ_{prob}. Thus, δ_{prob} estimates not the upper limit of the error but only its order of magnitude*.

Let us go back to the example of Sec. 8.3. The values listed in Tables 1 and 2 can be used for an approximate evaluation of variances

* Definition of the probable error is given in Sec. 9.7.

$D\eta$ for both methods of computation. The relevant computation formula is given in Sec. 6.1.*

The approximate variances $D\eta$, probable errors δ_{prob} calculated from these variances, and actual errors of calculations δ_{calc} are listed in Table 3.

TABLE 3

Method	$D\eta$	δ_{prob}	δ_{calc}
(a)	0.256	0.103	0.048
(b)	0.016	0.027	0.016

It can be seen that δ_{calc} are indeed of the same order as δ_{prob}.

---

* In the case (a):

$$D\eta \approx \frac{\pi^2}{9 \cdot 4}\left[\sum_1^{10} (\sin \xi_j)^2 - \frac{1}{10}\left(\sum_1^{10} \sin \xi_j\right)^2\right] =$$

$$= \frac{\pi^2}{36}(4.604 - 3.670) = 0.256.$$

In the case (b)

$$D\eta \approx \frac{\pi^4}{9 \cdot 64}\left[\sum_1^{10} \left(\frac{\sin \xi_j}{\xi_j}\right)^2 - \frac{1}{10}\left(\sum_1^{10} \frac{\sin \xi_j}{\xi_j}\right)^2\right] =$$

$$= \frac{\pi^4}{576}(6.875 - 6.777) = 0.016.$$

Appendix

Sec. 9. Proofs of Selected Statements

This section contains proofs of certain statements offered in the foregoing sections, these proofs either appear somewhat bulky for a popular text or expect a more extensive knowledge of the probability theory.

9.1. Proof of von Neumann's Method of Drawing a Random Variables (Sec. 4.3). The random point Γ is uniformly distributed in the rectangle $abcd$ (Fig. 27) whose area is $M_0 (b - a)$*. The probability for the point Γ to fall below the curve $y = p(x)$ and thus not to be rejected is equal to the ratio of the areas

$$\frac{\int_a^b p(x)\,dx}{M_0 (b-a)} = \frac{1}{M_0 (b-a)}.$$

The probability for the point Γ to fall below the curve $y = p(x)$ in the interval $a' < x < b'$ is also equal to the ratio of areas

$$\frac{\int_{a'}^{b'} p(x)\,dx}{M_0 (b-a)}.$$

Hence, the fraction of ξ values comprised within the interval (a', b') among all selected values of ξ, is equal to the ratio

$$\frac{\int_{a'}^{b'} p(x)\,dx}{M_0 (b-a)} : \frac{1}{M_0 (b-a)} = \int_{a'}^{b'} p(x)\,dx,$$

that is what is required to prove.

---

* Cf. Sec. 9.3.

9.2. Distribution Density of the Variable $\zeta' = a + \sigma\zeta$ (Sec. 4.4). It is assumed that the variable ζ is normal and

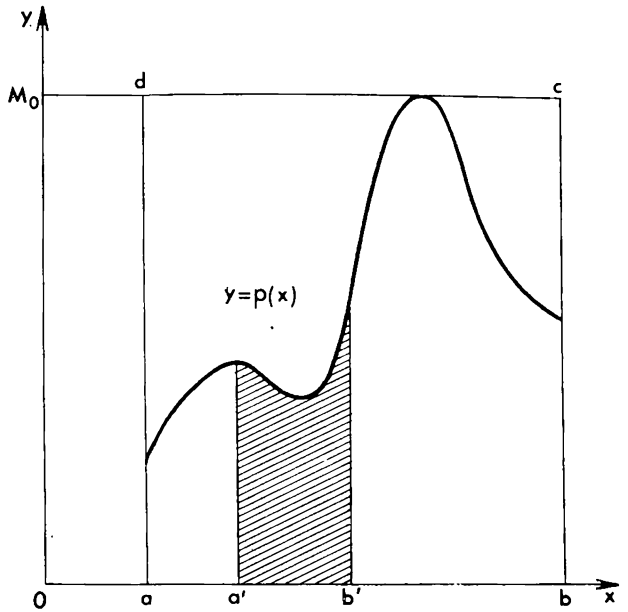

FIG. 27

has the expected value $\mathbf{M}\zeta = 0$ and the variance $\mathbf{D}\zeta = 1$, so that its density is

$$p_\zeta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

To calculate the distribution density of the variable ζ' we select two arbitrary numbers $x_1 < x_2$ and calculate the probability

$$\mathbf{P}\{x_1 < \zeta' < x_2\} = \mathbf{P}\{x_1 < a + \sigma\zeta < x_2\} = \\ = \mathbf{P}\left\{\frac{x_1 - a}{\sigma} < \zeta < \frac{x_2 - a}{\sigma}\right\}.$$

Hence

$$\mathbf{P}\{x_1 < \zeta' < x_2\} = \frac{1}{\sqrt{2\pi}} \int_{(x_1-a)/\sigma}^{(x_2-a)/\sigma} e^{-\frac{x^2}{2}}\, dx.$$

Let us change variable x for $x' = a + \sigma x$ in the last integral. We obtain

$$\mathbf{P}\{x_1 < \zeta' < x_2\} = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{x_1}^{x_2} e^{-\frac{(x'-a)^2}{2\sigma^2}}\, dx',$$

whence stems (cf. Eq. (14)) the normality of the variable ζ' with the parameters $\mathbf{M}\zeta' = a$, $\mathbf{D}\zeta' = \sigma^2$.

9.3. Uniform Distribution of Points (γ', γ'') Inside a Square (Sec. 4.5). Since the coordinates of the point (γ', γ'') are independent, its density $p(x, y)$ is equal to the product of densities

$$p(x, y) = p_{\gamma'}(x)\, p_{\gamma''}(y).$$

Each of these densities is identically equal to 1. Hence, $p(x, y) \equiv 1$ (for $0 < x < 1$ and $0 < y < 1$). This is exactly what we define as uniform distribution of the point (γ', γ'') in the unit square.

9.4. Selection of a Random Direction (Sec. 7.1). Let us specify the direction by a unit vector emerging from the origin of coordinates. The ends of these vectors lie on the surface of the unit sphere. The expression "any direction is equally probable" means that the end of the vector is a random point Ω uniformly distributed over the surface of this sphere: the probability that Ω will be found in any element of the surface dS is equal to $\frac{dS}{4\pi}$.

Let us choose spherical coordinates (φ, ψ) with the polar axis Ox on the surface of the sphere (Fig. 28). Then

$$dS = \sin\varphi \cdot d\varphi \cdot d\psi, \tag{36}$$

where $0 \leqslant \varphi \leqslant \pi$, $0 \leqslant \psi < 2\pi$.

Let us denote by $p(\varphi, \psi)$ the density of the random point (φ, ψ). It follows from Eq. (36) and from the requirement

$$p(\varphi, \psi)\, d\varphi \cdot d\psi = \frac{dS}{4\pi}$$

that

$$p(\varphi, \psi) = \frac{\sin\varphi}{4\pi}. \tag{37}$$

Densities of φ and ψ can be easily calculated from the joint density of these variables:

$$p_{\varphi}(\varphi) = \int_0^{2\pi} p(\varphi, \psi)\, d\psi = \frac{\sin \varphi}{2}; \tag{38}$$

$$p_{\psi}(\psi) = \int_0^{\pi} p(\varphi, \psi)\, d\varphi = \frac{1}{2\pi}. \tag{39}$$

The equality $p(\varphi, \psi) = p_{\varphi}(\varphi)\, p_{\psi}(\psi)$ demonstrates that φ and ψ are independent.

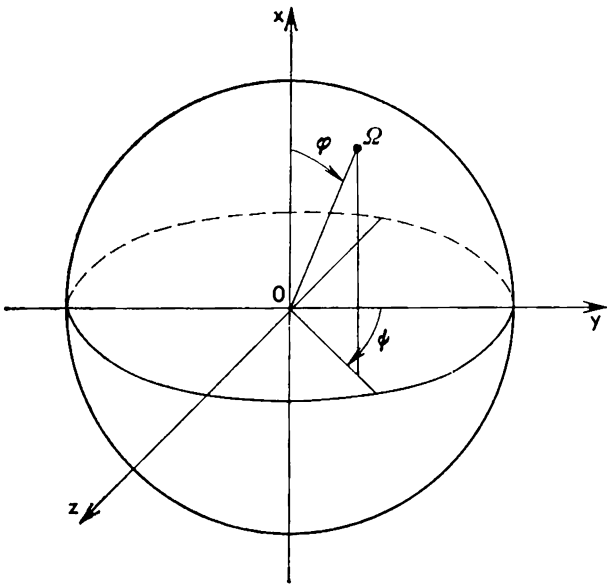

FIG. 28

It is obvious that ψ is uniformly distributed in the interval $(0, 2\pi)$ and the formula for drawing ψ should be written in the form

$$\psi = 2\pi\gamma. \tag{40}$$

The formula for drawing φ will be found by means of Eq. (23):

$$\frac{1}{2}\int_0^{\varphi} \sin x\, dx = \gamma,$$

whence

$$\cos \varphi = 1 - 2\gamma. \tag{41}$$

Eqs. (40) and (41) enable us to sample (draw) a random direction. Of course, the values of γ in these formulas must be different.

Equation (41) differs from the last formula in Sec. 7.1 only in that $1 - \gamma$ is replaced for γ, both variables having identical distributions.

9.5. Advantages of the Computation Method Using Weights (Sec. 7.3). Let us introduce random variables ν and ν' equal to the number (weight) of neutrons transmitted through the plate and obtained from one trajectory computed by the methods of Sec. 7.2 and 7.3, respectively.

It is physically obvious that

$$\mathbf{M}\nu = \mathbf{M}\nu' = p^+.$$

(This statement is rigorously proved in the author's monograph [3].)

Since ν can take on only two values, 0 or 1, the distribution of ν is specified by the table

$$\nu = \begin{pmatrix} 1 & 0 \\ p^+ & 1 - p^+ \end{pmatrix}.$$

Taking into account that $\nu^2 = \nu$, we easily find that $\mathbf{D}\nu = p^+ - (p^+)^2$.

It is clear that the variable ν' can take on an infinite set of values: $w_0 = 1, w_1, w_2, \ldots, w_k, \ldots$, as well as 0. Thus its distribution is specified by the table

$$\nu' = \begin{pmatrix} w_0 & w_1 & w_2 & \ldots & w_k & \ldots & 0 \\ q_0 & q_1 & q_2 & \ldots & q_k & \ldots & q \end{pmatrix}.$$

The values of q_i are of no interest for us because the variance can be written in all cases as

$$\mathbf{D}\nu' = \sum_{k=0}^{\infty} w_k^2 q_k - (p^+)^2.$$

We note that all $w_k \leqslant 1$ and that $\sum_{k=0}^{\infty} w_k q_k = \mathbf{M}\nu' = p^+$, and thus obtain the inequality $\mathbf{D}\nu' \leqslant p^+ - (p^+)^2 = \mathbf{D}\nu$.

The fact that the variance of ν' is always smaller than

that of ν demonstrates that the method of Sec. 7.3 is always better for computation of p^+ than that of Sec. 7.2.

The same conclusion is valid as regards the calculation of p^-, and, provided the absorption is not too high, as regards the calculation of p^0.

9.6. The Choice of the Best ξ (Sec. 8.2). An expression for the dispersion $\mathbf{D}\eta$ was obtained in Sec. 8.2. To find the minimum of this expression among all possible choices of $p_\xi(x)$ we will use a well known inequality

$$\left[\int_a^b |u(x)\,v(x)|\,dx\right]^2 \leqslant \int_a^b u^2(x)\,dx \cdot \int_a^b v^2(x)\,dx.$$

Let us set $u = g(x)/\sqrt{p_\xi(x)}$ and $v = \sqrt{p_\xi(x)}$; then this inequality yields

$$\left[\int_a^b |g(x)|\,dx\right]^2 \leqslant \int_a^b \frac{g^2(x)}{p_\xi(x)}\,dx \cdot \int_a^b p_\xi(x)\,dx = \int_a^b \frac{g^2(x)}{p_\xi(x)}\,dx.$$

Hence

$$\mathbf{D}\eta \geqslant \left[\int_a^b |g(x)|\,dx\right]^2 - I^2. \tag{42}$$

It now remains to prove that the lower boundary is taken on when $p_\xi(x)$ is proportional to $|g(x)|$.

Let

$$p_\xi(x) = \frac{|g(x)|}{\int_a^b |g(x)|\,dx}. \tag{43}$$

It can be easily calculated that for this density $p_\xi(x)$

$$\int_a^b \left[\frac{g^2(x)}{p_\xi(x)}\right] dx = \left[\int_a^b |g(x)|\,dx\right]^2,$$

and the variance $\mathbf{D}\eta$ is indeed equal to the right-hand side of the expression (42).

Note that it is practically impossible to select the "best" density (43) for these calculations: this would require that we know the value of the integral $\int_a^b |g(x)|\,dx$. However, the evaluation of the last integral is a problem equivalent to

that of evaluation of $\int_a^b g(x)\,dx$. Thus we have restricted ourselves to the suggestion given in Sec. 8.2.

9.7. Definition of the Probable Error (Sec. 8.4). Let ζ be the normal random variable defined in Sec. 2.3. It is easily calculated that whatever a and σ we have for $r = 0.675\sigma$

$$\int_{a-r}^{a+r} p_\zeta(x)\,dx = 0.5.$$

Hence

$$\mathbf{P}\{|\zeta - a| < r\} = \mathbf{P}\{|\zeta - a| > r\} = 0.5,$$

i.e. deviations exceeding r and those smaller than r are equally probable. Thus, the characteristic r is referred to as the *probable error* of the random variable ζ.

In Sec. 8.1 we calculate an approximately normal variable $\rho = (1/N)(\eta_1 + \eta_2 + \ldots + \eta_N)$. Its expected value is $a \equiv \mathbf{M}\rho = I$, and the variance is $\sigma^2 \equiv \mathbf{D}\rho = \mathbf{D}\eta/N$. Hence, the probable error of the variable ρ is approximately equal to $0.675\sqrt{\mathbf{D}\eta/N}$.

Sec. 10. On Pseudo-Random Numbers

Most of the algorithms for generation of pseudo-random numbers have the form

$$\gamma_{k+1} = F(\gamma_k). \tag{44}$$

If the starting number γ_0 is fixed, all the successive numbers $\gamma_1, \gamma_2, \ldots$ are calculated by the same formula (44) for $k = 0, 1, 2, \ldots$. Both algorithms considered in Sec. 3.3 also have form (44) although a set of operations, which are to be carried out over the argument x to obtain a value of y, was given instead of specifying the function $y = F(x)$ in an analytical form.

10.1. What Should Be the Function $F(x)$. The example below helps to elucidate the nature of one of the basic difficulties of selecting $F(x)$.

Example. Let us prove that the function $F(x)$ plotted in Fig. 29 cannot be used to generate pseudo-random numbers by means of formula (44).

Indeed, let us consider the points with Cartesian coordinates

$$(\gamma_1, \gamma_2), (\gamma_3, \gamma_4), (\gamma_5, \gamma_6), \ldots$$

inside the unit square $0 < x < 1$, $0 < y < 1$. Since here we have $\gamma_2 = F(\gamma_1)$, $\gamma_4 = F(\gamma_3)$, $\gamma_6 = F(\gamma_5)$, . . ., all these points are located on the curve $y = F(x)$. This is of course unacceptable because true random points must uniformly fill the whole square (see Sec. 9.3).

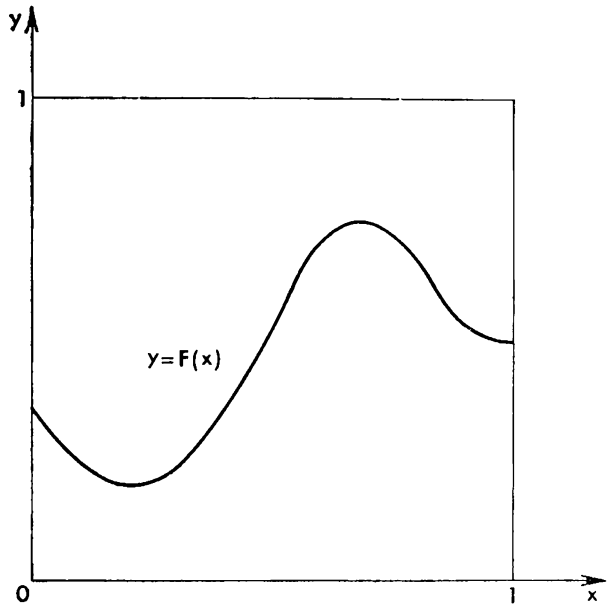

FIG. 29

The above example leads to the conclusion that the function $y = F(x)$ can be successfully used in formula (44) only if its graph provides sufficiently dense filling of the whole square!

An example of a function possessing this property is

$$y = \{gx\}, \tag{45}$$

where g is a very large number and $\{z\}$ is the fractional part of the number z. Function (45) is plotted in Fig. 30 for $g = 20$. The reader can imagine what does this plot look like at $g = 5^{17}$.

10.2. The Congruential Method (Method of Residues). The most widespread algorithm for generation of pseudo-

random numbers was suggested by D. Lehmer. This algorithm is based on function (45) but for the sake of convenient realization in computers the calculation formulae are written somewhat differently.

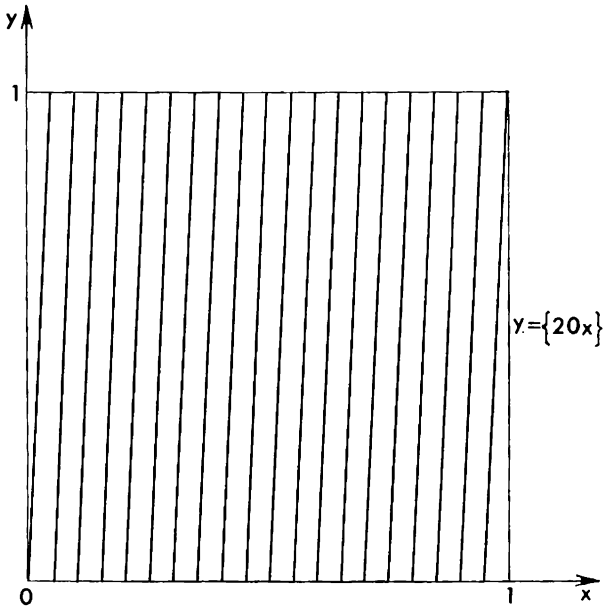

FIG. 30

A sequence of integer numbers m_k is generated in which the starting integer $m_0 = 1$ is fixed and all subsequent integers $m_1, m_2, \ldots$ are calculated by means of one formula

$$m_{k+1} \equiv 5^{17} m_k \pmod{2^{40}}, \tag{46}$$

for $k = 0, 1, 2, \ldots$; from the numbers m_k we calculate pseudo-random numbers

$$\gamma_k = 2^{-40} m_k. \tag{47}$$

Formula (46) means that the number m_{k+1} is equal to the residue obtained as a result of dividing $5^{17} m_k$ by 2^{40}. In the theory of congruence (see any textbook on the theory of numbers) this residue is referred to as the least positive residue modulo 2^{40}. Thus both terms for this algorithm—

congruential method and *method of residues*—originate from this theory.

Both formulas (46) and (47) are very easily realized in electronic computers, operating with 40-bit numbers, by means of the multiplication command with double precision: low-order digits of the product must be used*.

The period of the sequence m_k is equal to 2^{38}.

In computers operating with 36-bit numbers we use 5^{15} and 2^{36} instead of 5^{17} and 2^{40} in formulas (46) and (47); the period of the sequence m_k is in this case 2^{34}.

---

* The computer *Strela* (see Sec. 3.3) has no such a command.

RANDOM NUMBER TABLE

Table A. 400 random numbers *

86 515	90 795	66 155	66 434
69 186	03 393	42 502	99 224
41 686	42 163	85 181	38 967
86 522	47 171	88 059	89 342
72 587	93 000	89 688	78 416
52 452	42 499	33 346	83 935
76 773	97 526	27 256	66 447
04 825	82 134	80 317	75 120
87 113	84 778	45 863	24 520
84 754	57 616	38 132	64 294

Table B. 88 random normal variables **

0.2005	1.1922	—0.0077	0.0348
1.1609	—0.6690	—1.5893	0.5816
0.5864	—0.9245	0.0904	1.5068
0.1425	—0.2863	1.2809	0.4043
0.9516	—1.7708	2.8854	0.4686
—0.5863	0.8574	—0.5557	0.8115
1.1572	0.9990	—0.1032	0.5405
—1.4428	—0.5564	—0.5098	—1.1929
—0.3924	1.7981	0.6141	—1.3596
0.8319	0.4270	—0.8888	0.4167
0.9780	—0.7679	0.8960	0.5154

* Random numbers simulate the values of a random variable with
** Normal variables simulate the values of a normal (Gaussian) random

56 558	12 332	94 377	57 802
88 955	53 758	91 641	18 867
33 181	72 664	53 807	00 607
67 248	09 082	12 311	90 316
27 589	99 528	14 480	50 961
79 130	90 410	45 420	77 757
25 731	37 525	16 287	66 181
45 904	75 601	70 492	10 274
19 976	04 925	07 824	76 044
15 218	49 286	89 571	42 903

1.0423	−1.8149	1.1803	0.0033
1.8818	0.7390	−0.2736	1.0828
−1.1147	0.2776	0.1012	−1.3566
0.6379	−0.4428	−2.3006	−0.6446
1.4664	1.6852	−0.9690	−0.0831
−0.2676	−1.2496	−1.2125	1.3846
−0.6022	0.0093	0.2119	−1.4647
−0.0572	−0.5061	−0.1557	−1.2384
1.4943	−0.4406	−0.2033	−0.1316
−0.8513	1.1054	1.2237	−0.7003
0.7165	0.8563	−1.1630	1.8800

distribution (22) (see 3.1)
variable ζ with $a = 0$, $\sigma = 1$

BIBLIOGRAPHY

1. Buslenko, N. P., Golenko, D. I., Sobol, I. M., Sragovitch, V. G. Shreider, Yu. A. *The Method of Statistical Trials* (The Monte Carlo Method). Fizmatgiz, Mathematical Library Series, **1962.**
2. Yermakov, S. M. *The Monte Carlo Method and Relevant Problems*, "Nauka", **1971.**
3. Sobol, I. M. *Numerical Monte Carlo Methods*, "Nauka", **1973.**

TO THE READER

Mir Publishers welcome your comments on the content, translation and design of this book.

We would also be pleased to receive any proposals you care to make about our future publications.

Our address is:
USSR, 129820, Moscow I-110, GSP
Pervy Rizhsky Pereulok, 2
Mir Publishers

Printed in the Union of Soviet Socialist Republics

Other little mathematics library books:

DIVIDING A SEGMENT IN A GIVEN RATIO

N. M. Beskin

One of the series "Popular Lectures in Mathematics", widely used in the extracurricular maths clubs and circles of Soviet schools. Is broadly accessible to pupils doing "0" level mathematics. Describes various theories arrived at through deeper study of exercises in dividing a segment in a given ratio.
Through looking into this problem and related questions, the reader makes a short journey through mathematics, encountering on the way affine and projective geometry and group theory (generally without their labels).

THE METHOD OF MATHEMATICAL INDUCTION

I. S. Sominsky

One of the series "Popular Lectures in Mathematics", widely used in school extracurricular maths circles. Will interest sixth-formers doing "A" level mathematics, and first-year students whose courses include mathematics. Describes the method of mathematical induction so widely used in various fields of mathematics from elementary school courses to the most complicated investigations. Apart from being indispensible for study of mathematics, the ideas of induction have general educational value, and will interest a broad readership not familiar with mathematics.