



The contributions of Kolmogorov on the axiomatisation of the probability theory are fundamental, as Sundar's article-in-a-box shows, but his understanding of the applications of probability to practical problems in different fields was also deep; he could easily shift from abstract thinking to concrete problem formulations and finding solutions to them. In the first two sections of his article '*The Theory of Probability*', reproduced below, he discusses how to set up suitable statistical models to answer natural questions arising in studying the variations in the length of the life of an electric bulb and in finding the minimum number of hits of gunfire needed to hit a target with high probability.

S Natarajan, ISI, Bangalore

## THE THEORY OF PROBABILITY\*

### §1. The Laws of Probability

The simplest laws of natural science are those that state the conditions under which some event of interest to us will either certainly occur or certainly not occur; i.e., these conditions may be expressed in one of the following two forms:

1. If a complex (i.e., a set or collection) of conditions  $S$  is realized, then event  $A$  certainly occurs;
2. If a complex of conditions  $S$  is realized, then event  $A$  cannot occur.

In the first case the event  $A$ , with respect to the complex of conditions  $S$ , is called a "certain" or "necessary" event, and in the second an "impossible" event. For example, under atmospheric pressure and at temperature  $t$  between  $0^\circ$  and  $100^\circ$  (the complex of conditions  $S$ ) water necessarily occurs in the liquid state (the event  $A_1$  is certain) and cannot occur in a gaseous or solid state (events  $A_2$  and  $A_3$  are impossible).

An event  $A$ , which under a complex of conditions  $S$  sometimes occurs and sometimes does not occur, is called random with respect to the complex of conditions. This raises the question: Does the randomness of the event  $A$  demonstrate the absence of any law connecting the complex of conditions  $S$  and the event  $A$ ? For example, let it be established that lamps of a specific type, manufactured in a certain factory (condition  $S$ ) sometimes continue to burn more than 2,000 hours (event  $A$ ), but sometimes burn out and become useless before the expiration of that time. May it not still be possible that the results of experiments to see whether a given lamp will or will not burn for 2,000 hours will serve to evaluate

\* Reproduced from *Mathematics: Its contents, methods, and meaning*, edited by A D Aleksandrov, A N Kolmogorov, M A Lavrent'ev, Published by The MIT Press, 1963.



the production of the factory? Or should we restrict ourselves to indicating only the period (say 500 hours) for which in practice all lamps work without fail, and the period (say 10,000 hours) after which in practice all lamps do not work? It is clear that to describe the working life of a lamp by an inequality of the form  $500 \leq T \leq 10,000$  is of little help to the consumer. He will receive much more valuable information if we tell him that in approximately 80% of the cases the lamps work for no less than 2,000 hours. A still more complete evaluation of the quality of the lamps

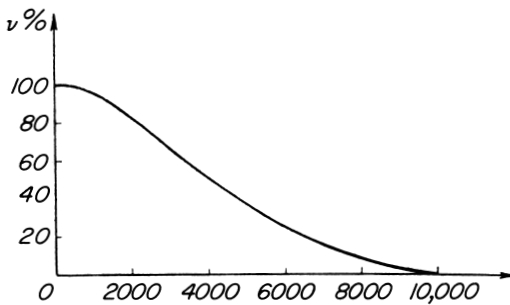


FIG. 1.

will consist of showing for any  $T$  the percent  $\nu(T)$  of the lamps which work for no less than  $T$  hours, say in the form of the graph in figure 1.

The curve  $\nu(T)$  is found in practice by testing with a sufficiently large sample (100-200) of the lamps. Of course, the curve found in such a manner is of real value only in those

where it truly represents an actual law governing not only the given sample but all the lamps manufactured with a given quality of material and under given technological conditions; that is, only if the same experiments conducted with another sample will give approximately the same results (i.e., the new curve  $\nu(T)$  will differ little from the curve derived from the first sample). In other words, the statistical law expressed by the curves  $\nu(T)$  for the various samples is only a reflection of the law of probability connecting the useful life of a lamp with the materials and the technological conditions of its manufacture.

This law of probability is given by a function  $P(T)$ , where  $P(T)$  is the probability that a single lamp (made under the given conditions) will burn no less than  $T$  hours.

The assertion that the event  $A$  occurs under conditions  $S$  with a definite probability

$$P(A/S) = p$$

amounts to saying that in a sufficiently long series of tests (i.e., realizations of the complex of conditions  $S$ ) the frequencies

$$\nu_r = \frac{\mu_r}{n_r}$$



of the occurrence of the event  $A$  (where  $n_r$  is the number of tests in the  $r$ th series, and  $\mu_r$  is the number of tests of this series for which event  $A$  occurs) will be approximately identical with one another and will be close to  $p$ .

The assumption of the existence of a constant  $p = \mathbf{P}(A/S)$  (objectively determined by the connection between the complex of conditions  $S$  and the event  $A$ ) such that the frequencies  $\nu$  get closer “generally speaking” to  $p$  as the number of tests increases, is well borne out in practice for a wide class of events. Events of this kind are usually called *random* or *stochastic*.

This example belongs to the laws of probability for mass production. The reality of such laws cannot be doubted, and they form the basis of important practical applications in statistical quality control. Of a similar kind are the laws of probability for the scattering of missiles, which are basic in the theory of gunfire. Since this is historically one of the earliest applications of the theory of probability to technical problems, we will return below to some simple problems in the theory of gunfire.

What was said about the “closeness” of the frequency  $\nu$  to the probability  $p$  for a large number  $n$  of tests is somewhat vague; we said nothing about how small the difference  $\nu - p$  may be for any  $n$ . The degree of closeness of  $\nu$  to  $p$  is estimated in §3. It is interesting to note that a certain indefiniteness in this question is quite unavoidable. The very statement itself that  $\nu$  and  $p$  are close to each other has only a probabilistic character, as becomes clear if we try to make the whole situation precise.

## §2. The Axioms and Basic Formulas of the Elementary Theory of Probability

Since it cannot be doubted that statistical laws are of great importance, we turn to the question of methods of studying them. First of all one thinks of the possibility of proceeding in a purely empirical way. Since a law of probability exhibits itself only in mass processes, it is natural to imagine that in order to discover the law we must conduct a mass experiment.

Such an idea, however, is only partly right. As soon as we have established certain laws of probability by experiment, we may proceed to deduce from them new laws of probability by logical means or by computation, under certain general assumptions. Before showing how this is done, we must enumerate certain basic definitions and formulas of the theory of probability.

From the representation of probability as the standard value of the frequency  $\nu = m/n$ , where  $0 \leq m \leq n$ , and thus  $0 \leq \nu \leq 1$ , it follows that



the probability  $P(A)$  of any event  $A$  must be assumed to lie between zero and one\*

$$0 \leq P(A) \leq 1. \quad (1)$$

Two events are said to be mutually exclusive if they cannot both occur (under the complex of conditions  $S$ ). For example, in throwing a die, the the occurrence of an even number of spots and of a three are mutually exclusive. An event  $A$  is called the *union* of events  $A_1$  and  $A_2$  if it consists of the occurrence of at least one of the events  $A_1, A_2$ . For example, in throwing a die, the event  $A$ , consisting of rolling 1, 2, or 3, is the union of the events  $A_1$  and  $A_2$ , where  $A_1$  consists of rolling 1 or 2 and  $A_2$  consists of rolling 2 or 3. It is easy to see that for the number of occurrences  $m_1, m_2$ , and  $m$  of two mutually exclusive events  $A_1$  and  $A_2$  and their union  $A = A_1 \cup A_2$ , we have the equation  $m = m_1 + m_2$ , or for the corresponding frequencies  $\nu = \nu_1 + \nu_2$ .

This leads naturally to the following axiom for the addition of probabilities:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2), \quad (2)$$

if the events  $A_1$  and  $A_2$  are mutually exclusive and  $A_1 \cup A_2$  denotes their union.

Further, for an event  $U$  which is certain, we naturally take

$$P(U) = 1. \quad (3)$$

The whole mathematical theory of probability is constructed on the basis of simple axioms of the type (1), (2), and (3). From the point of view of pure mathematics, *probability* is a numerical function of "events," with a number of properties determined by axioms. The properties of probability, expressed by formulas (1), (2), and (3), serve as a sufficient basis for the construction of what is called the elementary theory of probability, if we do not insist on including in the axiomatization the concepts of an event itself, the union of events, and their intersection, as defined later. For the beginner it is more useful to confine himself to an intuitive understanding of the terms "event" and "probability," but to realize that although the meaning of these terms in practical life cannot be completely formalized, still this fact does not affect the complete formal precision of an axiomatized, purely mathematical presentation of the theory of probability.

The union of any given number of events  $A_1, A_2, \dots, A_s$  is defined as the event  $A$  consisting of the occurrence of at least one of these events.

\* For brevity we now change  $P(A/S)$  to  $P(A)$ .



From the axiom of addition, we easily obtain for any number of pairwise mutually exclusive events  $A_1, A_2, \dots, A_s$  and their union  $A$ ,

$$P(A) = P(A_1) + P(A_2) + \dots + P(A_s)$$

(the so-called *theorem of the addition of probabilities*).

If the union of these events is an event that is certain (i.e., under the complex of conditions  $S$  one of the events  $A_k$  must occur), then

$$P(A_1) + P(A_2) + \dots + P(A_s) = 1.$$

In this case the system of events  $A_1, \dots, A_s$  is called a *complete system of events*.

We now consider two events  $A$ , and  $B$ , which, generally speaking, are not mutually exclusive. The event  $C$  is the intersection of the events  $A$  and  $B$ , written  $C = AB$ , if the event  $C$  consists of the occurrence of both  $A$  and  $B$ .\*

For example, if the event  $A$  consists of obtaining an even number in the throw of a die and  $B$  consists of obtaining a multiple of three, then the event  $C$  consists of obtaining a six.

In a large number  $n$  of repeated trials, let the event  $A$  occur  $m$  times and the event  $B$  occur  $l$  times, in  $k$  of which  $B$  occurs together with the event  $A$ . The quotient  $k/m$  is called the conditional frequency of the event  $B$  under the condition  $A$ . The frequencies  $k/m$ ,  $m/n$ , and  $k/n$  are connected by the formula

$$\frac{k}{m} = \frac{k}{n} \cdot \frac{m}{n}$$

which naturally gives rise to the following definition:

The conditional probability  $P(B/A)$  of the event  $B$  under the condition  $A$  is the quotient

$$P(B/A) = \frac{P(AB)}{P(A)}.$$

Here it is assumed, of course, that  $P(A) \neq 0$ .

If the events  $A$  and  $B$  are in no way essentially connected with each other, then it is natural to assume that event  $B$  will not appear more often, or less often, when  $A$  has occurred than when  $A$  has not occurred, i.e., that approximately  $k/m \sim l/n$  or

$$\frac{k}{n} = \frac{k}{m} \frac{m}{n} \sim \frac{l}{n} \frac{m}{n}.$$

\* Similarly, the intersection  $C$  of any number of events  $A_1, A_2, \dots, A_s$  consists of the occurrence of all the given events.



In this last approximate equation  $m/n = \nu_A$  is the frequency of the event  $A$ , and  $l/n = \nu_B$  is the frequency of the event  $B$  and finally  $k/n = \nu_{AB}$  is the frequency of the intersection of the events  $A$  and  $B$ .

We see that these frequencies are connected by the relation

$$\nu_{AB} \sim \nu_A \nu_B.$$

For the probabilities of the events  $A$ ,  $B$  and  $AB$ , it is therefore natural to accept the corresponding exact equation

$$\mathbf{P}(AB) = \mathbf{P}(A) \cdot \mathbf{P}(B). \tag{4}$$

Equation (4) serves to define the *independence* of two events  $A$  and  $B$ .

Similarly, we may define the independence of any number of events. Also, we may give a definition of the independence of any number of experiments, which means, roughly speaking, that the outcome of any part of the experiments do not depend on the outcome of the rest.\*

We now compute the probability  $P_k$  of precisely  $k$  occurrences of a certain event  $A$  in  $n$  independent tests, in each one of which the probability  $p$  of the occurrence of this event is the same. We denote by  $\bar{A}$  the event that event  $A$  does not occur. It is obvious that

$$\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A) = 1 - p.$$

From the definition of the independence of experiments it is easy to see that the probability of any specific sequence consisting of  $k$  occurrences of  $A$  and  $n - k$  nonoccurrences of  $A$  is equal to

$$p^k(1 - p)^{n-k}. \tag{5}$$

Thus, for example, for  $n = 5$  and  $k = 2$  the probability of getting the sequence  $A\bar{A}A\bar{A}A$  will be  $p(1 - p)p(1 - p)(1 - p) = p^2(1 - p)^3$ ,

By the theorem on the addition of probabilities,  $P_k$  will be equal to the sum of the probabilities of all sequences with  $k$  occurrences and  $n - k$  nonoccurrences of the event  $A$ , i.e.,  $P_k$  will be equal from (5) to the product of the number of such sequences by  $p^k(1 - p)^{n-k}$ . The number of such

---

\* A more exact meaning of *independent experiments* is the following. We divide the  $n$  experiments in any way into two groups and let the event  $A$  consist of the result that all the experiments of the first group have certain preassigned outcomes, and the event  $B$  that the experiments of the second group have preassigned outcomes. The experiments are called independent (as a collection) if for arbitrary decomposition into two groups and arbitrarily preassigned outcomes the events  $A$  and  $B$  are independent in the sense of (4).

We will return in §4 to a consideration of the objective meaning in the actual world of the independence of events.



sequences is obviously equal to the number of combinations of  $n$  things taken  $k$  at a time, since the  $k$  positive outcomes may occupy any  $k$  places in the sequence of  $n$  trials.

Finally we get

$$P_k = C_n^k p^k (1 - p)^{n-k} \quad (k = 0, 1, 2, \dots, n) \quad (6)$$

(which is called a binomial distribution).

In order to see how the definitions and formulas are applied, we consider an example that arises in the theory of gunfire.

Let five hits be sufficient for the destruction of the target. What interests us is the question whether we have the right to assume that 40 shots will insure the necessary five hits. A purely empirical solution of the problem would proceed as follows. For given dimensions of the target and for a given range, we carry out a large number (say 200) of firings, each consisting of 40 shots, and we determine how many of these firings produce at least five hits. If this result is achieved, for example, by 195 firings out of the 200, then the probability  $P$  is approximately equal to

$$P = \frac{195}{200} = 0.975.$$

If we proceed in this purely empirical way, we will use up 8,000 shells to solve a simple special problem. In practice, of course, no one proceeds in such a way. Instead, we begin the investigation by assuming that the scattering of the shells for a given range is independent of the size of the target. It turns out that the longitudinal and lateral deviations, from the mean point of landing of the shells, follow a law with respect to the frequency of deviations of various sizes that is illustrated in figure 2.

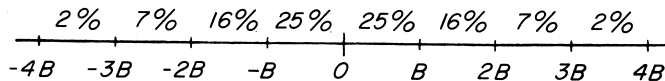


FIG. 2.

The letter  $B$  here denotes what is called the probable deviation. The probable deviation, generally speaking, is different for longitudinal and for lateral deviations and increases with increasing range. The probable deviations for different ranges for each type of gun and of shell are found empirically in firing practice on an artillery range. But the subsequent solution of all possible special problems of the kind described is carried out by calculations only.

For simplicity, we assume that the target has the form of a rectangle,



one side of which is directed along the line of fire and has a length of two probable longitudinal deviations, while the other side is perpendicular to the line of fire and is equal in length to two probable lateral deviations. We assume further that the range has already been well established, so that the mean trajectory of the shells passes through its center (figure 3).

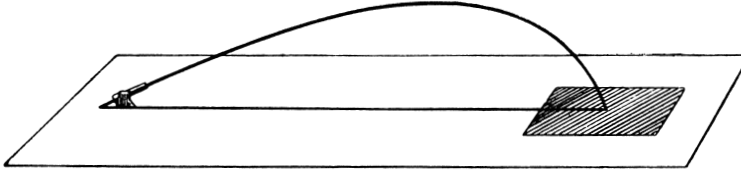


FIG. 3.

We also assume that the lateral and longitudinal deviations are independent.\* Then for a given shell to fall on the target, it is necessary and sufficient that its longitudinal and lateral deviations do not exceed the corresponding probable deviations. From figure 2 each of these events will be observed for about 50% of the shells fired, i.e., with probability  $\frac{1}{2}$ . The intersection of the two events will occur for about 25% of the shells fired; i.e., the probability that a specific shell will hit the target will be equal to

$$p = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4},$$

and the probability of a miss for a single shell will be

$$q = 1 - p = 1 - \frac{1}{4} = \frac{3}{4}.$$

Assuming that hits by the individual shells represent independent events, and applying the binomial formula (6), we find that the probability for getting exactly  $k$  hits in 40 shots will be

$$P_k = C_{40}^k p^k q^{40-k} = \frac{40 \cdot 39 \cdots (39 - k)}{1 \cdot 2 \cdots k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{40-k}.$$

What concerns us is the probability of getting no less than five hits, and this is now expressed by the formula

$$P = \sum_{k=5}^{40} P_k.$$

\* This assumption of independence is borne out by experience.



But it is simpler to compute this probability from the formula  $P = 1 - Q$ , where

$$Q = \sum_{k=0}^4 P_k$$

is the probability of getting less than five hits.

We may calculate that

$$P_0 = \left(\frac{3}{4}\right)^{40} \sim 0.00001,$$

$$P_1 = 40 \left(\frac{3}{4}\right)^{39} \frac{1}{4} \sim 0.00013,$$

$$P_2 = \frac{40 \cdot 39}{2} \left(\frac{3}{4}\right)^{38} \left(\frac{1}{4}\right)^2 \sim 0.00087,$$

$$P_3 = \frac{40 \cdot 39 \cdot 38}{2 \cdot 3} \left(\frac{3}{4}\right)^{37} \left(\frac{1}{4}\right)^3 \sim 0.0037,$$

$$P_4 = \frac{40 \cdot 39 \cdot 38 \cdot 37}{2 \cdot 3 \cdot 4} \left(\frac{3}{4}\right)^{36} \left(\frac{1}{4}\right)^4 \sim 0.0113,$$

so that

$$Q = 0.016, \quad P = 0.984.$$

The probability  $P$  so obtained is somewhat closer to certainty than is usually taken to be sufficient in the theory of gunfire. Most often it is considered permissible to determine the number of shells needed to guarantee the result with probability 0.95.

The previous example is somewhat schematized, but it shows in sufficient detail the practical importance of probability calculations. After establishing by experiment the dependence of the probable deviations on the range (for which we did not need to fire a large number of shells), we were then able to obtain, by simple calculation, the answers to questions of the most diverse kind. The situation is the same in all other domains where the collective influence of a large number of random factors leads to a statistical law. Direct examination of the mass of observations makes clear only the very simplest statistical laws; it uncovers only a few of the basic probabilities involved. But then, by means of the laws of the theory of probability, we use these simplest probabilities to compute the probabilities of more complicated occurrences and deduce the statistical laws that govern them.

Sometimes we succeed in completely avoiding massive statistical material, since the probabilities may be defined by sufficiently convincing

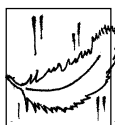


considerations of symmetry. For example, the traditional conclusion that a die, i.e., a cube made of a homogeneous material will fall, when thrown to a sufficient height, with equal probability on each of its faces was reached long before there was any systematic accumulation of data to verify it by observation. Systematic experiments of this kind have been carried out in the last three centuries, chiefly by authors of textbooks in the theory of probability, at a time when the theory of probability was already a well-developed science. The results of these experiments were satisfactory, but the question of extending them to analogous cases scarcely arouses interest. For example, as far as we know, no one has carried out sufficiently extensive experiments in tossing homogeneous dice with twelve sides. But there is no doubt that if we were to make 12,000 such tosses, the twelve-sided die would show each of its faces approximately a thousand times.

The basic probabilities derived from arguments of symmetry or homogeneity also play a large role in many serious scientific problems, for example in all problems of collision or near approach of molecules in random motion in a gas; another case where the successes have been equally great is the motion of stars in a galaxy. Of course, in these more delicate cases we prefer to check our theoretical assumptions by comparison with observation or experiment.

### Acknowledgements

*Resonance* thanks American Mathematical Society for the permission to reprint this article.



**The Calculus of Chance:** *"Interest in probability grew, encouraged by the researches of such eminent mathematicians as Leibniz, James Bernoulli, De Moivre, Euler, the Marquis de Condorcet, and above all, Laplace. The latter's epochal work on the analytic theory of probability brought the calculus to the point where Clerk Maxwell could say that it is "mathematics for practical men," and Jevons could wax quite lyrical (quoting without acknowledgement from Bishop Butler) that the mathematics of probability is "the very guide of life and hardly can we take a step or make a decision without correctly or incorrectly making an estimation of probability." And these opinions were uttered even before the calculus had achieved its most brilliant successes in physics and genetics as well as in more practical spheres. It was indeed remarkable, as Laplace wrote, that "a science which began with the considerations of play has risen to the most important objects of human knowledge."*

Edward Kasner and James R Newman, in *Mathematics and the Imagination*.

