CERTAINTY WITHIN UNCERTAINTY*

Louis H. Y. Chen
University of Singapore

Probability theory is a study of random phenomena, that is, phenomena whose outcomes cannot be predicted with certainty. Like many other branches of mathematics, it arose from attempts to solve physical problems. Its origin dates back to the seventeenth century. It is said that a French nobleman at that time was interested in several games played at Monte Carlo but tried unsuccessfully to derive mathematically the frequencies of bets that would be won. He was acquainted with two of the best mathematicians of that time, Fermat (1601 - 1665) and Pascal (1623 - 1662), and so he mentioned his difficulties to them. This began a famous exchange of letters between the two mathematicians concerning the applications of mathematics to games of chance. This exchange of letters is now considered as the beginning of probability theory as we know it today.

Although probability theory started more than three hundred years ago, a proper mathematical foundation of it was not laid until 1933 when a Russian mathematician, A. N. Kolmogorov, gave an axiomatic treatment of the subject. Since then great strides have been made in its development. Before the mathematical foundation was established, the notion of equal likelihood was used to calculate probabilities. If an experiment consists of a finite number of possible outcomes which are equally likely,

---

*This is the text of a public lecture organized by the *Society* delivered on October 22, 1975.

then the probability of an event occuring was taken to be the ratio of the number of outcomes favouring the event to the number of possible outcomes. This is known as the classical notion of probability. The French mathematician, Laplace (1749 - 1827), actually proposed to use this as a definition of probability. But as a definition it is logically inadequate, since equal likelihood depends on the notion of probability. However, from the technical point of view, it reduces probability to counting.

Let us see how probabilities can be calculated using the notion of equal likelihood. Suppose we roll a die which is balanced. Then there are six possible equally likely outcomes and the probability that 4 will show up is 1/6. Consider a slightly more complicated problem. Suppose we roll a balanced die n times and ask for the probability m 4's will show up. There are 6^n possible outcomes which are equally likely. Among these there are $\binom{n}{m}5^{n-m}$ outcomes with exactly m 4's. Therefore the probability that exactly m 4's show up is

$$\frac{\binom{n}{m}5^{n-m}}{6^n} = \binom{n}{m}\left(\frac{1}{6}\right)^m\left(\frac{5}{6}\right)^{n-m} .$$

Let us now consider the following problem. Suppose the probability of an event occuring in an experiment has been determined to be 1/6 and we wish to find the probability that the event occurs exactly m times in n repetitions of the experiment (or trials). To solve this problem, we can resort to the artifice of identifying the event with one face of a balanced die and proceed the argument as in the previous example. Of course we will end up with the same answer. This kind of argument can be extended to yield the following result: if the probability of an event occuring in an experiment is p where p is a fraction, then the probability that the event occurs exactly m times in n repetitions of the experiment is $\binom{n}{m}p^m(1 - p)^{n-m}$. Now we ask the next question. What happens if p is an irrational number, that is p is not a fraction ? For example $p = \sqrt{2}/3$. Such a situation can arise in real life problems. For example, consider a bomb being dropped randomly into an area of dimensions 50 metres by 50 metres. A particular target at the centre of the square is hit if the bomb falls within 10 metres of it. Then by applying an extension of equal likelihood, the probability that the target is hit is equal to $10^2\pi/50^2 = \pi/25$ which is an irrational number. In such a case we may identify the event with a face of a many-faced die and pass to an appropriate limit as the number of faces becomes infinitely large, since every irrational number has a sequence of fractions tending to it. This process may be awkward, but we will get the value of the probability which is the same as that for fractions. In fact, by the same process, the following more general result can be obtained: the probability that an event occurs $m_1, m_2, \ldots$ or m_k times in n trials, where the probability p of its occurance in each trial may or may not be a fraction, is

$$\sum_{i=1}^{k} \binom{n}{m_i} p^{m_i} (1-p)^{n-m_i} \quad . \quad (1)$$

The above method of calculating probabilities is only limited to a relatively small number of instances. There are far more other cases, paticularly those with infinitely many outcomes, to which this method is not applicable. This is where the axiomatic approach triumphs.

What is the axiomatic approach ? It is actually quite simple. We start with the set Ω of all possible outcomes and choose a suitable collection F of subsets of Ω. The set Ω is called the sample space and members of F are called events. A probability P is a nonnegative function defined on F such that it satisfies the following two axioms:

P1. If $A_1, A_2, \ldots$ are mutually exclusive events (i.e. no two of them have commom outcomes), then the probability of the event $\{A_1$ or A_2 or $\ldots\}$ is equal to the sum of the constituent events. Symbolically,

$P(A_1 \text{ or } A_2 \text{ or } \ldots) = P(A_1) + P(A_2) + \ldots$

P2. The whole sample space Ω is assigned (by convention) probability 1.

In applying this approach to problems involving repetitions of an experiment, one needs to introduce the notion of independence. Two events A and B are said to be independent if $P(A \text{ and } B) = P(A)P(B)$. This rule of multiplication is a mathematical formulation of independence (in the ordinary sense) of two events each of which has no influence on the occurance of the other. The notion of independence can be extended to more than two events: a given number of events are independent if the probability of the simultaneous occurance of any finite number of them is the product of the probabilities of the individual occurances of the corresponding events.

In the example of n repetitions of an experiment where the probability of a given event occuring in an experiment has been found to be a real number p, the sample space Ω can be taken to be the set of all n-tuples of 0's and 1's. We represent the occurance or non-occurance of the given event by "1" or "0" respectively. For example, if $n = 4$, then $(1,0,0,0)$ means that the event occurs only in the first trial. We then take F to be the collection of all subsets of Ω (there are 2^n of them). We now determine the probability P on F. The occurance of the events $A_i = \{(a_1,\ldots,a_n) : a_i = 1, a_k = 0 \text{ or } 1 \text{ for } k \neq i\}$ corresponds to the occurance of the given event in the ith trial. It is therefore natural to assume the following:

1. $P(A_i) = p$ for $i = 1,\ldots,n$
2. $A_1,\ldots,A_n$ are independent.

It can be shown that these two assumptions together with the axioms P1 and P2 uniquely determine the probability P. Simple calculations then show that the probability that the given event occurs $m_1, m_2, \ldots, m_k$ times in n trials is precisely the formula (1) given above.

There is yet another notion of probability - the relative frequency notion. Let k be the number of times an event occurs in n repetitions of an experiment. Since antiquity it has been observed that the ration k/n approaches a constant for very large n in many instances. This was particularly so in the field of demorgraphy where it had been found that the observed ratios of the number of girls to the number of births in large cities were all very close to 0.48. From the time of Fermat and Pascal, this statistical regularity has been observed in coin tossing, die rolling and many games of chance. In fact, there has been extensive experimental evidence to confirm this phenomenon.

It is therefore natural to take the relative frequency of the occurances of an event in a large number of trials to be an approximate value of the probability of the occurance of the event. This method of evaluating a probability is only empirical and may not be applicable in some instances. Nevertheless in cases where repetitions of an experiment can be performed the observed relative frequency has always been very close to the theoretical value of the probability calculated by using the classical notion of probability or the axiomatic approach. In the 1920's, the Austrian-born American mathematician R. von Mises, tried to axiomatize probability theory on the basis of relative frequency but his approach did not receive widespread acceptance among mathematicians.

Experiments have shown that the relative frequency over a large number of trials approaches the theoretical value of probability. The question now is whether this can be proved theoretically. The answer is yes. This was first proved by the Swiss mathematician Jacques Bernoulli (1654 - 1705). Let p be the probability of occurance of an event and k the number of its occurances in n trials. He proved that the probability that k/n differs from p by less than any preassigned number (no matter how small) approaches 1 as n increases indefinitely. Symbolically, for every $\varepsilon > 0$,

$$P\left(\left|\frac{k}{n} - p\right| > \varepsilon\right) \to 0 \quad \text{as } n \to \infty.$$

By the classical notion of probability, he showed that

$$P(|\frac{k}{n} - p| > \varepsilon) = \sum_{|\frac{k}{n}-p|>\varepsilon} \binom{n}{k} p^k (1-p)^{n-k},$$

where the summation is taken over those values of k for which $|\frac{k}{n} - p| > \varepsilon$. He then used Stirling's formula (1) to complete the proof. This result is known as the weak law of large numbers. Today its proof has been very much simplified and the law vastly generalized.

Two hundred years later, the French mathematician E. Borel proved (1909) the following astounding result: almost every number between 0 and 1 has asymptotically the same number of 0's and 1's in its binary expansion. Now every number t between 0 and 1 admits an expansion of the form

$$a_1/2 + a_2/2^2 + a_3/2^3 + \ldots$$

where each a_i is either 0 or 1. We write $t = 0.a_1a_2a_3\ldots$ and call it the binary expansion of t. For example, $5/8 = 1/2 + 0/2^2 + 1/2^3 + 0/2^4 + \ldots$ and therefore 5/8 is equal to 0.101000... Certain numbers have two distinct binary expansions. For example, 5/8 is also equal to $1/2 + 0/2^2 + 0/2^3 + 1/2^4 + 1/2^5 + \ldots$ so that 5/8 = 0.100111... Conversely every binary expansion $0.a_1a_2a_3\ldots$ corresponds to a number between 0 and 1. We can adopt a convention whereby every number corresponds to one and only one binary expansion, and vice versa. What Borel actually proved was that the set of numbers between 0 and 1 withthe property (in binary expansion)

$$\lim_{n\to\infty} (a_1 + a_2 + \ldots + a_n)/n = 1/2$$

has Lebesgue measure 1 (2) .

This result has a very significant probabilistic meaning. Consider repeating an experiment indefinitely where in each trial the probability of an event occuring is 1/2. Identify the occurance of the event by 1 and its non-occurance by 0. Then the sample space consists of all infinite sequences of 0's and 1's and therefore as observed earlier we can identify these sequences with the set of numbers between 0 and 1. The occurance of the event $A_i = \{0.a_1a_2a_3\ldots: a_i = 1,\ a_k = 0 \text{ or } 1 \text{ for } k \neq i\}$ corresponds to that of the given event in the ith trial. A probability P on any suitable collection of subsets of the set of numbers between 0 and 1 must be such that

1. $P(A_i) = \frac{1}{2}$ for all $i = 1,2,\ldots;$
2. $A_1, A_2, \ldots$ are independent.

It turns out that the conditions 1 and 2 uniquely determine P which equals the Lebesgue measure. Consequently, the event that the relative frequency of occurances of the given event approaches ½ has probability 1. This result is is known as the strong law of large numbers. Logically it implies the weak law of large numbers in the case $p = \frac{1}{2}$. Like the weak law, it has been vastly generalized and extended. In particular, one need not take the set of numbers between 0 and 1 as the sample space and the conclusion continues to hold if any number between 0 and 1, not necessarily ½, is taken.

In a very real sense, the weak and the strong laws of large numbers tell us that although the outcome of each repetition of an experiment cannot be predicted with certainty, the relative frequency of occurances of a particular event approaches with certainty the probability of its occurance as the number of trials increases indefinitely. Historically, these are the first two laws in probability theory which show that there is indeed *certainty within uncertainty*.

Nevertheless, these two laws are mathematical theorems. They do not in any way establish any physical truths. The important thing is that they agree with empirical results, thereby confirming the belief that probability theory is an adequate model for random phenomena.

What applications do these laws have ? An important application is that they provide a theoretical basis for the solution of deterministic problems by the Monte Carlo method. The Monte Carlo method is a method of solving numerical problems by means of the construction of a random process. Historically, the first example of the method is Buffon's celebrated problem of needle-tossing in 1777, whereby he calculated the value of π. But the name 'Monte Carlo' and its systematic development only dates back to the 1940's. The method was popularized by Fermi, Metropolis, von Neumann and Ulam.

Let us see what role the laws of large numbers play in Buffon's needle problem. A plane is partitioned by parallel lines separated by a distance of $2a$. The probability that a needle of length $2l$ ($l < a$) thrown at random onto the plane will intersect a line can be calculated to be $2l/a\pi$. If a needle is thrown n times and it intersects a line k times, then by the laws of large numbers, k/n approaches $2l/a\pi$ with high probability for large n. Then the value of π can be approximated by $2ln/ak$ for large n. Since the time of Buffon, a large number of needle throwing experiments have been carried out. The following are some of the recorded results.

Experimenter	Year	Number of Throws	Experimental Value of π
Wolf	1850	5000	3.1596
Smith	1855	3204	3.1553
Fox	1894	1120	3.1419
Lazzarini	1901	3408	3.1415929

It is possible to calculate the number of needles required in order to obtain the desired degree of accuracy. A formula for this is

$$P(|\frac{k}{n} - p| \leq \varepsilon) \geq 1 - 1/4n\varepsilon^2$$

where p in this case is $2l/a\pi$. But the value of n calculated by this formula is usually much larger than necessary. A more accurate formula for n can be derived using the normal approximation to the binomial distribution.

Another example where the Monte Carlo method finds application is the evaluation of integrals. Suppose we wish to evaluate the integral $\int_0^1 f(x)dx$ where $0 \leq f(x) \leq 1$. We use the computer to generate n pairs $(x_1,y_1),\ldots,(x_n,y_n)$ of random numbers between o and 1. The value of the integral is precisely the probability that each pair of the random numbers, which represents a point in the plane, will fall below the curve $y = f(x)$. Let k be the number of pairs (x_i,y_i) which fall below the curve $y = f(x)$, i.e. for which $y_i \leq f(x_i)$. By the laws of large numbers k/n is almost $\int_0^1 f(x)dx$ with high probability for large n and so k/n can be taken as an approximate value of the integral.

As an example, consider the integral $\int_0^1 x^2dx$. A total of three thousand pairs of random numbers between 0 and 1 were generated using the IBM 1130. The values of k/n corresponding to different values of n are tabulated below:

n	k/n
100	0.36000
200	0.36500
300	0.36333
400	0.35500
500	0.35600
600	0.35833
700	0.35285
1000	0.35000
1200	0.34333
1300	0.33538
1500	0.33466
1600	0.33500
1700	0.34117
2000	0.34100
2200	0.34090
2500	0.33920
2700	0.33814
2900	0.33655
3000	0.33900

Note that for large n the experimental value of the integral is reasonably close to 1/3 which is the theoretical value.

The laws of large numbers tell us that certain random observations must eventually lead to a constant. This may be the reason for the existence of many physical constants. If one pushes this argument further, even the deterministic laws of nature may be the consequences of the laws of large numbers. J. Bernoulli went even further. He wrote in his book *Ars Conjectandi*:

> "If all events from now through eternity were continually observed (whereby probability would ultimately become certainty), it would be found that everything in the world occurs for definite reasons and in definite conformity with law, and that hence we are constrained, even for things that may seem quite accidental, to assume a certain necessity and, as it were, fatefulness. For all I know that is what Plato had in mind when, in the doctrine of the universal cycle, he maintained that after the passage of countless centuries everything would return to its original state."

Notes

(1) Stirling's formula states that

$$n! \sim \sqrt{(2\pi)}\, n^{n+\frac{1}{2}} e^{-n}$$

where the symbol $\sim$ means that as n gets larger and larger, the ratio of the numbers on both sides of the symbol becomes closer and closer to 1.

(2) The Lebesgue measure was introduced by the twentieth century French mathematician Henri Lebesgue. It is an extension of the ordinary notion of length. For example the set of all numbers between 0 and 1 has Lebesgue measure equal to 1 (intuitively it is of length 1). The set of all fractions between 0 and 1 has Lebesgue measure 0 while the set of all irrational numbers between 0 and 1 has Lebesgue measure 1 (so that there are 'many more' irrational numbers than fractions).

Louis Chen Hsiao-Yun is a lecturer in mathematics at the University of Singapore and Honorary Secretary of the Singapore Mathematical Society. He graduated from the University of Singapore and received his Ph.D from Stanford University (U.S.A.). He has published research papers in the Annals of Probability *(U.S.A.) and the* Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete *(Germany). - Editor.*

* * * * * * * * * * * * *

A QUERY

Mr. A.D. Villanveva has sent in a query on the following geometrical problem.

Let ABCD be a quadrilateral, and let E and F be the points of intersection of the opposite sides produced. The segments AC, BD and EF are called the diagonals of the quadrilateral ABCD. Suppose we are given the lengths of the 3 diagonals of a quadrilateral which can be inscribed in a circle (i.e. concyclic). The problem is to construct the quadrilateral using compass and ruler only. Is the solution unique? Is the circumscribing circle unique? Is there a known relation between the lengths of the 3 diagonals?

Any reader who can help Mr. Villanveva is requested to write to the Editor.