INDIAN STATISTICAL INSTITUTE

TWENTY-FIRST CONVOCATION ADDRESS

Uncertainty, Randomness and Creativity

by

C. R. Rao

National Professor

5th March 1987

203 BARRACKPORE TRUNK ROAD

CALCUTTA 700 035

Mr. President, Mr. Chairman, Professor J. K. Ghosh, students and disguished guests,

It is the general tradition to invite a person not associated with the concerned organization to deliver the convocation address. The Director of the ISI has set aside this tradition in asking some one who had the longest association with the Institute and who is still connected with it to give the address. I recognise and appreciate the honour done to me.

To begin with, I would like to congratulate all those who received the degrees today, on the twenty first convocation of the Institute. I am sure you had the best of education and training available anywhere in the world, and there is a good demand for persons with your qualifications and expertize. You will be embarking on your career in a technologically fast changing world and your present knowledge can only ensure you a good start but not sustain you for a long period of time unless you constantly improve your knowledge and skills. I hope you would look for opportunities to do so by keeping in touch with your *alma mater* if necessary.

It has been customary with the ISI to invite only scientists to deliver the convocation address. This places an additional responsibility on me to give a scholarly address to be in line with the previous speakers. I hope you will bear with me if I attempt to do so.

Statistics—A Definition

There are several definitions of and misconceptions about statistics. The most appropriate definition, which has given statistics the status of a separate scientific discipline is that it deals with uncertainties and their quantification. We face with uncertainties every moment in the physical and social environment in which we live, we bear with uncertainties of nature and we suffer from catostrophies. Things are not deterministic as Gothe wished : 'Great, eternal, unchangeable laws prescribe the paths along which we all wander', or as Einstein, the greatest physicist in three centuries or possibly of all time, believed that 'God does not play dice with the universe'. Some theologians argue that nothing is random to God because He causes all to happen ; others say that even God is at the mercy of some random events. In his book on *The Garden of Epicurus*, Anatole France remarks, 'Chance is perhaps the pseudonym of God when He did not want to sign'.

Quantification of uncertainty is perhaps the greatest invention of the twentieth century which led to wise decision making under uncertainty and improved quality of life for the mankind. What are the tools employed in the study of uncertainty ? No doubt mathematics plays an improtant part. Strangely enough, the methodology for exploring uncertainty is based on a meaningless string of numbers called random numbers about which I would like to speak to you today.

Random Numbers

In 1927, a statistician by name L. H. C. Tippett produced a book entitled, *Random Sampling Numbers*. The contents of this book are 41,600 digits (from 0 to 9) arranged in sets of 4 in several columns and spread over 26 pages. It is said that the author took the figures of areas of parishes given in the British census returns, omitted the first two and last two digits in each figure of area and placed the truncated numbers one after the other in a somewhat mixed way till 41,600 digits are obtained. This book which is a haphazard collection of numbers became the best seller among technical books. A reproduction of a typical page of this book is given in an appendix to this article. This was followed by another publication by two great pioneering statisticians, R. A. Fisher and F. Yates, which contained 15000 digits formed by listing the 15–19th digits in some 20 figure logarithm tables.

A book of random numbers ! A meaningless and haphazard collection of numbers, neither fact nor fiction, of what earthly use is it ? This would have been the reaction of scientists and laymen in any earlier century. But a book of random numbers is typically a twentieth century invention arising out of the need for random numbers in solving problems of the real world. Now the production of random numbers is a multibillion dollar industry around the world involving considerable research and spohisticated high speed computers.

What is a sequence of random numbers ? It is difficult to say what it is, but easy to demonstrate how useful it is. There is no simple definition of a sequence of random numbers except that it does not follow any particular pattern. More technically, knowing all the numbers in the sequence up to any particular stage, it is impossible to predict the next number with any degree of success better than by any method of guessing. How does one generate such an ideal sequence of numbers ? For instance, you may toss a coin a number of times and record the sequence of 0's (for tails) and 1's (for heads) such as the following

$$0\,1\,1\,0\,1\,0\cdots\cdot$$

If you are not a magician who can exercise some control on tossing of coins, you get a random sequence of what are called binary digits (0's and 1's). When I was teaching the first year class at the Institute, I used to send my students to the Bon-Hooghly Hospital to get a record of successive male and female births delivered. Writing 1 for a male birth and 0 for a female birth we get a binary sequence as the one obtained above by repeatedly tossing a coin. One is a natural sequence of a biological phenomena and another is an artificially generated one. There was a remarkable similarity between the two indicating that the sex of a child is determined by a mechanism similar to that of coin tossing. God is tossing a coin ! In fact statistical tests showed that the male-female births provide a more faithful random binary sequence than the artificially generated one. Perhaps God is throwing a more perfect coin. In India one child is born every second, which is a cheap and expeditious source for generating random binary sequences.

In practice, besides computers, natural devices like the reverse-biased diode are used to generate random numbers based on the theory of quantum mechanics which postulates the randomness of certain events at the atomic level. Note that the theory itself is verifiable by comparing the numbers so observed with sequences generated through artificial divices.

We have already seen how artificially generated random sequences of numbers enable us to discover, by comparison, similar chance mechanisms in nature and explain the occurence of natural events (in biology and physics). There are a number of ways of exploiting randomness to make inroads on baffling questions, to solve problems that are too complex for an exact solution, to generate new information and also perhaps to help in creativity or evolving new ideas. I shall briefly describe some of them.

Monte-Carlo Technique

Karl Pearson, the British mathematician and one of the early contributors to statistical theory and methods, was the first to perceive the use of random numbers for solving problems in probability and statistics that are too complex for exact solution. If you know the joint distribution of a number of variables, say x_1, x_2, ..., x_n, how can we find the dsitribution of a given function $f(x_1, ..., x_n)$? The problem has a formal solution in the form of an incomplete multiple integral, but the computation is difficult. He found random numbers useful in obtaining at least an approximate solution to such problems and encourgaged L. H. C. Tippett to prepare a table of random numbers to help others in such studies. This method called *simulation* or *Monte-Carlo* technique has now become a standard device in statistics and all sciences to solve complicated numerical problems.

The basic principle of the simulation method is simple. Suppose that we wish to know what proportion of the area in the inside square of the ISI emblem is taken up by the picture of the banyan tree. You see the picture is complicated and there is no easy way of using a planimeter to measure the area. Now, let us take the square emblem and consider any two intersecting sides as the x and y axes. Choose a pair of random numbers (x, y) both in the range $(0, b)$, where b is greater than the length of the side of the square, and plot the point associated with (x, y) on the square emblem. Repeat the process a number of times and suppose that at some stage, a_m is the number of points that have fallen within the picture of the banyan tree and m is the total number of points that have fallen within the square. There is a theorem, called the law of large numbers, by the famous Russian probabilist A. N. Kolmogorov, which says that the ratio a_m/m tends to the true proportion of the area of the picture as m becomes large provided the pairs (x, y) chosen to locate the points are *truly* random. The success (or precision) of this method then depends on how faithful the random number generator is and how many we can produce subject to given resources.

Under the leadership of Karl Pearson the method was used by some of his students to find the distribution of some complicated sample statistics, but it did not catch up immediately except perhaps in India at the ISI where Professor P. C. Mahalonobis exploited the simulation technique, which he called random sampling experiments, to solve a variety of problems like the choice of the optimum sampling plans in survey work and optimum size and shape of plots in experimental work. The reason for delay in recognizing the potentialities of this method may be attributed to non-availability of devices to produce truly random numbers and in the requisite quantity both of which affect the precision of results. Also, in the absence of standard devices to generate random numbers, the editors of journals were reluctant to publish papers reporting simulation results. Now the situation is completely changed with the advent of reliable random number generators and easy access to them. We are able to undertake the investigation of complex problems and give at least approximate solutions for practical use. The editors of journals insist that every article submitted should report simulation results even when exact solutions are available. As a matter of fact, the whole character of research in statistics, perhaps in other fields too, is gradually changing with greater emphasis on what are called "number crunching methods", of which a typical example is the "bootstrap method" advocated by Efron, which has become very popular. You make the random numbers work for you. In India, for some unfortunate reasons, we have not been able to exploit the computer based simulation methods in research work and lost the opportunity of being in the forefront of statistical research in the world as we used to be before the advent of computers.

Sample Surveys

The next and perhaps the most important use of random numbers is in generating data in sample surveys and in experimental work. Consider a large population of individuals whose average income we wish to know. A complete enumeration or obtaining information from each individual and processing the data is not only time consuming and expensive but also undersirable due to organizational difficulties in getting accurate data. On the other hand, collection of data on a small proportion of individuals (a sample of individuals) can be obtained more expeditiously and under controlled conditions to ensure accuracy of data. Then the question arises as to how the sample of individuals should be chosen to provide data from which a valid and fairly accurate estimate of the average income could be obtained. One answer is the simple lottery method using random numbers. We first number all individuals, say 1, 2, ..., 1000, if there are one thousand individuals, choose the requisite number of random numbers in the range 1—1000 and select the individuals corresponding to these numbers. This is called a simple random sample of individuals. Again, statistical theory tells us that the average of the incomes of the individuals in a random sample tends to the true value as the sample size increases. In practice, the sample size can be determined to ensure a given margin of accuracy. Thus, random numbers enable us to acquire sample data which contain the desired information and to process the data in a specified way for making estimates of unknown parameters.

Design of Experiments

Randomization is an important aspect of scientific experiments such as those designed to test whether drug A is better than drug B for treating a certain disease or to decide which is a better yielding variety of rice among a given set of varieties. The object of these experiments is to generate data which provide valid comparisons of the treatments under consideration. R. A. Fisher, the statistician who initiated the subject of design of experiments, showed that by allocating individuals at random to the two drugs A and B in the medical experiment and assigning the varieties to the experimental plots at random in the agricultural experiment we can generate valid data for comparison of treatments. This has, indeed, baffled some statisticians who have challenged the randomization principle, but no alternative methodology without making extraneous assumptions has been proposed.

The early applications of random numbers for solving statistical problems paved the way for the use of random numbers in model building, prediction and estimating the margin of uncertainty in prediction. Some of the areas where such models are developed are for forecasting weather, demand for consumer goods, future needs of the society in terms of services like housing, schools, hospitals and transport facilities and so on.

Encryption of Messages

Some of the modern uses of random numbers which opened up a large demand for random number generators are in solving such complicated problems as the travelling salesman problem involving the determination of a minimum path connecting a number of given places to be visited starting from a given place and returning to the same place. Random numbers in large quantities are also required in cryptology or the secret coding of messages for transmission and in maintaining the secrecy of individual bank transactions.

Top-level diplomatic and military communications, where secrecy is is extremely important, are encrypted in such a way that any one illegally tapping the transmission lines can get only a random looking sequence of numbers. To achieve this, first a string of binary random digits called the key string is generated which is known only to the sender and the receiver but nobody else. The sender converts his message into a string of binary digits in the usual way by converting each character into its standard eight bit computer code (the letter a for example is 0110 0001). He then places the message string below the key string and obtains a coded string by changing every message bit to its alternative at all places where the key bit is 1 and leaving the others unchanged. The coded string which appears to be a random binary sequence is transmitted. The received message is decoded by making the changes in the same way as in encrypting using the key string which is known to the receiver. Here is an example.

Key	0	1	0	0	0	1	1	Sender
Message	1	0	1	1	1	0	0	
Encrypted message	1	1	1	1	1	1	1	Transmitted message
Key	0	1	0	0	0	1	1	Receiver
Recovered message	1	0	1	1	1	0	0	

Banks use secret codes based on random numbers to guarantee the privacy of transactions made by automatic teller machines. For this purpose a random number is generated as a key with a rule of coverting a message into a code which is decipherable only with the knowledge of the key. Later after the key is given to both the central computer and the teller machine, the two devices communicate freely by telephone without fear of evesdroppers. After receiving the message from the teller machine regarding the customer's number and the amount he wants to withdraw, the central computer verifies the customer's account and instructs the teller machine to make or not make the payment.

Gambler's Fallacy

It is an interesting property of random numbers that, like the Hindu concept of God, it is patternless and yet has all the patterns in it. That is, if we go on generating strictly random numbers we will encounter any given pattern sometime or other. Thus, if we go on tossing a coin, we should not be surprised if 1000 heads appear in successive tosses at some stage. So we have the proverbial monkey which if allowed to type continuously can produce the entire works of Shakespeare in a finite though a long period of time. Non-realization of this has led many to two common mistakes. One may be called the parents' blunder (usually known as gambler's fallacy) by which a couple expects to have a boy after having three girls in succession. Such false expectations may lead to undesirable consequences. Another is the hasty conclusion drawn about natural phenomena by encountering a pattern in a sequence of events. Professor Narlikar, in his sixteenth convocation address, has referred to a controversy between Fred Hoyle and Martin Ryle arising out of such an error. Professor Narlikar mentioned that his simulation or Monte-Carlo experiments showed that a steady and homogeneous system can exhibit local inhomogeneities, and Ryle's observations of such inhomogeneities in the density of radio sources does not contradict Hoyle's steady state theory of the universe.

Let me give you another example. It is found that population sizes of a large variety of animals exhibit roughly a three year cycle, i.e., the average time that elapses between two sucessive peak years of population size is about 3 years. (A peak year is defined as a year in which there are more animals than in the immediately preceding and immediately succeeding years). The ubiquity of such phenomenon led some to believe that perhaps a new law of nature has been uncovered. The belief was dealt a moral blow when it was noted that if one plots random numbers at equidistant points, the average distance between peaks approaches 3 as the series of numbers gets large. In fact, such a property is easily demonstrable by using the fact that the probability of the middle number being larger than the others in a set of 3 random numbers is 1/3. This gives an average distance of 3 years between the peaks.

Statistical Physics

For a long time it was believed that all natural events have an unambiguously predetermined character which implied that occurence of particular events could be predicted with certainty given certain specified premises. It was realized during the middle of the last century that the quest for such deterministic laws of nature is strewn with both logical and practical difficulties and attempts should be made to explain natural phenomena through appropriate chance mechanisms.

Theree major developments took place about the same time in three different fields of enquiry. They are all based on the premises that chance is inherent in nature. Adolph Quetelet (1869) used concepts of probability in describing social and biological phenomena. Gregor Mendel (1870) formulated his laws of heredity through chance mechanisms like rolling a die. Boltzman (1866) gave a statistical interpretation to one of the most famous propositions of theoretical physics, the second law of thermodynamics. The ideas propounded by these stalwarts were revolutionary in nature. Although they were not accepted immediately, rapid advances took place in all these areas using statistical concepts during the present century.

The contributions to statistical physics are particularly important since they led to a revolutionary change in our whole conception of the universe. Reference may be made to the explanation of Brownian motion and the scintillations caused by radio-activity, Heisenberg's uncertainty principle, Maxwell's velocity distributions for molecules of equal mass, all of which blazed the trail for quantum mechanics of the present day. The concept of statistical laws displacing the deterministic laws did not find favour with Einstein who said that 'God does not play dice with the universe'. He maintained even towards the end of his life :

> "But I am quite convinced that some one will eventually come up with a theory, whose objects, connected by laws, are not probabilities but considered facts, as was until recently taken for granted."

Randomness and Creativity

We have seen how randomness is inherent in nature requiring natural laws to be formulated in statistical terms. We have also seen how in sample surveys and experimental work, randomization helps in generating data with the desired information and in extracting the information. We have have also seen how randomness can be exploited to solve complicated problems like the travelling salesman problem and in maintaining the secrecy of communications during transmission. Can randomness help in developing new ideas, or what is its relationship to creativity ?

What is creativity ? At least in the realm of science, it is the birth of a new idea which is qualitatively different and not deducible from existing ideas and which provides a key for explaining a wide variety of natural phenomena. The observations we get from nature appear to us like random numbers, as an encrypted message I talked about earlier in my talk. Creativity is the ability to discover the key to decode and read the underlying message. Is creativity mechanizable ? We have no answer at present as all attempts to do so have not yet succeeded even in a small measure. It is believed that creativity results from random thinking by allowing the human mind to wander unfettered by the rigidities of accepted knowledge or conventional rules. Writing on this subject, Hofstadter says :

> "It is a common notion that randomness is an indispensable ingredient of creative arts. This may be true, but it does not have any bearing on the mechanizability—or rather programmability ! — of creativity. ..... Randomness is an intrinsic feature of human thought, not something which has to be artificially inseminated, whether through dice, decaying nuclei, random number tables, or what-have-you. It is an insult to human creativity to imply that it relies on arbitrary sources".

Perhaps, random thinking is only one ingredient of creativity. Other elements are required such as the preparedness of the mind, ability to identify important and significant problems, quick perception of what ideas can lead to fruitful results and above all certain amount of confidence to pursue difficult problems. The last aspect is what is lacking in the bulk of scientific research today, which Einstein emphasized :

> "I have little patience with scientists who take a board of wood, look for the thinnest part and drill a great number of holes where drilling is easy"

I leave these thoughts with you, which may be relevant in the context of current discussions on the quality of scientific research in India and how it can be improved.

Part of page xiv of Tippett's Random Sample Numbers

1 2 3 4	5 6 7 8	9 10 11 12	13 14 15 16	17 18 19 20	21 22 23 24	25 26 27 28	29 30 31 32
7 8 1 6	6 5 7 2	0 8 0 2	6 3 1 4	0 7 0 2	4 3 6 9	9 7 2 8	0 1 9 8
3 2 0 4	9 2 4 3	4 9 3 5	8 2 0 0	3 6 2 3	4 8 6 9	6 9 3 8	7 4 8 1
2 9 7 6	3 4 1 3	2 8 4 1	4 2 4 1	2 4 2 4	1 9 8 5	9 3 1 3	2 3 2 2
8 3 0 3	9 8 2 2	5 8 8 8	2 4 1 0	1 1 5 8	2 7 2 9	6 4 4 3	2 9 4 3
5 5 5 6	8 5 2 6	6 1 6 6	8 2 3 1	2 4 3 8	8 4 5 5	4 6 1 8	4 4 4 5
2 6 3 5	7 9 0 0	3 3 7 0	9 1 6 0	1 6 2 0	3 8 8 2	7 7 5 7	4 9 5 0
3 2 1 1	4 9 1 9	7 3 0 6	4 9 1 6	7 6 7 7	8 7 3 3	9 9 7 4	6 7 3 2
2 7 4 8	6 1 9 8	7 1 6 4	4 1 4 8	7 0 8 6	2 8 8 8	8 5 1 9	1 6 2 0
7 4 7 7	0 1 1 1	1 6 3 0	2 4 0 4	2 9 7 9	7 9 9 1	9 6 8 3	5 1 2 5
5 3 7 9	7 0 7 6	2 6 9 4	2 9 2 7	4 3 9 9	5 5 1 9	8 1 0 6	8 5 0 1
9 2 6 4	4 6 0 7	2 0 1 2	3 9 2 0	7 7 6 6	3 8 1 7	3 2 5 6	1 6 4 0
5 8 5 8	7 7 6 6	3 1 7 0	0 5 0 0	2 5 9 3	0 5 4 5	5 3 7 0	7 8 1 4
2 8 8 9	6 6 2 8	6 7 5 7	8 2 3 1	1 5 8 9	0 0 6 2	0 0 4 7	3 8 1 5
5 1 3 1	8 1 8 6	3 7 0 9	4 5 2 1	6 6 6 5	5 3 2 5	5 3 8 3	2 7 0 2
9 0 5 5	7 1 9 6	2 1 7 2	3 2 6 7	1 1 1 4	1 3 8 4	4 3 5 9	4 4 8 8
7 9 0 0	5 8 7 0	2 6 0 6	8 8 1 3	5 5 0 9	4 3 2 4	0 0 3 0	4 7 5 0
3 6 9 3	9 2 1 2	0 5 5 7	7 3 6 9	7 1 6 2	9 5 6 8	1 3 1 2	9 4 3 8
0 3 8 0	3 3 3 8	0 1 3 8	4 5 6 0	4 2 0 3	6 4 9 6	3 8 0 6	0 3 4 7
0 2 4 6	4 4 6 9	9 7 1 9	8 3 1 6	1 2 8 5	0 3 5 7	2 3 8 9	2 3 9 0
7 2 6 6	0 0 8 1	6 8 9 7	2 8 5 1	4 6 6 6	0 6 2 0	4 5 9 6	3 4 0 0
9 3 1 2	4 7 7 9	5 7 3 7	8 9 1 8	4 5 5 0	3 9 9 4	5 5 7 3	9 2 2 9
6 1 1 1	6 0 9 8	0 9 6 5	7 3 5 2	6 8 4 7	3 0 3 4	9 9 7 7	3 7 7 0
2 3 1 0	4 4 7 6	9 1 4 8	0 6 7 9	2 6 6 2	2 0 6 2	0 5 2 2	9 2 3 4
9 8 2 6	8 8 5 7	8 6 7 5	6 6 4 2	5 4 7 1	8 8 2 0	4 3 0 8	2 1 0 5
6 7 0 3	8 2 4 8	6 0 6 4	6 9 6 2	0 0 5 3	8 1 8 8	6 4 9 4	4 5 0 9
1 1 1 0	9 4 8 6	6 5 3 3	3 9 5 4	1 9 4 4	1 5 1 6	1 6 8 2	3 4 0 4
9 6 5 1	1 4 5 6	5 6 1 3	0 3 5 7	4 2 4 4	3 3 4 1	9 6 0 5	3 5 6 7
8 3 5 0	5 7 2 8	4 3 3 8	0 8 2 4	7 8 9 9	1 3 0 7	5 8 1 4	8 6 8 8
6 9 8 2	5 1 2 6	7 7 3 6	3 3 8 3	6 2 1 5	3 4 4 1	8 5 7 8	2 2 7 7
6 4 9 0	7 6 4 4	7 0 8 5	8 3 6 1	5 6 6 2	4 1 4 1	9 8 7 7	3 7 4 7
8 5 7 0	2 1 5 0	8 1 4 0	4 3 5 5	5 3 2 1	2 5 4 8	0 2 8 0	7 5 4 3
9 1 6 9	0 4 0 8	4 3 5 3	6 1 2 2	8 9 1 3	9 9 3 0	4 1 6 9	6 0 3 2
2 1 2 7	0 1 6 2	6 1 7 6	4 9 6 9	8 1 8 5	9 3 1 2	8 7 4 8	8 5 7 5
8 0 9 0	9 8 7 2	1 9 6 8	0 2 6 3	0 6 8 1	2 6 6 2	6 8 3 1	3 1 0 6
2 9 5 9	9 0 1 1	1 4 4 8	4 3 4 6	7 0 1 9	8 1 4 8	1 5 5 7	8 4 0 0