# On Randomness and Probability

## How To Mathematically Model Uncertain Events

*Rajeeva L Karandikar*

**Whether random phenomena exist in nature or not, it is useful to think of the notion of randomness as a mathematical model for a phenomenon whose outcome is uncertain. Such a model can be obtained by exploiting the observation that, in many phenomena, even though the outcome in any given instance is uncertain, collectively there is a pattern. An axiomatic development of such a model is given below. It is also shown that in such a set-up an interpretation of the probability of an event can be provided using the 'Law of Large Numbers'.**

**Rajeeva Karandikar is with Indian Statistical Institute, New Delhi. He received his Ph.D. from Indian Statistical Institute in 1982. His research interests include stochastic processes and filtering theory.**

What is randomness? Do random phenomena exist outside of casinos and gambling houses? How does one interpret a statement like "*there is a 30 per cent chance of rain tonight*" — a statement we often hear on the news?

Such questions arise in the mind of every student when she/he is taught probability as part of mathematics. Many students who go on to study probability and statistics in college do not find satisfactory answers to these questions. Those who did not (and some of those who did) study probability as part of their curriculum are generally sceptical when it comes to the notions of probability and randomness. But many of them still rely on these notions — like physicists when it comes to statistical mechanics and quantum theory and engineers when it comes to communications, design of reliable systems and so on.

How does one interpret the statement that "there is a 30 per cent chance of rain tonight"?

Let us look at the question: What is a random phenomenon? Some accept that the outcome of a toss of a coin is a random event since it is not known whether the coin will come up *Heads* or *Tails*. But if one were to write down all the parameters involved, like

Some accept that the outcome of a coin toss is random since it cannot be predicted. But if one were to write down all the parameters involved, then it is conceivable that the exact path of the coin can be described by equations of motion. And if one can solve these equations, the outcome is deterministic, not random.

the exact force applied, the point where the force is applied, the wind velocity, the density of air, ... then it is conceivable that the exact behaviour of the coin can be described by equations of motion; and if one is able to solve them, the outcome can be determined. Thus it can be argued that the outcome is deterministic, not random. We may not be able to determine it easily though!

The above argument prompts us to think: *Are we calling certain events random out of sheer ignorance?*

## Randomness as a Model for Uncertainty

One view which is not open to such criticism is to think of the notion of randomness as a mathematical model for events whose outcome, even in principle, is not completely specified. Immediately two questions arise. How can we model an event if its outcome is uncertain? And why should we model such events?

Think of the notion of randomness as a mathematical model for events whose outcome is not completely specified.

Let us look at the second question first. One can think of many situations where we have to make decisions under uncertainty. We need to take a train at 6.30 pm at the railway station and we need to decide when to start from home — we know that it may take anywhere between 30 minutes and 1 hour depending on the traffic; in any case before we start we don't know exactly how long it will take. Consider another situation: A drug company has come out with a drug which it claims is better than chloroquin for the treatment of malaria, and the government agency needs to decide whether to allow the company to sell the drug in the market. No one can say with certainty which medicine is more effective and what side effects the medicine may have. This is the case with most medicines. Take another situation which all of us face or have faced in the past: at the end of a school year, the teacher needs to decide which students deserve to be promoted to the next class — it is not feasible for the teacher to ask each student to do everything that has been taught in the class. We all

accept the solution in this case because we grew up with it: the teacher chooses some questions related to the material that has been taught during the year; and based on the answers to these chosen questions, the decision is made. Everyone knows that the final marks obtained by a student depend on the questions that are asked. It is possible that a question paper set by another teacher will yield a very different result. Yet, the marks obtained still give an indication of what the student has learned. We believe that it is extremely unlikely that a student who has got the highest marks in a test will fail the test if the paper were set by another teacher. One last example — the government wants to decide if enough foodgrains will be produced in the country this year or whether there will be a shortfall (in which case it has to import). In this case, as the data on food production will be available only after the harvest, when it may be too late to import, the government needs to estimate the foodgrain production and make a decision in time. The cost of an error in this case is very high for the country as we know from recent experience.

One can think of many situations where we have to make decisions based on incomplete information. So if we can mathematically model the uncertainty, it may help us in decision-making.

We can think of many more situations where we have to make decisions when we do not have complete information — may be because the event is a future event, or our understanding of the underlying phenomenon is incomplete, or it is too expensive to gather the information. Thus, if we can mathematically model the uncertainty, it may help us in decision-making.

Now let us examine the other question. How can we model uncertain events mathematically? Over the centuries, mankind has observed many phenomena in which the outcome in any given instance is uncertain, but collectively the outcomes conforms to a pattern.

An example of this is: Though, to start with, one could not tell whether an unborn child would be a boy or a girl, the total number of births in a town over a year showed a pattern — the number of male children and female children were approximately the same.

Over the centuries, mankind has observed many phenomena where the outcome in any given instance is uncertain; but collectively it conforms to a pattern. We cannot say whether a particular unborn baby will be a boy or a girl, but the number of male and female children born worldwide is always almost the same.

And this was observed in different towns, across the continents. The situation has changed marginally. Today, by medical tests, it can be determined a few months before birth if the unborn child is a boy or a girl; but even today, there is no deterministic model which can tell us the sex of an unborn child at the moment of conception.

The next example is from physics — about radioactive substances. It is known that certain substances like radium and uranium spontaneously emit particles like alpha and beta particles and/or electromagnetic radiation like gamma rays. This phenomenon is called 'radioactive decay'. This happens because some of the nuclei (i.e. radioactive nuclei) of such substances are unstable. It is also observed that the rate of this decay is proportional to the number of radioactive nuclei present in the substance, and does not depend on other factors such as the shape of the substance and other physical conditions of the environment. In fact it has been observed that the number of radioactive nuclei present in a sample of a radioactive substance is reduced to half the initial number in a fixed length of time. This time is called the *half life*. We seem to know exactly what will happen if we have, say, 1 gram of radioactive material. Yet there is no (deterministic) model at the atomic level for determining when a specified atom will disintegrate.

We seem to know exactly what will happen if we have a gram of radioactive material. Yet there is no deterministic model at the atomic level to predict when a specified atom will disintegrate.

Similar is the case with the kinetic theory of gases. It deals with the collective behaviour of gas molecules, but there is no deterministic model for the behaviour of an individual gas molecule. There are many such instances in physics.

## The Model

Let us assign to an uncertain event a number between 0 and 1 which we call its *probability of occurrence* with the understanding that higher the number, the higher is the chance that it will occur. Also, let us postulate that the *certain* event (i.e. an event that will

definitely happen) has probability 1 and the *null* event (i.e. an event that will never occur) has probability 0. We also postulate that if two events cannot occur simultaneously (such events are called mutually exclusive), then the probability that one of the two events will occur is the sum of their respective probabilities.

Let us consider experiments that can result only in one of countably many outcomes (finite or infinite) — we exclude, for now, experiments which can result in one of uncountably many outcomes. Let us represent the outcomes as $\omega_i$ and let

An infinite set can be *countable*: e.g. the set {1, 2, ... *n*, ...} or *uncountable*: e.g. the set of all points on the unit interval (0, 1).

$$\Omega = \{\omega_1, \omega_2, \omega_3, \cdots, \omega_N\}$$

if the total number of outcomes is $N < \infty$ or

$$\Omega = \{\omega_1, \omega_2, \omega_3, \cdots, \omega_n, \cdots\}$$

if the total number of outcomes is countably infinite.

The probability of occurrence of an event is a number between 0 and 1. The higher the number, the higher is the chance that the event will occur.

Subsets of $\Omega$ are called events. We say that the event $A$ has occurred if the experiment results in an outcome $\omega_i \in A$. A probability allocation for this experiment is given by a real-valued function $P$ defined on the set of all subsets of $\Omega$ such that

$$0 \leq P(A) \leq 1 \qquad \forall A \subseteq \Omega.$$

Further, if $A, B$ are mutually exclusive (i.e. $A \cap B = \Phi$), then

$$P(A \cup B) = P(A) + P(B). \tag{1}$$

Let $p_i$ be the probability of occurrence of the event $\{\omega_i\}$; i.e. $p_i = P(\{\omega_i\})$. Then the postulates stated above imply that :

1. $p_i \geq 0 \qquad \forall i$
2. $\sum_{i:\omega_i \in \Omega} p_i = 1$
3. $P(A) = \sum_{i:\omega_i \in \Omega} p_i, \quad \forall A \subseteq \Omega.$

Thus once we choose $p_i = P(\omega_i)$, the probabilities of all events $A \subseteq \Omega$ are determined. How does one go about choosing $p_i$?

Well, this is where the *modelling* aspect comes into the picture. The $p_i$'s of the probability model should reflect all the information we have on the phenomenon or should at least be a close approximation of the same. We will begin with the simplest situation and draw conclusions in this case. We will get an interpretation for the numerical value of the assigned probability of an event and this in turn will help us in modelling more complicated phenomena.

The probability model used should reflect all the information available on the phenomenon.

Let us now consider the situation where $\Omega$ is a finite set with $\Omega = \omega_1, \omega_2, \omega_3, \cdots, \omega_N$, and where given all the information about the phenomenon, we have reason to believe that all outcomes are *equally likely*. In this case, the appropriate choice of probabilities is

$$P(\omega_i) = \frac{1}{N} \quad \forall i \in \Omega.$$

This is clearly the case when there is an inherent symmetry in the phenomenon; for example, most of us will agree that "the *chance* that the first child born in a given nursing home the next day is a boy" is the same as "the chance that the child will be a girl". Thus the events $\omega_1$ = *the child is a boy* and the event $\omega_2$ = *the child is a girl* are equally likely, and we can model the probabilities for this experiment as

$$P(\omega_1) = \frac{1}{2}, \quad P(\omega_2) = \frac{1}{2}.$$

Similarly, if we are told that a family has 3 children but we have no further information, we are justified in postulating that all the 8 possibilities *GGG, GGB, GBG, GBB, BGG, BGB, BBG, BBB* are equally likely and hence the probability of each of these events is $1/8$. This is based on the observation that knowing that the first

child is a girl (or a boy) does not give any information about the sex of the next child.

Let us look at the following experiment. Consider an urn containing 12 balls of the same size and weight, numbered 1 to 12. Suppose that the balls with numbers 1, 2 and 3 are red balls, and the rest are blue. If the balls in the urn are mixed well and one ball out of them is drawn without looking at the colour/number, then the 12 events (that the ball with number $i$ on it is drawn, $1 \leq i \leq 12$ ) can be modelled as equally likely — each with probability $1/12$. As a result the probability that the ball so drawn is red is $1/4$. Now even if the balls are not numbered, but the urn contains 3 red and 9 blue balls, then the probability of drawing a red ball is still $1/4$. Thus even if the balls are not numbered, we can always pretend that they are numbered.

If a given experiment can result in $N$ equally likely outcomes and a given event occurs in $M$ out of $N$ outcomes, then its probability of occurrence can be modelled to be $M/N$.

We can thus draw the following conclusion: if a given experiment can result in $N$ outcomes, and based on all the information that we have on the phenomenon, they seem to be equally likely, and if a given event occurs in $M$ out of the $N$ outcomes, then its probability (corresponding to the model that the $N$ outcomes are equally likely) is $M/N$. Note that we are not adopting this as a definition, but as a model for the phenomenon. If another person has more information on the experiment, his model, i.e. allocation of probabilities, could be quite different.

Now let us consider two urns, both like the one considered above. The experiment consists of drawing one ball from each of the urns. This time, all the $12 \times 12 = 144$ outcomes are equally likely. Out of these, $3 \times 3 = 9$ outcomes determine the event that both the balls drawn are red, and hence its probability is $9/144 = 1/16$. Note that this is equal to the product of the probabilities of the event that the first ball is red and of the event that the second ball is red. Here we are in a situation where the two events are independent — *i.e. the occurrence or otherwise of the first event does not change our perception of the second event.* In such a situation, the

events are said to be independent and the probability that both events occur can be taken to be the product of the two events. This is a very important notion and very useful in model building.

Consider two experiments,

$$\Omega_1 = \{\omega_i^{(1)} : i \in I^{(1)}\} \text{ and } \Omega_2 = \{\omega_j^{(2)} : j \in I^{(2)}\}$$

$I^{(1)}$, $I^{(2)} \subseteq N$, and suppose that we have a model for each of them, namely

$$P(\{\omega_i^{(1)}\}) = p_i^{(1)} \text{ and } P(\{\omega_j^{(2)}\}) = p_j^{(2)}.$$

If two events are independent, the occurrence or otherwise of the first event does not change our perception of the second event. The probability that both events occur can therefore be taken to be the product of the individual probabilities of the two events.

The set of possible outcomes for the *joint* experiment is

$$\Omega = \{(\omega_i^{(1)}, \omega_j^{(2)}) : i \in I^{(1)}, j \in I^{(2)}\}.$$

If the experiments are such that the outcome of one has no bearing on the outcome of the other, then it is reasonable to model the joint experiment as follows :

$$P(\{\omega_i^{(1)}, \omega_j^{(2)})\}) = p_i^{(1)} p_j^{(2)}.$$

Similarly, if we have a model for each of finitely many experiments and if these experiments are independent of each other, then we can construct a model for the experiment which consists of performing all these experiments together.

Now we are in a position to provide an interpretation for the probability of an event related to an experiment. We shall show that if this experiment can be repeated again and again (independently) then the limit of the proportion of occurrences of this event is exactly its probability.

## Law of Large Numbers

Let us now fix a set of outcomes, $\Omega = \{\omega_i : i \in I\}$, $\omega_i \neq \omega_j$ for $i \neq j$ where $I = \{1, 2, 3, \ldots, N\}$ or $I = \{1, 2, 3, \ldots, N, \ldots\}$. Also let

us fix an assignment of probabilities for subsets of $\Omega$. Recall that such an assignment is determined by

$$p_i = P(\{\omega_i\})$$

and then for any *event* $A$,

$$P(A) = \sum_{i:\, \omega_i \in A} p_i$$

A function $X$ from $\Omega$ into $\boldsymbol{R}$ is called a *random variable.* We think of $X$ as follows : $X$ represents a certain numerical characteristic of the outcome of the experiment, and after the experiment is conducted we get to observe the function $X$ at the outcome (we may or may not actually observe the outcome).

A random variable represents a certain (measurable) characteristic of the outcome of the experiment.

Let $X$ be a random variable and let $f$ be a function from the real line into itself. We denote by $f(X)$ the random variable given by

$$f(X)\,(\omega_i) = f(X(\omega_i)).$$

Also, let $R(X)$ denote the range of $X$ i.e. $R(X) = \{x \in \mathrm{R}$: there exists $\omega \in \Omega$ with $\mathrm{X}(\omega) = x\}$. For a subset A of $\boldsymbol{R}$, we will write $X \in A$ for the set $\{\omega_i: X(\omega_i) \in A\}$. When $A = \{x\}$ we will also write $X = x$ for $X \in A$. The function $x \to P(X = x)$ is called the distribution of $X$. It is easy to check that

$$P(\{X \in A\}) = \sum_{x \in A} P(\{X=x\}).$$

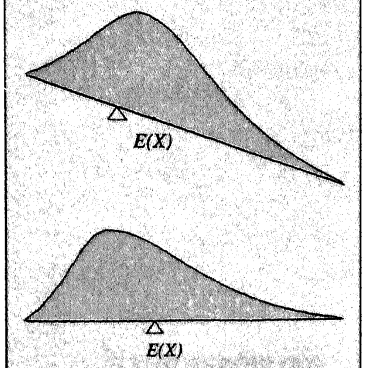
A random variable $X$ is said to be bounded if there exists a finite constant $K$ such that

$$P(-K \le X \le K) = 1.$$

For a bounded random variable $X$, we define its *expected* value $E(X)$ by

$$E(X) = \sum_{i \in I} X(\omega_i)\, p_i. \qquad (2)$$

**Expectation and Variance**




Think of the expected value as the centre of gravity of the probability distribution. Imagine placing mass $P(\{X=x\})$ at the point $x$ (for each $x$) on a beam; the balance point of the beam is the expected value of $X$.

The variance of a probability distribution indicates how dispersed the distribution is about its centre of gravity or how spread out on the average are the values of the random variable about its expected value.

$E(X)$ represents the quantity we expect to observe on the average, if we repeat the experiment (independently) a large number of times. Hence the name *expected value*. A justification of the statement made above is given towards the end of this article.

Let us observe that for a random variable $X$,

$$E(X) = \sum_{x \in R(X)} x\, P(|X = x|). \tag{3}$$

To see this, let $A_x = |\omega_i\colon X(\omega_i) = x|$. Then

$$\begin{aligned} E(X) &= \sum_{i \in I} X(\omega_i)\, p_i \\ &= \sum_{x \in R(X)} \sum_{\omega_i \in A_x} x\, P(|\omega_i|) \\ &= \sum_{x \in R(X)} x\, P(\{|X = x|). \end{aligned}$$

For a bounded random variable $X$ such that $E(|X|^2) < \infty$, let us define the *variance* of $X$ by

$$\mathrm{Var}(X) = E(X-\mu)^2$$

where $\mu = E(X)$.

Let us note that for a positive random variable $Y$,

$$\begin{aligned} \lambda P(|Y \geq \lambda|) &= \lambda \sum_{y \in R(Y):\, y \geq \lambda} P(|Y = y|) \\ &\leq \sum_{y \in R(Y):\, y \geq \lambda} y P(|Y = y|) \\ &\leq \sum_{y \in R(Y)} y P(|Y = y|) \\ &= E(Y), \end{aligned}$$

and as a consequence, one has (for positive random variables $Y$)

$$P(Y \geq \lambda) \leq \frac{1}{\lambda} E(Y). \qquad (4)$$

Using this for $Y = (X - \mu)^2$ (where $\mu = E(X)$), one has

$$P(\{|X - \mu| \geq t\}) \leq \frac{1}{t^2} \mathrm{Var}(X). \qquad (5)$$

This inequality is known as *Chebychev's inequality.*

Chebychev's Inequality gives a bound on the probability of the tails of a distribution.

## Independence

We say that two events A, B are independent if

$$P(A \cap B) = P(A)\, P(B).$$

A collection of random variables $X_1, X_2, \cdots, X_n$ is said to be a collection of independent random variables if

$$P(\cap_{j=1}^{n} \{X_j = x_j\}) = P(\{X_1 = x_1\})\, P(\{X_2 = x_2\}) \cdots P(\{X_n = x_n\})$$

for all $x_j \in R(X_j)$: $1 \leq j \leq n$.

**Lemma 1:** Let $X, Y$ be independent random variables. Let $f, g$ be bounded functions on the real line. Then

$$E(f(X)\, g(Y)) = E(f(X))\, E(g(Y)).$$

**Proof :** First consider the case when both $f, g$ are positive functions. Then it can be checked that

$$E(f(X)\, g(Y)) = \sum_{x \in R(X), y \in R(Y)} f(x)\, g(y)\, P(\{X = x, Y = y\}).$$

Using independence of $X, Y$, it follows that

$$E(f(X)g(Y)) = \sum_{x \in R(X), y \in R(Y)} f(x)g(y)P(\{X=x\}), P(\{Y=y\}).$$

The required identity now follows from this.

**Theorem 2:** Let $X,Y$ be bounded independent random variables. Then

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y).$$

The essential content of the Weak Law of Large Numbers is that if our experiment can be repeated again and again, independently, then the limit of the proportion of occurrences of this event is exactly its probability!

**Proof :** Let $\mu = E(X)$, $\nu = E(Y)$, $U = X - \mu$, $V = Y - \nu$. It is easy to see that $\text{Var}(X) = \text{Var}(U)$, $\text{Var}(Y) = \text{Var}(V)$, $\text{Var}(X + Y) = \text{Var}(U + V)$. Also, $E(U) = E(V) = 0$. Thus

$$\begin{aligned}\text{Var}(U+V) &= E((U+V)^2) \\ &= E(U^2) + E(V^2) + 2E(UV) \\ &= \text{Var}(U) + \text{Var}(V)\end{aligned}$$

where we have used the previous lemma in deducing that $E(UV) = 0$. The required result now follows from this.

We are now in a position to prove the *Weak Law of Large Numbers*.

**Theorem 3:** Let $X_1, X_2, \cdots, X_n \cdots$ be a sequence of bounded random variables such that for each $n$, $X_1, X_2, \cdots, X_n$ is a collection of independent random variables and such that for all $i \geq 1$, $R(X_i) = R(X_1)$ and

$$P(\{X_i = x\}) = P(\{X_1 = x\}) \quad \forall x \in R(X_1).$$

There is also the *Strong Law of Large Numbers* which deals with a different and actually stronger mode of convergence of $Z_n$, the proof of which is, however, beyond the scope of this article.

Let $\mu = E(X_1)$ and let $Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then for all $\varepsilon > 0$,

$$\lim_{n \to \infty} P(\{|Z_n - \mu| > \varepsilon\}) = 0.$$

**Proof** : Using Theorem 2, it follows that

$$\text{Var}\,(Z_n) = \frac{1}{n^2}\left\{\sum_{i=1}^{n} \text{Var}\,(X_i)\right\}$$

$$= \frac{1}{n^2}\, n\, \{\text{Var}\,(X_1)\}$$

$$= \frac{1}{n}\, \{\text{Var}\,(X_1)\}$$

Now using the inequality 5, we obtain for $\varepsilon > 0$

$$P(\{|Z_n - \mu| > \varepsilon\}) \leq \frac{1}{\varepsilon^2}\,\frac{1}{n}\,\text{Var}\,(X_1).$$

The required conclusion follows from this.

## Interpretation of Probability of an Event

Let us consider an experiment with the space of outcomes $\Omega = \{\omega_i : i \in I\}$ and with assignment of probabilities $P(\{\omega_i\}) = p_i$. (Here, I is either equal to $\{1,2, \cdots, N\}$ or is the set of natural numbers.)Let us fix an *event* A (i.e. a subset of $\Omega$) with $P(A) = \theta$.

Let us consider repeating the experiment $n$ times, in such a way that the outcome of the previous trials has no influence on the next trial. This time the set of outcomes of this repeated experiment can be taken to be

$$\Omega^n = \{(\omega_{i_1}, \omega_{i_2}, \omega_{i_3}, \cdots, \omega_{i_n}) : i_1, i_2, \cdots, i_n \in I\}.$$

Since we have assumed that the experiments have been performed independently of each other, we are justified in assigning the probabilities as follows:

$$P(\{(\omega_{i_1}, \omega_{i_2}, \omega_{i_3}, \cdots, \omega_{i_n})\}) = p_{i_1}\, p_{i_2} \cdots p_{i_n}.$$

### Suggested Reading


W Feller. An Introduction to Probability Theory and Its Applications. Vol. 1. (Third Edition). Wiley-Eastern, New Delhi. 1985.

P G Hoel, S C Port and C J Stone. Introduction to Probability Theory. Universal Book Stall, New Delhi. 1991.

K L Chung. Elementary Probability Theory and Stochastic Processes. Narosa Publishing House, New Delhi. 1978.

The statement that "there is a 30 per cent chance of rain tonight" simply means that under a given weather forecasting probability model, the probability of the event that it will rain tonight is 0.3.

Let us define random variables $X_1, X_2, \cdots X_n$ as follows:

$$X_i((\omega_{i_1}, \omega_{i_2}, \omega_{i_3}, \cdots, \omega_{i_n})) = 1_A(\omega_i)$$

where $1_A$ denotes the indicator function of the set $A$ i.e. $1_A(\omega_i) = 1$ if $\omega_i \in A$ and $1_A(\omega_i) = 0$ if $\omega_i \notin A$. Then it follows that $X_1, X_2, \ldots, X_n$ satisfy the conditions of Theorem 3 with $E(X_1) = \theta$. It thus follows that given $\varepsilon > 0, \eta > 0$, we can choose $n_0$ such that for $n \geq n_0$, one has

$$P(|\frac{1}{n}(X_1 + X_2 + \cdots + X_n) - \theta| > \varepsilon) \leq \eta.$$

Let us note that $(X_1 + X_2 + \cdots + X_n)/n$ is the proportion of the times the event $A$ occurred in the $n$ independent repetitions of the experiment. We have seen above that for large $n$, this observed proportion is close to the probability of $A$.

This gives us an interpretation of $P(A)$. Similarly, we can get an interpretation for $E(X)$—namely, if we repeat the experiment a large number of times and compute the average of the observed values of $X$, then, with a high probability, this average is close to the expected value $E(X)$ of $X$.

Let us briefly return to the question posed at the beginning of the article: how does one interpret a statement like *there is a 30 per cent chance of rain tonight*?

From some theoretical reasoning and some observational data, weather forecasters (and other forecasters) usually have probability models for forecasting. The above statement simply means that under such a model, the probability of the event that it will rain tonight is 0.3.

*Address for correspondence*
Rajeeva L Karandikar,
Statistics and
Mathematics Unit,
Indian Statistical Institute,
7 S J S Sansanwal Marg
New Delhi 110 016, India.

**Niels Bohr said ...** "It is difficult to predict, especially about the future".