

GENERAL SCIENCE NOTES

PROBABILITY AND ITS APPLICATION TO THE ORIGIN OF LIFE

*By David G. Kissinger, Center for Health Promotion,
Loma Linda University*

WHAT THIS ARTICLE IS ABOUT

Probability theory has been developed and verified using physical devices such as coins or cards that meet the assumptions of fairness and independence inherent in probability theory. Because these assumptions may not apply to steps postulated in a scenario for the origin of life and because some of the steps may be nearly unique, probabilistic arguments regarding the origin of life do not seem to be applicable except as gross generalizations.

In this article the author quotes non-creationist scientists who suggest that time, space, and raw materials in the universe may be too limited to permit the origin of life by chance alone.

Probability is a mathematical construct that can be demonstrated to model well-behaved non-deterministic phenomena such as coin tosses, and is accepted as being useful in modeling and analyzing masses of data from well-designed scientific studies of less well-behaved random processes. The application of probability analysis to events which may be nearly unique and happen so seldom as to be rarely observed seems questionable from a practical viewpoint.

Probability is essential in comparing or predicting the outcome of events based on a particular model. In the case of the origin of life suitable models amenable to scientific investigations have not yet been proposed or evaluated.

The ability to foretell the future and to know and understand the past has been coveted by man for a long time. This is illustrated by the TV serial Star Trek fictional character Mr. Spock, who could state the probability of a unique, future event with great precision. Of course, the script invested Spock with an aura of authority as Chief Science Information Officer and assigned him superhuman mental abilities because of his "race."

In the real world we are faced with unique past or future events to which we would like to assign probabilities because, even though in a scientific sense we do not "know" the truth about the matter, we want to talk about it with a degree of certainty and to discover incontrovertible evidence, if possible.

This note looks at some basic properties of probability and considers the appropriateness of using probability to demonstrate the impossibility or inevitability of the origin of life.

What is probability? and what do we mean by the term? are philosophical questions to which no clear or entirely satisfactory answers have been proposed (Theobald 1968). The frequency theory of probability is popular among many scientists and some philosophers. This theory defines probability objectively in terms of the frequency of occurrences in long runs. The advantage is that the definition is at the same time its measure (Theobald 1968).

Probability is a complex mathematical topic (Feller 1968, 1971; de Finetti 1974; Noether 1974). One way to define probability is with three axioms:

$$(1) P(A) \geq 0, (2) P(S) = 1, \text{ and } (3) P(A+B) = P(A) + P(B), \text{ if } AB = 0.$$

Statement (1) tells us that no probability can be negative; $P(A)$ means the probability that the event “A” will occur; if $P(A) = 0$, “A” will not occur.

In statement (2), $P(S)$ means the probability of the occurrence of the entire set of events or outcomes that can occur for a particular situation; for instance, when a coin is tossed the possible outcomes are “heads” or “tails”; other possibilities such as landing on its edge or disappearing are ignored; the biggest value that probability can take is 1 which is certainty. In statement (3), $P(A+B)$ is the probability that 2 disjoint events will occur and indicates that this is the sum of the individual probabilities; by disjoint we mean that either “A” or “B” occurs and not some fractional happening of “A” and “B” simultaneously.

Because these are mathematical axioms, there is no point in talking about their “true nature” or “definition”; these are like the set of rules which define a game of chess (Feller 1968).

Not all events in the real world need to be treated probabilistically. Some phenomena always produce the same deterministic outcome under specified identical conditions. An illustration is the way an object falls to the ground with constant acceleration due to the force of gravity. Thus constant, static conditions yield a result which can be predicted with a great degree of certainty for a deterministic process.

In contrast are nondeterministic or random events where repeated observation under constant static conditions do not always lead to the same outcome. A popular example is the result of a coin toss. Consider a coin that is tossed many times. Heads or tails result in a seemingly erratic and unpredictable manner. Many such nondeterministic phenomena show a statistical regularity based on the concept of probability. By statistical regularity we mean that the outcome of a suitably large number of observations of a non-deterministic phenomenon can be predicted accurately before the observations are made if the statistics of the phenomenon are known. This means that a model has been proposed for the phenomenon

and that the model has been tested, accepted, and is known to explain the situation adequately.

In the simple coin toss certain assumptions are inherent in the probabilistic interpretation. First, the coin must be fair so there is an equal opportunity for each outcome, heads or tails in this case, to occur at each toss. This means that the coin is not unbalanced, double headed, or biased in any way. Most recently minted, unaltered coins will be fair because of manufacturing procedures.

Second, the coin is independent of its past. It has no memory of its past outcomes, and the past cannot influence the outcome of the current toss. For example, the probability of a head on any coin toss is 0.5. If a series of 10 heads in a row has been tossed for a coin, then for the 11th toss the probability of heads is still 0.5; the coin has not run up a deficit of tails which it is obligated to repay.

The assumptions of fairness and independence have been realized in other nondeterministic phenomena used to investigate probability. Examples are containers with fixed numbers of distinguishable but identical objects such as an urn with red or black balls or gambling devices such as cards or a roulette wheel.

There is a considerable gap between simple cases such as the coin toss and situations which are important in the real world. Again the assumptions of fairness and independence are important, but the assumptions may be compromised or ignored. Examples would be life insurance, risk management, and the interpretations of the results of scientific experiments.

An important reason why these results can be evaluated probabilistically is that there is a sufficient number of instances under consideration so that analysis is possible. In a statistical sense we would say that the sample size is large enough. How large is large enough is the subject of controversy, but usually it is on the order of 25 or 100.

In the analysis of the results of scientific experiments, one cannot use statistics if the number of occurrences under consideration is too small, because there is a direct relationship between the sample size and one's confidence in the interpretation of the outcome. Statistical theory tends to impose this limitation on us. As a consequence we really know little about using statistical theory and methods to evaluate the probability of unique events.

Two important concepts are directly involved with the application of probability to the origin of life. These are RANDOMNESS and IMPOSSIBILITY, concepts that are intuitively understood but for which concrete definitions are difficult to find.

Knuth (1969) in his monumental 3-volume series, *The Art of Computer Programming*, devotes 160 pages to his treatment of the generation of pseudorandom numbers by a computer and to the evaluation of such techniques. Certain tests can be applied to a series of numbers to determine if the series meets the criterion of randomness. Even in a series of statistically acceptable random numbers, a non-random pattern may be detected.

The point he makes is that it is not always easy to distinguish between random and non-random even under ideal conditions.

Impossibility is another relative term. Borel (1962) discusses probabilities that are negligible on 4 different scales. On the human scale events rarer than one in a million are essentially ignored. On the terrestrial scale he suggests that one in 10^{15} is negligible, since this is about a billion times as small as the probability ignored by one man. On the cosmic scale he sets one in 10^{50} as being either impossible or at least would never be observed. On the supercosmic scale he uses a number on the order of one in 10^n , where n is a number of more than 10 figures.

Consider the following situation. Using current actuarial practice, what is the probability that a man can live to be 1000 years old? According to formulas on which modern mortality tables are based, the proportion of men surviving 1000 years is of the order of magnitude of one in 10^{35} (10 to the 10th power to the 35th power). This statement makes no sense from a biological or sociological point of view, but considered exclusively from statistical considerations it certainly does not contradict any experience. Since fewer than 10^{10} people are born in a century, it would require 10^{25} centuries to test the contention statistically which is 10^{34} times the supposed lifetime of the earth. Such small probabilities are compatible with our notion of impossibility (Feller 1968).

Another example of a small probability involves the following argument proposed to show that a pattern is needed to make a biologically active enzyme (Pardee 1962). This argument was proposed before it was known that protein molecules could contain subunits; however, the enzyme prokaryotic DNA I polymerase is a single polypeptide chain with MW = 110,000 (White et al. 1978).

Suppose we consider a protein of molecular weight 100,000 which is composed of 830 amino acids in a particular order. The number of possible ways that 830 amino acids can be arranged to form a protein of this size is 20^{830} . A sphere constructed from one of each of these 20^{830} molecules would have a radius of $10^{34.5}$ light years (Pardee 1962). The visible universe has a diameter of about 10^{12} light years.

How should one approach the problem of using very small probabilities to bolster the concept of the apparent impossibility of the origin of life

from non-living sources? At the present time the theory involves the spontaneous union of amino acids to form postulated prebiologically significant proteins, given the necessary precursor amino conditions. Does this really establish the impossibility of abiogenesis? We do not know the exact conditions that may have prevailed. Considering this from a purely probabilistic viewpoint we do not know that the growth or breakup of a polypeptide or other macromolecule was strictly random or that some type of autocatalysis would mean the process was not random; this might make certain sequences of amino acids more probable or it might make biologically desirable sequences less probable.

The difficulties of actually applying probability to the events postulated by some to have occurred in the origin of life have been noticed previously. The following notes from S. W. Fox (ed.), 1965, *The Origins of Prebiological Systems and of Their Molecular Matrices*, will support this contention. I have selected passages that are especially interesting to me; I am not trying to make a statement about the philosophy or beliefs of the particular author.

J. B. S. Haldane in his paper, "Data Needed for a Blueprint of the First Organism," postulates a very primitive kind of "organism" and makes the following statement:

If the minimal organism involves not only the code for its one or more proteins, but also twenty types of soluble RNA, one for each amino acid, and the equivalent of ribosomal RNA, our descendants may be able to make one, but we must give up the idea that such an organism could have been produced in the past except by a similar pre-existing organism or by an agent, natural or supernatural, at least as intelligent as ourselves, and with a good deal more knowledge (p 12).

Haldane suggests that something like a generalized phosphokinase may have been involved which may have contained about 25 amino residues. In talking about the generation of such a molecule from existing amino acids, he states:

But even this would mean one out of 1.3×10^{30} possibilities. This is an unacceptable, large number. If a new organism were tried out every minute for 10^8 years, we should need 10^{17} simultaneous trials to get the right result by chance. The earth's surface is 5×10^{18} cm². There just isn't, in my opinion, room. Sixty bits, or about 15 amino acids, would be more acceptable probabilistically, but less so biochemically (p 14).

Peter T. Mora in his paper, "The Folly of Probability," points out some problems and limitations inherent in present-day science when trying

to account for the origin of life. I quote the following as an example of his statements:

A further aspect I should like to discuss is what I call the practice to avoid facing the conclusion that the probability of a self-reproducing state is zero. This is what we must conclude from classical quantum mechanical principles, as Wigner demonstrated....These escape clauses postulate an almost infinite amount of time and an almost infinite amount of material (monomers), so that even the most unlikely event could have happened This is to invoke probability and statistical considerations when such considerations are meaningless. When for practical purposes the condition of infinite time and matter has to be invoked, the concept of probability is annulled. By such logic we can prove anything, such as that no matter how complex, everything will repeat itself, exactly and innumerably (p 45).

In the discussion following Mora's paper, Carl Sagan takes Mora to task for suggesting that 5 billion years is an infinite period of time. Mora replies, "That is a matter of opinion" (p 60).

J. D. Bernal comments:

In the first place, the questions may be wrongly put; such a question, for instance, as 'could life have originated by a chance occurrence of atoms' clearly leads as our knowledge, and also the limitations of the time and space available, increase, to a negative answer (p 52-53).

H. H. Pattee makes some comments concerning probability:

The concept of probability I don't believe is properly used here, at least the way Laplace and others represent it. The idea is that two models which are sufficiently well defined in order to apply a probability measure may then be objectively compared with probability theory, which is only a mathematical theory. In this sense, probability cannot possibly explain anything. It is an objective way to compare two alternative models. And in this sense, I don't believe it is folly to use probability (p 58).

Other remarks by Pattee include:

I think we agree that the chance hypothesis for the origin of life is unsatisfactory. It is not only conceptually barren, but also untestable empirically. However, if we create an alternative model which is sufficiently well defined to apply probability theory, it may then be correctly applied. It is not the fault of probability theory that a good model hasn't been made yet (p 58).

A Szutka suggests that more than chance was responsible for events leading to living systems. He mentions the possibility of several (unknown)

parameters acting and increasing the probability that the event would occur (p 60). In terms of our previous discussion this means that the two or more molecules are not independent, and therefore it is going to be difficult to apply probability measures to such an instance.

Mora's response is:

I hope I don't give the impression that by pure chance it [the origin of life] could have happened just by itself, without there being some particular yet unknown attributes or physicochemical properties in the interacting molecules (p 60).

In conclusion, probability is a mathematical construct which can be demonstrated to model well-behaved non-deterministic phenomena such as the results of coin tosses and is accepted as being useful in modeling and analyzing masses of data from well-designed scientific studies of less well-behaved random processes. The application of probability analysis to events which may be nearly unique and happen so seldom as to be rarely observed seems questionable from a biological point of view. Forgetting about problems of bias and independence which are inherent in the discussion, some scientists other than creationists agree that the appearance of life through the sole action of random events on molecules is so excessively close to being impossible that other, possibly supplemental, explanations must be sought.

LITERATURE CITED

- Borel E. 1962. Probabilities and life. NY: Dover Publ.
- Feller W. 1968-1971. An introduction to probability theory and its applications. 2 vols. (Vol. 1, 3rd ed., 1968; Vol. 2, 2nd ed, 1971). NY: John Wiley & Sons.
- di Finetti B. 1974. Interpretation of probability. In: Kruskal WH, Tanur JM, editors. International Encyclopedia of Statistics, Vol. 2. NY: Free Press, p 744.
- Haldane JBS. 1965. Data needed for a blueprint of the first organism. In: Fox SW, editor. The Origins of Prebiological Systems and of Their Molecular Matrices. NY: Academic Press, p 11-15.
- Knuth DE. 1969. The art of computer programming. Reading, MA: Addison-Wesley.
- Mora PT. 1965. The folly of probability. In: Fox SW, editor. The Origins of Prebiological Systems and of Their Molecular Matrices. NY: Academic Press, p 39-52.
- Noether GE. 1974. Formal probability. In: Kruskal WH, Tanur JM, editors. International Encyclopedia of Statistics, Vol. 2. NY: Free Press, p 734.
- Pardee AB. 1962. The synthesis of enzymes. In: Gunsalus IC, Stanier RY, editors. The Bacteria: A Treatise on Structure and Function. NY: Academic Press.
- Theobald DW. 1968. Introduction to the philosophy of science. London: Methuen.
- White A, Handler P, Smith EL, Hill RL, Lehman IR. 1978. Principles of biochemistry. 6th ed. NY: McGraw-Hill.