CS 480 **Computational Theory**

Descriptive Complexity

From "Randomness and Mathematical Proof" by Gregory J. Chaitin in Scientific American, May, 1975

Although randomness can be precisely defined and can even be measured, a given number cannot be proved to be random. This enigma establishes a limit to what is possible in mathematics.

Almost everyone has an intuitive notion of what a random number is. For example, consider these two series of binary digits: 01010101010101010101 and 01101100110111100010.

The first is obviously constructed according to a simple rule; it consists of the number 01 repeated ten times. If one were asked to speculate on how the series might continue, one could predict with considerable confidence that the next two digits would be 0 and 1. Inspection of the second series of digits yields no such comprehensive pattern. There is no obvious rule governing the formation of the number, and there is no rational way to guess the succeeding digits. The arrangement seems haphazard; in other words, the sequence appears to be a random assortment of 0's and 1's.

The second series of binary digits was generated by flipping a coin 20 times and writing a 1 if the outcome was heads and a 0 if it was tails. Tossing a coin is a classical procedure for producing a random number, and one might think at first that the provenance of the series alone would certify that it is random. This is not so. Tossing a coin 20 times can produce any one of 2^{20} (or a little more than a million) binary series, and each of them has exactly the same probability. Thus it should be no more surprising to obtain the series with an obvious pattern than to obtain the one that seems to be random; each represents an event with a probability of 2^{-20}. If origin in a probabilistic event were made the sole criterion of randomness, then both series would have to be considered random, and indeed so would all others, since the same mechanism can generate all the possible series. The conclusion is singularly unhelpful in distinguishing the random from the orderly.

Clearly a more sensible definition of randomness is required, one that does not contradict the intuitive concept of a "patternless" number. Such a definition has been devised only in the past 10 years. It does not consider the origin of a number but depends entirely on the characteristics of the sequence of digits. The new definition enables us to describe the properties of a random number more precisely than was formerly possible, and it establishes a hierarchy of degrees of randomness. Of perhaps even greater interest than the capabilities of the definition, however, are its limitations. In particular the definition cannot help to determine, except in very special cases, whether or not a given series of digits, such as the second one above, is in fact random or only seems to be random. This limitation is not a flaw in the definition; it is a consequence of a subtle but fundamental anomaly in the foundation of mathematics. It is closely related to a famous theorem devised and proved in 1931 by Kurt Gödel which has come to be known as Gödel's incompleteness theorem. Both the theorem and the recent discoveries concerning the nature of randomness help to define the boundaries that constrain certain mathematical methods.

Algorithmic Definition

The new definition of randomness has its heritage in information theory, the science, developed mainly since World War II, that studies the transmission of messages. Suppose you have a friend who is visiting a planet in another galaxy, and that sending him telegrams is very expensive. He forgot to take along his tables of trigonometric functions, and he has asked you to supply them. You could simply translate the numbers into an appropriate code (such as the binary numbers) and transmit them directly, but even the most modest tables of the six functions have a few thousand digits, so that the cost would be high. A much cheaper way to convey the same information would be to transmit instructions for calculating the tables from the underlying trigonometric formulas, such as Euler's equation $e^{ix} = cos\ x + i\ sin\ x$. Such a message could be relatively brief, yet inherent in it is all the information contained in even the largest tables.

Suppose, on the other hand, your friend is interested not in trigonometry but in baseball. He would like to know the scores of all the major-league games played since he left the earth some thousands of years before. In this case it is most unlikely that a formula could be found for compressing the information into a short message; in such a series of numbers each digit is essentially an independent item of information, and it cannot be predicted from its neighbors or from some underlying rule. There is no alternative to transmitting the entire list of scores.

In this pair of whimsical messages is the germ of a new definition of randomness. It is based on the observation that the information embodied in a random series of numbers cannot be "compressed," or reduced to a more compact form. In formulating the actual definition it is preferable to consider communication not with a distant friend but with a digital computer. The friend might have the wit to make inferences about numbers or to construct a series from partial information or from vague instructions. The computer does not have that capacity, and for our purposes that deficiency is an advantage. Instructions given the computer must be complete and explicit, and they must enable it to proceed step by step without requiring that it comprehend the result of any part of the operations it performs. Such a program of instructions is an algorithm. It can demand any finite number of mechanical manipulations of numbers, but it cannot ask for judgments about their meaning.

The definition also requires that we be able to measure the information content of a message in some more precise way than by the cost of sending it as a telegram. The fundamental unit of information is the "bit," defined as the smallest item of information capable of indicating a choice between two equally likely things. In binary notation one bit is equivalent to one digit, either a 0 or a 1.

We are now able to describe more precisely the differences between the two series of digits presented at the beginning of this article: 01010101010101010101 and 01101100110111100010.

The first could be specified to a computer by a very simple algorithm, such as "Print 01 ten times." If the series were extended according to the same rule, the algorithm would have to be only slightly larger; it might be made to read, for example, "Print 01 a million times." The number of bits in such an algorithm is a small fraction of the number of bits in the series it specifies, and as the series grows larger the size of the program increases at a much slower rate.

For the second series of digits there is no corresponding shortcut. The most economical way to express the series is to write it out in full, and the shortest algorithm for introducing the series into a computer would be "Print 01101100110111100010." If the series were much larger (but still apparently patternless), the algorithm would have to be expanded to the corresponding size. This "incompressibility" is a property of all random numbers; indeed, we can proceed directly to define randomness in terms of incompressibility: A series of numbers is random if the smallest algorithm capable of specifying it to a computer has about the same number of bits of information as the series itself.

This definition was independently proposed about 1965 by A. N. Kolmogorov of the Academy of Science of the U.S.S.R. and by me, when I was an undergraduate at the City College of the City University of New York. [. . .] During the past decade we and others have continued to explore the meaning of randomness. The original formulations have been improved and the feasibility of the approach has been amply confirmed.

[. . .]

Formal Systems

It can readily be shown that a specific series of digits is not random; it is sufficient to find a program that will generate the series and that is substantially smaller than the series itself. The program need not be a minimal program for the series; it need only be a small one. To demonstrate that a particular series of digits is random, on the other hand, one must prove that no small program for calculating it exists.

It is in the realm of mathematical proof that Gödel's incompleteness theorem is such a conspicuous landmark; my version of the theorem predicts that the required proof of randomness cannot be found. The consequences of this fact are just as interesting for what they reveal about Gödel's theorem as they are for what they indicate about the nature of random numbers.

Gödel's theorem represents the resolution of a controversy that preoccupied mathematicians during the early years of the 20^{th} century. The question at issue was: “What constitutes a valid proof in mathematics and how is such a proof to be recognized?” David Hilbert had attempted to resolve the controversy by devising an artificial language in which valid proofs could be found mechanically, without any need for human insight or judgement. Gödel showed that there is no such perfect language.

Hilbert established a finite alphabet of symbols, an unambiguous grammar specifying how a meaningful statement could be formed, a finite list of axioms, or initial assumptions, and a finite list of rules of inference for deducing theorems from the axioms or from other theorems. Such a language, with its rules, is called a formal system.

A formal system is defined so precisely that a proof can be evaluated by a recursive procedure involving only simple logical and arithmetical manipulations. In other words, in the formal system there is an algorithm for testing the validity of proofs. Today, although not in Hilbert's time, the algorithm could be executed on a digital computer and the machine could be asked to “judge” the merits of the proof.

Because of Hilbert's requirement that a formal system have a proof-checking algorithm, it is possible in theory to list one by one all the theorems that can be proved in a particular system. One first lists in alphabetical order all sequences of symbols one character long and applies the proof-testing algorithm to each of them, thereby finding all theorems (if any) whose proofs consist of a single character. One then tests all the two-character sequences of symbols, and so on. In this way all potential proofs can be checked, and eventually all theorems can be discovered in order of the size of their proofs. (The method is, of course, only a theoretical one; the procedure is too lengthy to be practical.)

Unprovable Statements

Gödel showed in his 1931 proof that Hilbert's plan for a completely systematic mathematics cannot be fulfilled. He did this by constructing an assertion about the positive integers in the language of the formal system that is true but that cannot be proved in the system. The formal system, no matter how large or how carefully constructed it is, cannot encompass all true theorems and is therefore incomplete. Gödel's technique can be applied to virtually any formal system, and it therefore demands the surprising and, for many, discomforting conclusion that there can be no definitive answer to the question “What is a valid proof?”

Gödel's proof of the incompleteness theorem is based on the paradox of Epimenides the Cretan, who is said to have averred, “All Cretans are liars” [see “Paradox,” by W. V. Quine; Scientific American, April, 1962]. The paradox can be rephrased in more general terms as “This statement is false,” an assertion that is true if and only if it is false and that is therefore neither true nor false. Gödel replaced the concept of truth with that of provability and thereby constructed the sentence “This statement is unprovable,” an assertion that, in a specific formal system, is provable if and only if it is false. Thus either a falsehood is provable, which is forbidden, or a true statement is unprovable, and hence the formal system is incomplete. Gödel then applied a technique that uniquely numbers all statements and proofs in the formal system and thereby converted the sentence “This statement is unprovable” into an assertion about the properties of the positive integers. Because this transformation is possible, the incompleteness theorem applies with equal cogency to all formal systems in which it is possible to deal with the positive integers [see “Gödel's Proof,” by Ernest Nagel and James R. Newman; Scientific American, June, 1956].

The intimate association between Gödel's proof and the theory of random numbers can be made plain through another paradox, similar in form to the paradox of Epimenides. It is a variant of the Berry paradox, first published in 1908 by Bertrand Russell. It reads: “Find the smallest positive integer which to be specified requires more characters than there are in this sentence.” The sentence has 114 characters (counting spaces between words and the period but not the quotation marks), yet it supposedly specifies an integer that, by definition, requires more than 114 characters to be specified.

As before, in order to apply the paradox to the incompleteness theorem it is necessary to remove it from the realm of truth to the realm of provability. The phrase "which requires" must be replaced by "which can be proved to require," it being understood that all statements will be expressed in a particular formal system. In addition the vague notion of "the number of characters required to specify" an integer can be replaced by the precisely defined concept of complexity, which is measured in bits rather than characters.

The result of these transformations is the following computer program: "Find a series of binary digits that can be proved to be of a complexity greater than the number of bits in this program." The program tests all possible proofs in the formal system in order of their size until it encounters the first one proving that a specific binary sequence is of a complexity greater than the number of bits in the program. Then it prints the series it has found and halts. Of course, the paradox in the statement from which the program was derived has not been eliminated. The program supposedly calculates a number that no program its size should be able to calculate. In fact, the program finds the first number that it can be proved incapable of finding.

The absurdity of this conclusion merely demonstrates that the program will never find the number it is designed to look for. In a formal system one cannot prove that a particular series of digits is of a complexity greater than the number of bits in the program employed to specify the series.

A further generalization can be made about this paradox. It is not the number of bits in the program itself that is the limiting factor but the number of bits in the formal system as a whole. Hidden in the program are the axioms and rules of inference that determine the behavior of the system and provide the algorithm for testing proofs. The information content of these axioms and rules can be measured and can be designated the complexity of the formal system. The size of the entire program therefore exceeds the complexity of the formal system by a fixed number of bits c. (The actual value of c depends on the machine language employed.) The theorem proved by the paradox can therefore be stated as follows: In a formal system of complexity n it is impossible to prove that a particular series of binary digits is of complexity greater than $n + c$, where c is a constant that is independent of the particular system employed.

Limits of Formal Systems

Since complexity has been defined as a measure of randomness, this theorem implies that in a formal system no number can be proved to be random unless the complexity of the number is less than that of the system itself. Because all minimal programs are random the theorem also implies that a system of greater complexity is required in order to prove that a program is a minimal one for a particular series of digits.

The complexity of the formal system has such an important bearing on the proof of randomness because it is a measure of the amount of information the system contains, and hence of the amount of information that can be derived from it. The formal system rests on axioms: fundamental statements that are irreducible in the same sense that a minimal program is. (If an axiom could be expressed more compactly, then the briefer statement would become a new axiom and the old one would become a derived theorem.) The information embodied in the axioms is thus itself random, and it can be employed to test the randomness of other data. The randomness of some numbers can therefore be proved, but only if they are smaller than the formal system. Moreover, any formal system is of necessity finite, whereas any series of digits can be made arbitrarily large. Hence there will always be numbers whose randomness cannot be proved.

The endeavor to define and measure randomness has greatly clarified the significance and the implications of Gödel's incompleteness theorem. That theorem can now be seen not as an isolated paradox but as a natural consequence of the constraints imposed by information theory. In 1946 Hermann Weyl said that the doubt induced by such discoveries as Gödel's theorem had been "a constant drain on the enthusiasm and determination with which I pursued my research work." From the point of view of information theory, however, Gödel's theorem does not appear to give cause for depression. Instead it seems simply to suggest that in order to progress, mathematicians, like investigators in other sciences, must search for new axioms.