Infinite monkey theorem

From Wikipedia, the free encyclopedia

Given an infinite length of time, a typing monkey will eventually produce any given text, such as Hamlet, or almost surely type out all of Wikipedia.

The **infinite monkey theorem** states that a monkey hitting keys at random on a typewriter keyboard for an infinite amount of time will almost surely type a given text, such as the complete works of William Shakespeare. In fact the monkey not only will produce this text, but it will produce every possible finite text an infinite number of times. However, the probability that monkeys filling the observable universe would type Shakespeare's complete works is so astronomically low that it is for all practical purposes assumed to be impossible. It would require the expenditure of hundreds of megaliters (billions of megawatt-hours) per letter typed, and the sun will burn out before it is typed.

In this context, "almost surely" is a mathematical term with a precise meaning, and the "monkey" is not an actual monkey, but a metaphor for an abstract device that produces an infinite random sequence of letters. The theorem can be shown to be true or false, and the use of the "monkey metaphor" is that of French mathematician Emile Borel in 1913.[1] but the first instance may have been even earlier.

Variants of the theorem include multiple and even infinitely many typists, and the requirement that the typing text between an entire library and nothing but meaningless gibberish. Typing techniques such as those used in this idea from Aristotle's *On Generation and Corruption* and Cicero's *De natura deorum* (On the Nature of the Gods; through Blaise Pascal and Jonathan Swift, up to modern statements with their iconic simians and typewriters. In the 20th century, Claude Shannon, Borel and Arthur Eddington used the theorem to illustrate the timescales implicit in the foundations of statistical mechanics.

Contents

1 History
  1.1 Direct proof
  1.2 Infinite strings
    1.2.1 Infinite strings
  1.3 Infinite sequences
    1.3.1 Almost surely
    1.3.2 Correspondence between strings and numbers
  1.4 Infinite monkeys
    1.4.1 Statistical mechanics
    1.4.2 Origins and "The Total Library"
2 Applications
  2.1 Evolution
  2.2 Literary theory
  2.3 Random document generation
  2.4 Testing of random number generators
  2.5 Popular culture
3 See also
4 Notes
5 Bibliography
6 External links

Direct proof

There is a straightforward proof of this theorem. As an introduction, recall that if two events are statistically independent, then the probability of both happening equals the product of the probabilities of each happening independently. For example, if the probability of a randomly typed letter is a vowel is 0.4, and the chance of an earthquake in San Francisco on any particular day is 0.00003, then the chance of both happening on the same day is 0.4 × 0.00003 = 0.00012, assuming that they are indeed independent.

Suppose the typewriter has 50 keys, and the word to be typed is banana. If the keys are pressed randomly and independently, it means that each key has an equal chance of being hit, and that the probability of, for example, the first letter to be a b, and then the chance that the second letter typed is a a also 1/50, and so on. Therefore, the probability of the entire word banana being typed is

$$(1/50) \times (1/50) \times (1/50) \times (1/50) \times (1/50) \times (1/50)^6 = 1/15\,625\,000\,000\ldots$$

less than one in 15 billion, but not zero, hence a possible outcome.

From the theorem, the chance of not typing banana in a given block of 6 letters is 1 − (1/50)6. Because each block is typed independently, the chance X_n of not typing banana in any of the first n blocks of 6 letters is

$$X_n = \left(1 - \frac{1}{50^6}\right)^n$$

As n grows, X_n gets smaller. For an n of a million, X_n is roughly 0.9999, but for an n of 10 billion X_n is roughly 0.53 and for an n of 100 billion it is roughly 0.0017. As n approaches infinity, the probability X_n approaches zero, that is, by typing enough blocks, the chance of not typing banana can be made as small as desired.[2][3][4][5] and the chance of typing banana approaches 100%.

The probability calculated in this way and least one of infinitely many monkeys will produce a text as quickly as it would be produced by a perfectly accurate human typist copying it from the original. In this case $X_n = (1 - (1/50)^6)$ where X_n represents the probability that none of the first n monkeys types *banana* correctly on their first n blocks of 6 letters (or any other specified sequence of letters). When n is 1 billion, X_n is 0.9999, or 99.99%. But 1 billion typing monkeys will not wait for the last one. Suppose the typewriter produces 1 keystroke per second, and the monkeys are allocated 1 each. The last, for n going to infinity, is zero. So the probability of the word *banana* appearing at some point in an infinite sequence of keystrokes is equal to one.

Infinite strings

This can be stated more generally and compactly in terms of strings, which are sequences of characters chosen from some finite alphabet:

- Given an infinite string where each character is chosen uniformly at random, any given finite string almost surely occurs as a substring at some position.
- Given an infinite sequence of infinite strings, where each character is chosen uniformly at random, any given finite string almost surely occurs as a prefix of one of these strings.

Both follow from the second Borel-Cantelli lemma. For the second theorem, let E_k be the event that the kth string begins with the given text. Because this has some fixed nonzero probability of occurring, the E_k are independent, and the Borel sum diverges.

$$\sum_{k=1}^{\infty} \Pr(E_k) = \sum_{k=1}^{\infty} p = \infty,$$

The probability that infinitely many of the E_k occur is 1. The first theorem is shown similarly; one can divide the random string into nonoverlapping blocks matching the given fixed text, and make E_k the event where the kth block equals the desired string.[citation needed]

Probabilities

However, for physically meaningful numbers of monkeys typing for physically meaningful lengths of time the results are reversed. If there were as many monkeys as there are atoms in the observable universe typing extremely fast for trillions of times the life of the universe, the probability of the monkeys replicating even a single page of Shakespeare is unfathomably small.

Ignoring punctuation, spacing, and upper-lower case distinction, a monkey typing letters uniformly at random has a chance of one in 26 of correctly typing the first letter of Hamlet, then a chance of one in 676 (26 × 26) of correctly typing the first two letters. A typical page of Hamlet will require about 1 800 letters. Such a page has a chance of one in 26^1 800 = 19,928,148,895,209,409,152,340,197,376 (almost 2 × 10^28). In the case of the entire text of Hamlet, the probabilities are so vanishingly small as to be incalculable. The text of Hamlet contains approximately 130,000 letters.[6] Thus there is a probability of one in 3 × 10^181,000 to get the text right at the first trial. The average number of letters that needs to be typed until the text appears is also 3 × 10^181,000.[3][4] or including punctuation, 4.4 × 10^185,700 typed letters.[7]

If one uses a modern typewriter keyboard with at least 100 randomly typed keys will on average first see Hamlet after 97 billion years on the average (and twice that much longer until, they would still need a bill for greater amount of time – more than three hundred and sixty thousand orders of magnitude longer – to have even a 1 in 10 chance of success, to have a 50% chance they would need 2 years, and to have a 99% chance, they would need 664 years).[citation needed] Much of Hamlet will be typed in much less time. The best strategy for the monkeys is not to try to predict what has been typed already. In the field of information theory, this is expressed as the fact that maximum entropy of typing means that the probability of Hamlet is therefore zero in any operational sense of an event,[...] and the statement that the monkeys will eventually succeed is meaningless, or, in the language of measure theory, false.

In fact there is less than a one in a trillion chance of success that such a universe made of monkeys could type any particular document a mere 79 characters long.[citation needed]

Almost surely

The probability that an infinite randomly generated string of text will contain a particular finite substring is 1. However, this does not mean the substring's absence is impossible, even though the infinite string will almost surely contain all such finite substrings. The infinite string where every character is a b is an example. The probability that the infinite string begins with five a characters is 1/32 × 1/32 × 1/32 × 1/32 × 1/32 = 1/32^5 = 1/33 554 432, so the probability that it does not is 33 554 431/33 554 432 = 0.999…969…875. Similarly, the probability that it does not begin with ten a's is 1/32^10 = 1/1 152 921 504 606 846 976, and so forth. The probability that it does not begin with any finite string of all b's is the product of all these numbers, or 0.999…969…875 × 0.999…999…968…750 × 0.999…999…999…992…187 × … = 0.288…788…095…086…189…119…684…925…291…528…992…584…226…031…792…179…201…953…125. However, this infinite product is equal to the probability that the infinite string does not consist entirely of b's. As shown above, this probability is 1. Similarly, the probability that the infinite string does not contain the substring abcdefghijklmnopqrstuvwxyz is 1, and the probability that it does not contain the substring abcdefghijklmnopqrstuvwxyzabcdefghijklmnopqrstuvwxyz is 1, and so forth. In this way, the infinite string where every character is a b is shown to be not only possible, but almost surely not. In fact, the infinite string almost surely will contain every possible finite substring, and almost surely will have every possible finite substring appear infinitely many times, almost surely will have every possible finite substring appear with the expected frequency, and almost surely will have every possible finite substring appear with every possible frequency.

Correspondence between strings and numbers

In a simplification of the thought experiment, the monkey could have a typewriter with just two keys: 1 and 0. The infinitely long string thusly produced would correspond to the binary digits of a particular real number between 0 and 1. For example, a typewriter that produces the sequence of characters 011010111... would correspond to the decimal expansion 0.011010111... Because the set of rational numbers, which can be expressed as fractions of whole numbers, is countable, almost all real numbers are irrational, and so correspond to non-repeating sequences. Some of the rational numbers do correspond to repeating sequences, such as 011111111... = 0.1. Such a typewriter would almost surely produce a random real number.

Meaningful, there is an uncountably infinite set of strings which do not end in such repetition, these correspond to the irrational numbers. These can be sorted into meaningful categories, such as those which are normal, which almost surely all finite strings appear, and so on. Almost all real numbers are normal, and almost all strings created from Hamlet, but which contain every other possible string of any length, and with equal distribution of such substrings. These irrational numbers are called normal. Borel showed that almost all real numbers are normal, and so almost any random real number is normal. Although this means that almost all real numbers have the property that no matter what finite sequence one is looking for, one will appear at some point in the number, almost all such numbers are also such that no matter what finite sequence one is looking for, it will appear infinitely many times and with the expected frequency. Almost all real numbers also have the property that no matter what finite sequence one is looking for, it will appear at some point in the number, almost all such numbers are also such that no matter what finite sequence one is looking for, it will appear infinitely many times and with every possible frequency. This property is also satisfied by the digits of pi or e, which is just another way of saying that almost all real numbers are normal.

The same applies to the event of typing a particular version of Hamlet followed by endless copies of itself, or Hamlet immediately followed by all the digits of pi, these numbers correspond to particular real numbers. It is almost certain that a random real number will be normal, and therefore any finite string will almost surely be part of its fate, any finite string followed by any other finite string will almost surely be followed by it, even though the result will almost surely be irrational.

This is an extension of the principle that a finite string of random text has a lower and lower probability of being a particular string the longer it is (though all specific strings are equally improbable, even if they happen to be meaningful when typed by a monkey). However, the probability that the string will eventually be typed, or the gambler's fallacy.[citation needed] This is a weaker form of the theorem that almost all strings will have the property that no matter what finite string one is looking for, it will appear at some point in the string, and almost all strings will have every possible finite substring appear with the expected frequency, but almost all strings will have every possible finite substring appear infinitely often and the strings present in it so that total strings appear with frequency infinitely and thus the original theorem is correct