Tales that numbers tell

NUMBERS LEAVE AN INTERESTING TRAIL THAT STATISTICIANS CAN SNIFF AND FOLLOW, WRITES **S ANANTHANARAYANAN**

Natural processes have characteristics that get disturbed when there is motivated action. Numerical markers of "normality" can then signal anything unusual, in a way that those responsible may find it difficult to conceal or where normal detection may take more time or effort.

Professor Shankar Venkatagiri, mathematician and member of the decision sciences and information systems area at the Indian Institute of Management, Bangalore during the annual meeting of the Indian Railway Accounts Service at the Rail Wheel Factory in Bangalore, described features of numbers and the way that fraud detection agencies, as well as the world of business, make use of patterns to detect threats and opportunities.


Shankar Venkatagiri

A little known property of numbers that arise in natural processes is that the first digit of these numbers is not uniformly distributed, but tends to be low, like one, two, or three rather than high, like eight or nine. For example, the height of mountains in feet, or of buildings in millimeters, would be numbers, typically, from a few hundreds to many thousands. Now, the first digits of actual numbers, which may be 12,335 or 8,322 or 6,345, for instance, are one, eight and six, in these examples. Would there be a tendency for this first digit to be preferentially in some range, rather than be uniformly distributed from one to nine?

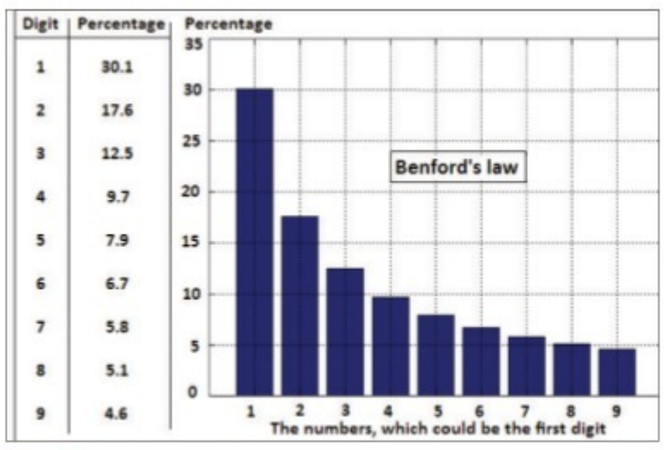
While one would normally expect that all the digits from one to nine are equally likely to be the first digit in long lists of numbers, which cover many orders of magnitude (rather than stay in a limited range), Venkatagiri explained that there was a counter-intuitive law which said this was not so. The Benford's Law, he said, was that one was the first digit as often as 30 per cent of the time and nine appeared at the first place only 4.6 per cent of the time. The percentage of times that all the digits arise and a graph of how they fall, from 30.1 per cent to 4.6 per cent, is shown in the picture.

This rule about how the first digit is more often not a lower number has been verified in a great many instances, like the area of lakes in a district, population sizes, birth or death rates, electricity bills and commodity prices. It will be noticed that these are numbers that arise "naturally" or without a design that affects the first digit.

This would not be the case, say, in the height, in inches, of the average 12-yr-old, which would be between 50 and 60 inches, with five as the most common first digit. The area of a lake, in square metres, or populations, for instance, could be anything from a few hundred to thousands or even hundreds of thousands.

While this feature of the first digit being low numbers rather than high ones would seem surprising at first, it can be understood with a little analysis. The number 1 we can see, occurs as the first digit, first, by itself, then from the numbers, 10 to 19 and then from 100 to 199, and so on. The number, 'two,' similarly, occurs as the first digit first by itself, then from 20 to 29, and then from 200 to 299 and so on. The same sequence is true of 'three.' What we notice is that 'one' gets repeated first within nine numbers of its first appearance, and then after just the next 80 numbers. But the number, 'two,' has to wait for 18 numbers before the first repetition and then for 170 numbers before the second repetition. The wait before repetition keeps extending, like this, for the numbers, 'three' to 'nine'—it being from 100 to 899, or 799 numbers before the second repetition of 'nine' as first digit. That is ten times the wait of only 80 numbers for 'one.' When we reach higher numbers, the distance between successive appearances of the higher digits extends exponentially or the greater the number, the more marked the higher separation of occurrence.

Digit	Percentage
1	30.1
2	17.6
3	12.5
4	9.7
5	7.9
6	6.7
7	5.8
8	5.1
9	4.6



This is the reason that in a collection of numbers that cover a wide range, the distribution of first digits follows Benford's law. A direct application is to capture the numbers generated in a system and to keep checking if the first digits follow Benford's law. One kind of fraud in banks, for instance, is with the daily interest calculated on balances. The fraudster manipulates the system to add some small figure to the interest worked out on a thousand accounts and transfer the total amount to a separate account that the fraudster can access.

If the bank had a "Benford's law check system" in place, it would regularly inspect the first digits, and also some other features of the numbers in the bank's records. If all is well, the numbers follow Benford's law. But if there is a systematic change being made, this would reflect in how the first digits appear and alert the bank's auditors.

A similar application could be in the data collected through surveys. Figures that arise from honest surveys show features that do not appear in fictitious data or even in data where there have been errors in sampling. Applying statistical checks on the numbers could then show that corrections need to be applied. This kind of check could be vitally important in statistical quality checks or checks that ensure safety.

Venkatagiri went on to describe other uses of capturing and analysing numbers, like in maintaining law and order, public health, scheduling material movement or public transport. An area of great use was in advertising and marketing.

The clicks on pages of search engines like *Google*, or in the course of purchases on the Internet were captured and made use of to send specifically sele-cted advertisement messages to individual users, based on their browsing behaviour.

Venkatagiri also described how *Google* may be able to detect an epidemic before the health administration of a state came to know of it. Particularly in countries where medical help or dispensing was expensive, the occurrence of symptoms was revealed first in the way Internet users carried out searches rather than in the records of their visits to doctors or hospitals.

Google could hence use its data to alert governments of apparent rise in the incidence of body pain and fever, for instance, to set in motion a process of investigation and containment.

THE WRITER CAN BE CONTACTED AT
response@simplescience.in

Benford's law

THE logarithm of a number is the power to which the number 10 needs to be raised to result in that number. The use of the logarithm is that if we know the logs of numbers, we can multiply them simply by adding the logarithms. The logarithms of all numbers, starting from zero to many decimal places, have hence been worked out and tabulated. These tables were exceedingly useful for work in science and technology before we had calculators and computers.

In 1881, astronomer Simon Newcomb noticed that it was the early pages of a book of logarithms that were most worn with use, rather than the later pages. This suggested that the numbers that arose in the course of scientific work started with lower digits like one or two, rather than the larger digits like eight or nine.

The physicist, Frank Benford again noticed this phenomenon in 1938 and he tested numbers that arose in different domains, "like surface areas of 335 rivers, the sizes of 3,259 US populations, 104 physical constants, 1,800 molecular weights, 5,000 entries from a mathematical handbook, 308 numbers contained in an issue of *Reader's Digest*, the street addresses of the first 342 persons listed in *American Men of Science* and 418 death rates." (*Wikipedia*)

Benford then established the rule, which has been named after him.

CIRCULAR REPLICATION

TAPAN KUMAR MAITRA EXPLAINS HOW RING CHROMOSOMES ARE FORMED IN DIFFERENT ORGANISMS

Ring chromosomes have been described in a number of organisms from viruses to humans, but it is now evident that as a group they are similar only in a superficial, morphological sense. Some are naturally circular, others can be artificially produced while others must be considered sporadically occurring aberrations.

The circularity of the lambda bacteriophage chromosome is due to the complementary redundancy of single polynucleotide sequences, which terminate the chromosome, thus permitting it to exist either as a ring or a linear structure. The chromosome of Escherichia coli, which is one mm in length and possesses a molecular weight of about two billion, replicates in a circular form although the manner of unwinding of the double helix of DNA during replication still remains a puzzle. The genetic map of such chromosomes would, of course, be similarly circular, for the topography of the map and the physical nature of the chromosome must be consistent with each other. However, such consistency is not found in the T2 or T4 bacteriophage — the map is circular, but the chromosome is a linear structure, 56 μ in length.

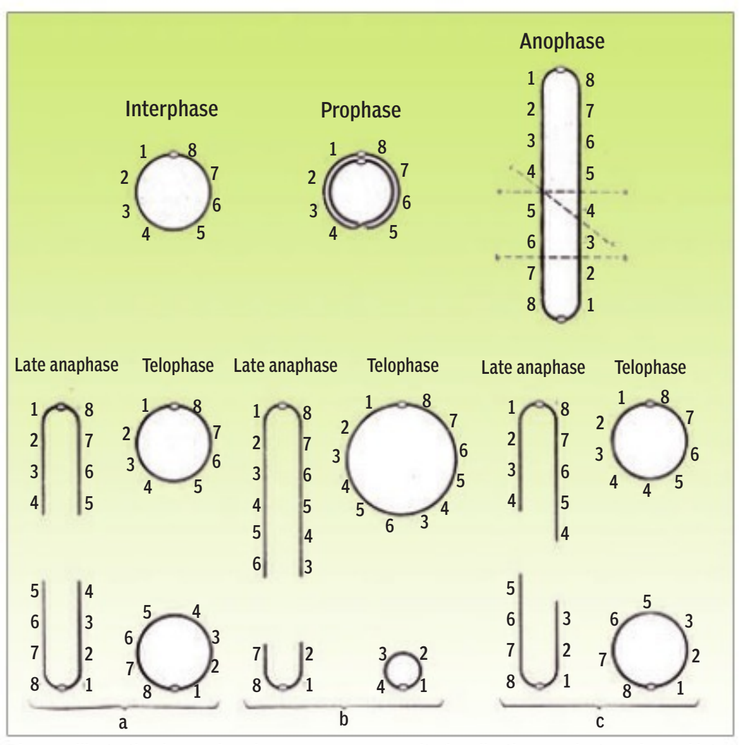


Diagram illustrating how a ring chromosome, as a result of replication, can be transformed into a double-sized, di-centric ring. The ring would either be hung between the poles at anaphase or, as is illustrated, fragment into two pieces, each with a centromere. The broken ends can re-fuse in telophase, to repeat the same cycle in the next cell division. Some ring chromosomes particularly small ones can separate cleanly at anaphase without forming a dicentric ring or becoming interlocked (not illustrated), but how this is accomplished is not known.

The T2 chromosomes are circularly and genetically permuted and possess a terminal redundancy amounting to about two percent of the total length. Partial enzymatic digestion followed by annealing produces circular chromosomes, providing proof of the terminal redundancy. Therefore, the circularity which occurs naturally in lambda bacteriophage can be artificially induced in the T2 virus.

The ring chromosomes of higher organisms, however, must be viewed as aberrant types. They have been studied in Drosophila, maize, and humans, and although they can be perpetuated in certain experimental stocks, they would, in the long run, tend to be eliminated in any situation where they compete with their normal linear homologues. This fact is evident from their behavior, for although they can reproduce in such a manner as to give two un-entangled rings, they also give rise to interlocked rings and to double-sized, di-centric single rings.

The fact that cleanly separating rings can be formed at all is, by itself, surprising if it is assumed that the DNA molecule replicates semi-conservatively. In maize, however, small ring chromosomes freely separate a fair percentage of the time while larger rings have a greater tendency to be interlocked or to form double-sized di-centric rings.

Ring chromosomes in humans have involved the X chromosome, some other members of the six-12 group of chromosomes, and chromosomes 17 or 18. All are associated with phenotypic abnormalities, and it seems reasonable to suppose that the loss of chromatin accompanying formation of the rings and the irregularity of transmission of these chromosomes are responsible for the observed abnormalities.

THE WRITER IS ASSOCIATE PROFESSOR, HEAD, DEPARTMENT OF BOTANY, ANANDA MOHAN COLLEGE, KOLKATA, AND ALSO FELLOW, BOTANICAL SOCIETY OF BENGAL, AND CAN BE CONTACTED AT
tapanmaitra59@yahoo.co.in

It's all in your blood

SCIENTISTS HAVE FOUND THAT 'VAMPIRE' TREATMENTS THAT MAKE PEOPLE LOOK YOUNGER AND LIVE LONGER DON'T ACTUALLY WORK. **ANDREW GRIFFIN** REPORTS

The dream of using "vampire" techniques to look younger and live longer might have delighted goths and the vain alike. But scientists have warned them not to get too excited. They have said that the treatments, which involve taking young blood and injecting it into older people in an attempt to rejuvenate people's bodies, don't actually work.

But the research has found that old blood does in fact perform an important part of the ageing process and contributes to declining health. That might mean that there are treatments for older blood that helps relieve those effects. to 199, and so on. The number, 'two,'

The same team had previously found that giving older mice younger blood seemed also to give them a new lease of life. They published their work in a study in 2005.

It quickly led to talk of vampires and hopes that similar techniques could be used to improve the lives of people, too. But the study wasn' t able to control the flow of blood precisely enough to be sure about how the effect was working. For the new study, precise measurements were made of the way old mice responded to young blood, and vice-versa. It showed that young blood made little or no difference to indicators of ageing and health in older mice.

In contrast, young mice receiving older blood experienced significant deterioration of their tissues and organs. The rapid changes occurred within 24 hours and affected multiple tissues including muscle, liver and brain.

Lead scientist Irina Conboy, from the University of California at Berkeley, US, said, "Our study suggests that young blood by itself will not work as effective medicine. It's more accurate to say that there are inhibitors in older blood that we need to target to reverse ageing."

Mice in the original experiment not only shared blood but also organs, so that older animals benefited from young lungs, immune systems, hearts, livers and kidneys.

The new study, reported in the journal *Nature Communications*, removed the influence of shared organs. In a series of trials, blood was exchanged between an old mouse and a young one until each animal had half its blood from the other. Various indicators of ageing were then tested including liver growth, scarring and fattiness, brain cell development affecting learning and memory, and muscle strength and repair.

The most telling results came from the brain tests. Older mice showed no improvement in neural regeneration from stem cells after receiving young blood but young mice given old blood saw a more than two-fold drop in brain cell replacement.

Conboy added, "Under no circumstances did young blood improve brain neuro-genesis in our experiments. Old blood appears to have inhibitors of brain cell health and growth, which we need to identify and remove if we want to improve memory."

THE INDEPENDENT

PLUS POINTS

Mars on earth

Scientists are hoping to discover evidence of past life on Mars by studying a series of hot springs in Northern Chile.

Steve Ruff and Jack Farmer from Arizona State University found remarkable similarities between ancient hot springs in the Gusev crater on Mars and the El Tatio hot springs near the edge of the Atacama Desert.

The Atacama Desert is one of the highest places on earth and is seen as the most Mars-like area on earth. The springs themselves are at an altitude of 4,300 metres above sea level. El Tatio experiences sub-freezing temperatures every night and endure lots of ultraviolet light during the day through thin dry air — conditions very similar to that on Mars.

The site near the Martian Home Plate plateau was first discovered by Nasa's Spirit rover and show finger-like structures that form in hot spring deposits through biological and non-biological processes.

Ruff said, "We went to El Tatio looking for comparisons with the features found by Spirit at Home Plate. Our results show that the conditions at El Tatio produce silica deposits with characteristics that are among the most Mars-like of any silica deposit on earth."

The ASU team are studying bio-signatures at the Chilean springs that could also be present on Mars which could help explain the planet's natural history.

Bio-signatures are traces that indicate the presence of life right now or in the past. While fossils are the most obvious form of bio-signatures, others include organic molecules trapped in rocks or structures made up of compacted microorganisms.

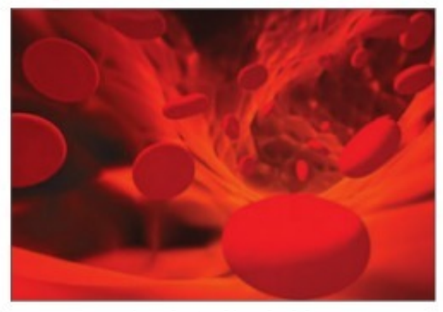
As no fossils have yet been found on Mars, any evidence of past or present life is expected to be microscopic.

MATT PAYTON/THE INDEPENDENT

Gene transport

Haemophilia A is a rare genetic condition in which the blood does not clot properly.

Scientists in Singapore are studying the use of cord-lining stem cells to transport a specific gene into the body that will then go on to produce a clotting protein known as factor VIII, which patients lack.



In animal studies led by Professor Kon Oi Lian of the National Cancer Centre, Singapore, stem cells carrying factor VIII were introduced into mice with the condition. Scientists found that the animals began to produce the protein — the mice to which the gene was introduced also bled less when their tails were clipped.

"We were able to show, at least in a small animal model, that these cord-lining epithelial cells not only secreted factor VIII, but were able to mitigate, though not cure, the haemophilic mice," said Kon.

Adeno-associated viruses, which are not known to cause disease in humans, are now the commonly used vector for transporting the gene, but there have been some challenges.

One problem is that people infected before by the virus already produce neutralising antibodies against it, which means they would not be able to receive the gene.

"The AAV also does not insert itself into the genome, so it is possible the effect will wear off after a while," added Kon. Other viral vectors studied that insert themselves into the patient's genome have had adverse side-effects.

Kon noted that cord-lining stem cells could provide an alternative method. Her team started studies on dogs last year.

SAMANTHA BOH/THE STRAITS TIMES