

The Great Aadhaar Game: Don't be Arbitrary, be Cleverly Random

thewire.in/government/the-great-aadhaar-game-dont-be-arbitrary-be-cleverly-random



Representational image. Photo: PTI/Reuters.

opinion

Government

The Centre, through arbitrary subversion of legalities, is turning us into random number generators.



Rajendran Narayanan



Government

Law

05/Apr/2018

Prologue: A wise man told me this article should have appeared on April 1.

Loaded: 6.62%

Remaining Time -10:00

The Supreme Court's interim order of March 13 appears incongruous to the earlier unanimous verdict on the right to privacy and is at odds with the constitutional right to equality. The deadline to link Aadhaar with bank accounts and phones have been indefinitely extended while they aren't extended for any welfare schemes for the poor such as old age pensions and the rural job guarantee scheme, MGNREGA.

This implies that Aadhaar will be voluntary for the rich while being mandatory to access critical survival measures for the poor. In other words, sadly, right to privacy seems to crucially hinge on the class position of a citizen. In essence, it reinstates the Orwellian cliché that some of us are more equal than others.

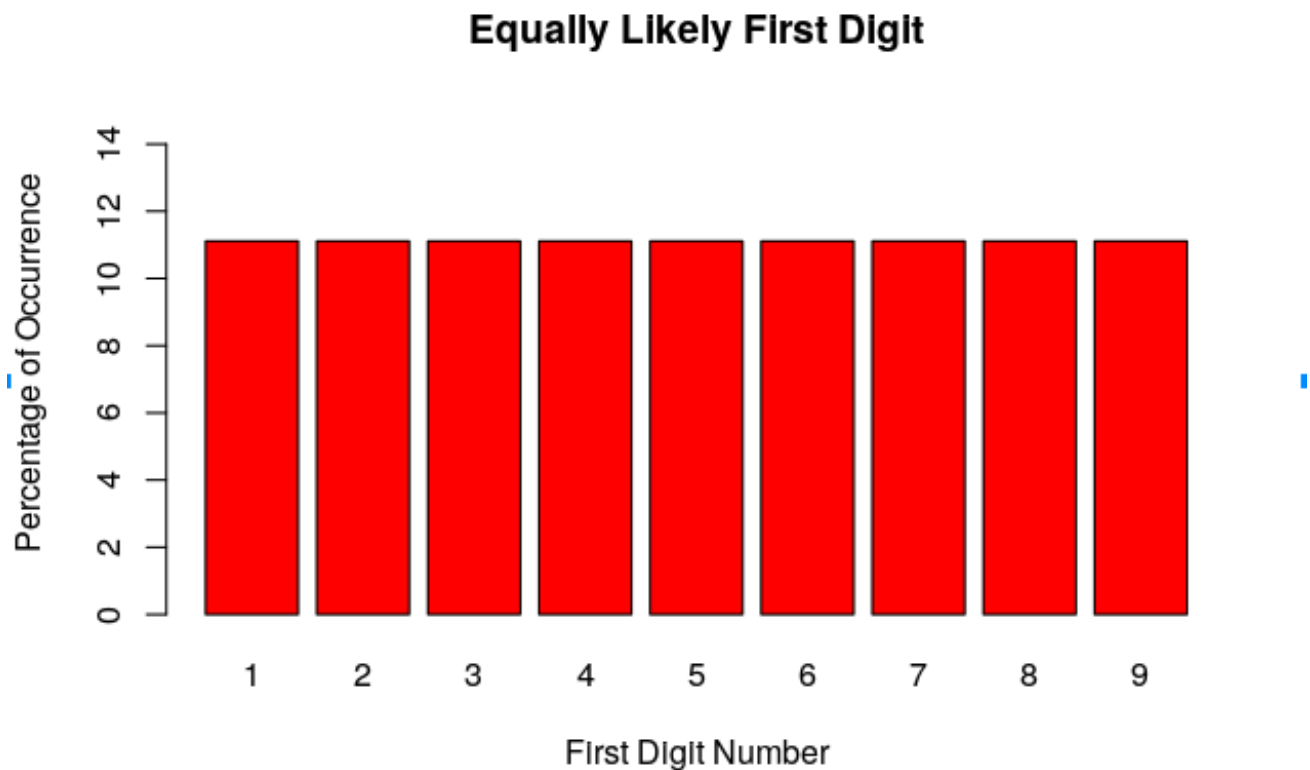
Be that as it may, this differential access to "freedom" implies that we can happily choose to file our taxes without Aadhaar. There is a small catch though. For some inexplicable and rather bizarre reason, nobody can file their returns electronically without an Aadhaar number. The income tax returns website does not let you file your returns unless you furnish a 12-digit number. This seems to be in contravention of the March 13 interim order of the apex court. As such, even conscientious taxpayers, who do not have Aadhaar, are reportedly being forced to provide an arbitrary 12-digit number as a proxy for Aadhaar.

What does a diligent taxpayer who does not have an Aadhaar do in such a Catch 22 situation? Anecdotal evidence suggests that such taxpayers are trying out their hand at imagining a random Aadhaar number so that the system lets you submit your returns. The Centre, through arbitrary subversion of legalities, is turning us into random number generators. It is not the intention to suggest that anyone *should* do that, but it is in this context that it might be instructive learn about Benford's law (also called First Digit Law). Perhaps it may help to learn how to imagine a "good" large random number? Let's get into some mathematical melodrama to learn one way to fake randomness in situations that deal with large numbers (12 digits, for example).

So, what is Benford's law? Benford's law states that in many real life situations when data is observed in high magnitude, more numbers begin with the digits 1, 2, or 3, as opposed to 7,8, or 9. For example, consider a 250-page book. There are about 109 pages whose page numbers begin with the digit 1 (pages 1,11-19,100-199), 60 pages beginning with the digit 2 while only 10 pages beginning with the digit 9. As another example — consider the number

of people in various age groups. As a proportion of the population, chances are there are more people with ages that start with the digit 1 than with the digit 9. At first, Benford's law appears to be counter-intuitive because if the distribution of the first digit in datasets are truly random, then each digit between 1 and 9 have the same chance, i.e., 1 out 9 (11.1%) of being the first digit, i.e., be uniformly spread as in Figure 1.

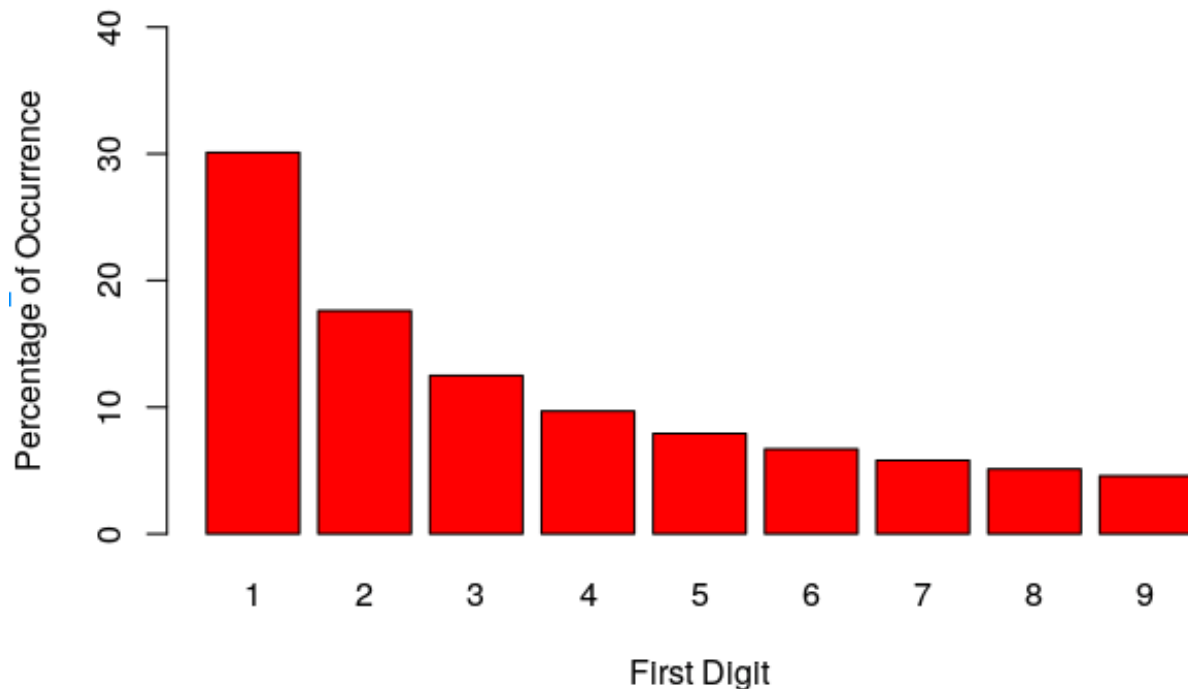
Figure 1: If each digit had the same chance of being the first digit



However, the frequency distribution of beginning digits in many data settings indicate otherwise (see Figure 2). In particular, this holds true for datasets that grow exponentially (doubles or triples) such as bacterial colony data, populations of cities, and not to forget income tax data. It is empirically observed that, on an average, the first digit is 1 in about 30% of the cases, it is 2 in about 17.5% of the cases, and it is 3 in about 12% of the cases. The systematic pattern of the first digit being further away from 1 continues and number 9 is the first digit in only about 5% of all the data points.

Figure 2: Distribution of first digit in datasets varying in large orders of magnitude

Benford's Law



The fact is a consequence of digits in such datasets being non-uniformly spread on the original measured scale but uniformly distributed on a logarithmic scale – used to rescale exponentially growing data where the interest is more on the number of digits. Very simply, logarithms (\log) tell us the number of digits after the first digit in a number. So $\log 10 = 1$ and $\log 1000 = 3$. It denotes the power to which a number is to be raised to get another number.

This phenomena was first highlighted by the astronomer Simon Newcomb in 1881 when he observed that the initial pages of the logarithm tables were yellower and more smudged than the latter pages. This led him to conjecture that the logarithms of numbers beginning with the digit 1 were more prevalent than numbers beginning with higher digits. The same principle was later tested and verified across several datasets by the physicist, Frank Benford of GE Research in 1938. He tested the hypothesis by looking at surface areas of rivers, US population, numbers in *Reader's Digest* magazine, street addresses of hundreds of people etc.

In the 1990s, researcher Mark Nigrini used Benford's law to track accounting frauds by reviewing sales figures, insurance claims and reimbursement claims. For example, owing to a policy threshold of \$100,000, a fraudster wrote several cheques to himself just below this threshold, i.e., with a first digit in the cheque amount being 9. This obvious departure from

the expectation of 5% of numbers beginning with the digit 9 was a red signal to catch accounting fraud. Benford's law has been similarly used to look at fudged numbers in income tax returns and is permissible as evidence for criminal cases in the US.

However, Benford's law does not apply to data that have been assigned. Consequently, this law will not apply to the oft-compared twin (quite incorrectly so) of Aadhaar-Social Security Numbers (SSN) in the US. The 9-digit SSN are not randomly generated but have a well-defined structure of assignation.

The questions for us is – given that Aadhaar is a 11-digit random number (the 12 is technically not), will Aadhaar numbers also follow Benford's law? In case it does, then just to follow the Supreme Court order, and be a diligent citizen, should innocent people without Aadhaar have had to learn all this jugglery just to be “random”?

Recently, the central government/UIDAI's lawyer claimed in the Supreme Court that Aadhaar data centres are kept protected inside walls that are “13 feet tall and 5 feet thick”. Given such *strong* and exemplary data security and data protection features, would one wish to be caught just because one sprinkled equal number of ones and nines to create a make-believe 12 digit number? Isn't it absurd that we have to fake to establish honesty? Till the crimes of the state's arbitrariness are brought to proper punishment, the so-called *misdemeanour* of randomness may continue.

Moral of the story: Don't be arbitrary like UIDAI and the current government. Be cleverly random.

Rajendran Narayanan is a faculty member at Azim Premji University, Bangalore. The views expressed are personal.

This is an updated version. The initial version of the article appeared in the *Business Standard* on March 31, 2018.

The founding premise of The Wire is this: if good journalism is to survive and thrive, it can only do so by being both editorially and financially independent. This means relying principally on contributions from readers and concerned citizens who have no interest other than to sustain a space for quality journalism. For any query or help write to us at support@thewire.in

I would like to contribute

Once

Monthly

Yearly

Select amount

₹200

₹1000

₹2400

Type an amount