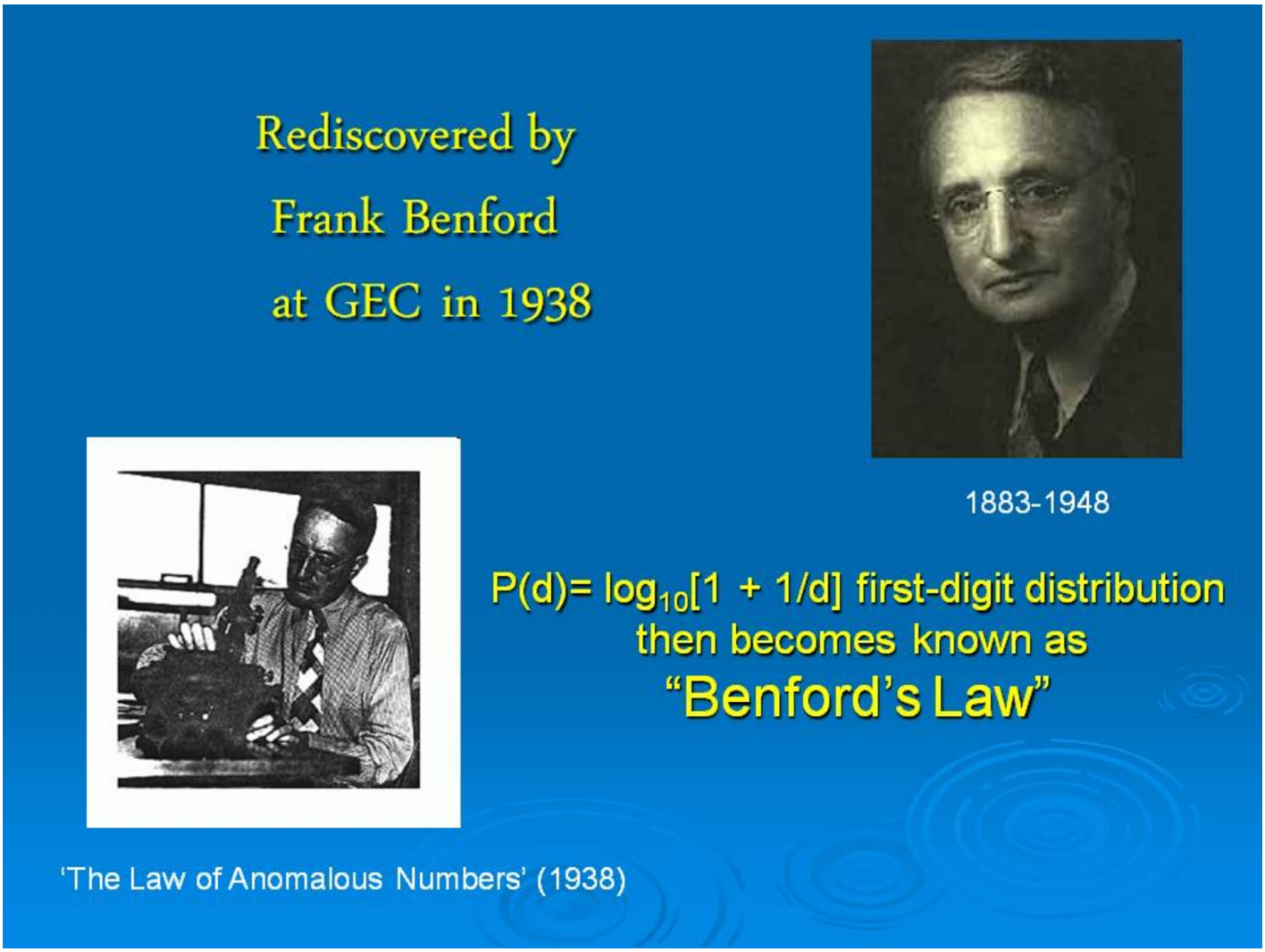
GRESHAM COLLEGE
Founded 1597

Benford's Very Strange Law Transcript

Date: Tuesday, 1 February 2011 - 1:00PM

Location: Museum of London



Gresham Lecture, 1 February 2011

Benford's Very Strange Law

Professor John Barrow

Today, we are going to take a look at something rather unusual that always intrigues me whenever I return to it. We will see how this unusual probability distribution that is sometimes known as Benford's Law arose, why it is a reasonable thing to expect to be true about numbers that you encounter in real life, and then some of the unusual applications that have been made of the idea to accountancy and other matters of everyday life.

The first person to notice that there was something rather unequal about numbers, or certain bits of them, was Simon Newcomb. Newcomb tends to be famous for making one or two rather bad predictions. He sometimes bears the blame for saying that he believed that heavier than air flying machines were impossible, and he also said in 1888 that mankind was probably reaching the limit of all it could know about astronomy. For all that, he was actually rather a good astronomer, and, like many intellectuals of the period, was very much a polymath and someone who worked in all sorts of different areas of science and mathematics. In 1881, he wrote a little paper – I think it appeared in Atlantic Mathematical Monthly or something like that, which had the title "On the Frequency of Use of Different Digits in Natural Numbers". It sounds a bit strange, and he said he was motivated by taking a careful look at his log tables.

When I looked up log tables on the internet - I found I no longer owned a set of log tables – I was offered a picture of a table made out of logs!

Anyway, I then moved on and found mathematical log tables. People beyond a certain age will remember these. They are a complete mystery to anyone under a certain age. If you study mathematics at school today, there are certain things, like log tables and a slide rule, which you will never see and have probably never heard of. When I was at school, they were the only thing that you really needed in order to be able to carry out any calculation.

Newcomb noticed that his log tables were always slightly more worn for the smaller numbers, and you can see it on this page. They are more thumbed down for the lower numbers than they are for the higher numbers. This made him suspicious that there was a different distribution, other than just the uniform one, for the occurrence of particular digits in the first digit place of numbers that you tended to come across and use. He noticed that people were tending to look up the logarithms of numbers that began with 1 rather more often the numbers that began with 2 and 3, and certainly much more often the numbers that began with 8 and 9. This sounds very odd, and led to what I call Newcomb's Law.

He said that the 10 digits from 0 to 9 do not occur with equal frequency, and he said that this is evident to anyone making use of logarithmic tables and you notice how much faster the first pages wear out than the last ones. He then made an observation rather more particularly about the first digit of a number – i.e. take the digit after the decimal place, for example, if the number was 3.1414, the first digit would be 1. He said it was more often 1, that first digit, than any other digit, and then the next most common one was 2, then 3, and then 4, and so on, up to 9. He then noticed that there was a good reason for this, and that the probability distribution of the mantissae (the fractional parts of the logarithms), was that they were equally probable. He was saying that if you changed the numbers into 10^x, as it were, the distribution of those x's in the power would be uniform.

Let us look at this in a number of different ways, and we shall prove it later on.
He was saying that the data on the first digits are evenly spread out if you use a logarithmic scale. Log 2 is 0.3010, so he was saying for 30% or 30.1% of numbers, the first digit is a 1, and then you work down. The probability for the digit being 2 is 18%, for 3 is 12% and so on. So his evident law is that p(d) = [log(d+1) – log(d)]/log(10) – log(1)]. This is in log base 10 so log(1) = 0 and log (10) = 1 so it simplifies to log(1 + 1/d).

That is the rule. So the probability that the first digit is a 1 = log(1 + 1) = log(2) = 0.3; the probability that it is a 2 = log(1 + ½) = log(1.5) and so on all the way down to the probability of 9 = log(1 + 1/9).

You can see that something odd is going on here, that the distribution of the digits is not uniform. The situation is rather like having a clock face, with the numbers 1 to 10 marked around that dial, and as you go to larger and larger numbers, of course, you first meet 1, and then you go all the way round to 9. Then you come back to 10, 11 etc and you start to get more numbers beginning with 1 again, and as you go round and round and round this circle, eventually, in the long run, the distribution of the appearance of any digit is uniform, but it is uniform on this circular distribution, not on the simple counting linear distribution.

If the number starts to get large and towards 9, if x is small, then log(1+x) = x. That is a Taylor expansion. So, when d is getting up towards 9, the probability begins to look like 1/d. Here is the distribution as a bar chart with log 2, log 1.5 all the way down to log 10/9.

The obvious thing you notice about this is that it is not as you would have guessed at all. So, if someone had stopped you in the street, and you did not know about this, and asked you the likelihood of any number that you start to pick out of the newspaper or pick in the lottery or somewhere else, at random, having a first digit equal to 2 or equal to 3, you would have thought that the probabilities would all be the same, and that they would all be

just equal to about 0.11111 or 1/9.

Newcomb's claim was that that was not the case at all. This is what his distribution tells you for these probabilities, so, to 2 decimal places, 30% of numbers that you just pick at random in some way, uncontrolled way, will begin with 1. So you can see you might win a lot of money if you knew this and other people did not.

Nobody took much notice of this, in the sense of writing papers about it. I think most people felt fairly obviously that it was reasonable at the end of the 19th Century and that there was not much more to say so it was just forgotten.

It was rediscovered in 1938 by Frank Benford, who was an electrical engineer working at GEC in the United States. He made essentially the same discovery, but he decided to test it out much more systematically by gathering up all sorts of collections of data and numbers and seeing whether the result was really true. He derived this same distribution - he did not know of Newcomb's work before – and it is as a result of this that it became known as Benford's Law. Sometimes I will call it the Newcomb-Benford Law; sometimes I will call it Benford's Law for short. His paper was called "The Law of Anomalous Numbers". It is probably not a good title in that the numbers are not anomalous at all.

This is a collection of some of the things that Benford gathered. He gathered nearly 20,000 pieces of numerical data and tested them out to see if they followed this law. So, using his atlas and geographical information, he looked at lists of the the areas, of all the leading lakes in the United States, the American continent and in Europe, and he wrote down the numbers for the areas, and then he looked at the first digit after the decimal place of these areas. This was a big sample of information. It was information of the same sort, that had units of area, and he found, lo and behold, 31% of the river areas had 1 as their first digit. He then showed the others, all the way down to 9. He looked at the baseball averages for past seasons, systematically recorded in the US, rather like Wisden gives you information about cricket, averages and so forth. He started looking through those and seeing what the batting averages and the bowling averages and so forth were for players, back as far as records were kept, and again, he had the same trend: 32.7% of those numbers began with a 1.

Next, he looked through random magazines and newspapers, and every time a number appeared, other than the page numbers, he made a note of it. You can see, again, it was not exactly 30, but there was pretty much the same trend. It is rather spooky. Then he did something systemic: he worked out all the powers of 2 – 2, 4, 8, 16, 32 and 64 to generate a huge list of numbers that are just powers of 2. Well, he got, bang on, 30% of them started with a 1, and down to 5% started with 9. He looked in the world of physics and engineering at the back of his book with all the tables about the strengths of materials, the coefficients of expansion of different metals, all the standard properties or whatever, and made a frequency distribution from those. Again, he arrived at exactly the same type of answer; for 30.6% of them, first digit was 1. He next looked at the half-life of all sorts of radioactive materials against alpha decay. This was a real physical situation, governed by intrinsically random quantum mechanical process but also a systematic aspect of nuclear physics. Finally, these are the predictions that you would expect if you followed the Benford-Newcomb Law, and you can see the agreement is good enough.

There is obviously something going on here. So, is there some secret, universal law of nature, as it were, behind this?

Here are some other trials that people tried later. If you take the street addresses of all the people in your mailing list and look at the digits of the street numbers 30% of those begin with ones. Here is some of the information in Benford's original list, brought up-to-date. So Benford's Law is the blue bar; numbers in newspapers, the green one; the Census data in 1990 in the United States the red one; and the financial information from Dow Jones, day by day, is the orange one. You can see, again, it shows exactly the same type of trend. To rub it in, here is another one: consumer price index variations, the Census, the birth rate, the areas of particular countries in the world, lottery winning tickets. Here is Benford's Law, with the solid line, and this dotted line is what would happen if the probabilities were 0.11.

Where does this magic 30% come from? There are all sorts of ways of looking at this, and all sorts of people have had different goes at trying to produce the most convincing way of looking at it.

I think this is quite a good way of thinking about it. Suppose that you are picking raffle tickets and you're interested again in the probability that you have a raffle ticket where the first digit is 1. So it might be the ticket with 1 on it, or 11, or 111, or 109 etc. Well, you can see, if there are just 2 tickets in the raffle the probability of having a 1 as the first digit is ½ – it is this ticket, but not that ticket. If there are 3 tickets in a raffle, the probability is 1/3, and so on. If you go up and you have 9 tickets, numbered 1 to 9, the probability of having a 1 is 1/9. However, once you have 10 tickets, things change. A new number has appeared as there is another number that begins with a 1, and the probability jumps up to 1/5 because there are now 2 candidates out of 10..2/10 = 1/5.

If we add an 11, 12 and 13, you can see the probability beginning with 1 is steadily going up each time, because each of the new numbers begins with a 1. Once you get to 20 tickets, it will start dropping again because all the new numbers do not begin with a 1. So the probability of having a ticket that begins with a 1 is that, at first the probability is 50%; as you go up to 9 tickets, the probability drops; and then, when you add the 10th ticket, it

goes up to 20%; you keep adding, all the way up to 20, it will go up all the way up to here; and then, once you go between 20 tickets onward, it starts to drop. However, when you get to 100, okay, it starts to go up again. You can see, as you add more and more and more numbers, you get this funny increasing oscillatory change in the probability. So what this means, obviously, is the probability of having a ticket that begins with 1 depends on the number of tickets.

You might take the average. Suppose there is a huge number of tickets, hundreds and hundreds of thousands. The average of this funny spiky curve is, if you take the average carefully, 30.1%. So, you can understand from this picture how you get such a probability.

If we were to run through the same argument for the 2s it would be slightly different because, with the 2 tickets, we would have a probability of 50% again, and then, as you went all the way up to 10, you would still have no new case. You would have to go all the way up to 12 before you got a new one, and then, when you started with the 20s, you would have a rising probability. If you carried out the same argument, we would get the other probability, log(1 + ½). Thus, this is a simple argument as to why we expect this type of pattern to work.

However, there are one or two other ways of looking at that distribution of Newcomb and Benford's. It has to have certain properties. Take the areas of the lakes in the world that Benford took. That has units. When he did it, he probably worked them out in square miles. However, if there is a universal distribution about digits, it should not care whether you measure the areas in square miles, square kilometres or square inches or whatever. That means that any universal distribution governing such things must be invariant under multiplying it by some overall factor, every piece of data by some number, which would be the conversion factor, say between square kilometres to square miles, to square inches.

What that means in sort of more formal mathematical language, mathematicians would say the distribution has to be scale invariant. So if you scale all the numbers by multiplying or dividing them by the same factor, you must get the same overall result. This means that if there is a distribution for these first digits of numbers which have units then the probability distribution of a number multiplied by a constant – call it k – has to look the same as the probability distribution for the numbers, just multiplied by some overall factor. Therefore, the shape of the distribution does not change. It is just like looking at it under a magnifying glass or under a microscope – it just scales. Therefore, $p(kx) = f(k)p(x)$

It is a probability distribution, so $\int p(x)\,dx = 1$ over all the x values. That means that if you substitute x for kx $\int p(kx)\,dx = 1/k$ over all the x values. Therefore, as $p(kx) = f(k)p(x)$, $\int p(kx)\,dx = 1/k = f(k)\int p(x)\,dx$ so $1/k = f(k)$.

We can now set this up as a differential equation. We know that $p(kx) = f(k)p(x)$ and $f(k) = 1/k$ so $d(p(kx)/dk = -p(x)/k^2$ so $xp'(kx) = -p(x)/k^2$. If we have $k = 1$, then $xp'(x) = -p(x)$ and the solution is that $p(x) = 1/x$ ($p'(x) = -1/x^2$). Therefore, the answer to this little problem is that the function $p(x) = 1/x$ so $p(kx) = 1/kx = 1/k \times 1/x$.

So there is a requirement that the distribution be scale invariant - it does not matter in which units you measure the quantities that you line up in your tables, as long as you measure them all in the same units. The probability distribution that we are looking for is 1/x.

There is a detail here in that if you did try to check this, if the distribution went all the way from 0 to infinity, we would have a problem here, because this would be log x, and the log of infinity is not 1, it is infinity. In practice, when we do these problems we do not go all the way to infinity. The areas of the lakes have a finite maximum size, and they do not go to zero either.

So, this tells us that $p(d) = (\int_d^{d+1} 1/x\,dx)/(\int_1^{10} 1/x\,dx)$. The integral underneath is to normalise it to get a probability. So the top integral is the probability that it is in this little interval, and the bottom integral is the probability that you are in the whole interval from 1 to 10. $\int 1/x\,dx = \log x$ so you get $(\log(d + 1) – \log(d))/\log 10 – \log 1 = \log((d + 1)/d) = \log(1 + 1/d)$. That was the result that Newcomb guessed, in a sense, just by rather advanced commonsense, and Benford was also able to work out. Here, we see it arises as a requirement that things be scale invariant under changing the units in which you do the problem.

Other people after Newcomb, and Newcomb himself wondered what would happen with the second digit. It is clear that there would be a distribution for the second digit. In fact, there is a distribution for any digit. So the probability, for example, of the second digit being a 3 is just the sum of the probabilities of the number being 1.3, 2.3, and 3.3, all the way up to 9.3. They are the possible ways in which you could have a combination of first and second digits so that the second digit was a 3. If we look at the logarithmic spread of the possibilities from 1 up to 9 and up to 10 here, the probability of having these situations is the sum of the bits between 1.3 and 1.4, 2.3 and 2.4, all the way up. The fraction, for example, in this little niche 1.4 to 1.3 = $(\log(1.4) – \log(1.3))/(\log 10 – \log 1) = \log(14/13)$. Therefore, you just do the same calculation to find the probability of being in the next little interval and then the next one and add them all together. There is nothing very difficult about this; it is just a bit laborious.

We already know the probability of the first digit = $\log(1 + 1/d)$. The second digit will be the sum of the probabilities that you are in these little intervals. Newcomb himself understood that so you can work out, if you wish, the third digit and so on.

The interesting thing is that this reveals what statisticians would call the joint distribution – the probability the second digit has a particular value given that the first digit has another value; this probability is not independent. So, the probability that the second digit has a particular value and the first digit has a particular value is not just equal to the product of the separate probabilities for the first and second digits. The second digit probability depends on the first digit probability.

The expression you get is: $P(1^{st} = d_1,..., k^{th} = d_k) = \log[1 + (\sum_{i=1}^{k} d_i \times 10^{k-i})^{-1}]$. Suppose you had the factional part of Pi, so 3.14, call it 0.314. The probability that the 1^{st} digit is 3, the 2^{nd}digit is 1 and the 3^{rd} digit is 4 means it is $\log(1 + 1/314) = 0.0014$.

You can start asking what would happen if there was no conditioning. The unconditional probability that the second digit is 1 is 0.109. The probability that the first digit is 1 is 0.30. The conditional probability that it the second digit is 1, given that the first is 1 is a little bit bigger, 0.115.

As the number of digits increases, the dependence of, for example, the second digit on the first falls off, and they become increasingly de facto independent. So, rather quickly, as the number of digits in the number gets bigger, this formula shows that the probability just becomes equal to 1 over the number of numbers. Thus, it becomes uniform because the log of 1 + something small tends to something small.

We have seen that Benford's distribution gets picked out by the criteria scale invariance. It is also picked out by another requirement that you might imagine you would want. We have been talking about taking numbers out of the newspapers, looking them up randomly, and there are always cases where we would be using our ordinary base 10 arithmetic. However, you might wonder whether there is some universal distribution governing first digits. You ought to get a similar sort of result using any base arithmetic and that turns out to be the case. I will not go through the algebra, but you can show that the only distribution that you could get which would have the same form, irrespective of base, is this one which has the arithmetic base b, where b need not be 10. The first digit distribution approaches the Benford form of $\log_b(1 + 1/d)$, but with the logarithm to the base b, because, whereas we had had that log 10 - log 1 as the denominator, you now have log b - log 1.

Here is a graph showing all the probabilities varying on the base. As you see, there is a collection of distributions that have the same basic shape, but with slightly different numerics.

Although these two invariances pick out Benford's form of the law, you might wonder why there should be a distribution like this at all. It does not really tell you why there is a distribution or under what conditions you would expect there to be such a distribution. If you look far enough on the web, you will find mysterious speculations that Benford's Law is the secret of the universe and some deep secret about numbers.

Not everything follows Benford's Law. Therefore, it is interesting to try and understand why that might be, because it might be the key to understanding why it is so ubiquitous. Here is a collection of numbers taken from United States tax returns, and the first digit of sums of total income and you can see it follows the leading digit Benford Law pretty well. However, this is a collection of results from a random number generator. You can find such things on the web, and sometimes people need such numbers for various purposes - you can see that these numbers do not follow Benford's Law.

Here is another example of something else that does not follow Benford's Law, which we met a when we had a look at continued fractions. You remember, we could expand any real number as a continued fraction in this staircase of descending fractions, and this collection of whole numbers can be listed and they give a unique definition of the real number. If the number is irrational then this continued fraction never ends. Remarkably, there is a probability distribution which holds for the distribution of these numbers which is true for almost every number in a rather precise sense. That probability distribution is $P(k) = [\ln(1 + 1/(k(k+2))]/\ln 2$. You remember the Benford one was $P(k) = \ln(1 + 1/k)$. This is not the same, but it shares many features. Again, the 1 is the most common digit – 41% of continued fraction numbers are 1s, 17% are 2s, and so on. This distribution, when k becomes very large, again, is the log of 1 plus something that gets small, so $\ln(1 + x)$ looks like x when x is small. Therefore, when k is big, $P(k) = \ln(1 + 1/k^2)$, so this distribution looks like $1/k^2$ so this is slightly steeper than Benford. The probabilities fall off much faster. So this is another non-Benford distribution.

It is possible to understand in a fairly clear way what you need of a list of numbers in order that they're going to be Benford distributed.

We want the data that we line up, with all these digits on, to measure the same thing. So, we take a collection of areas, a collection of prices, but we do not mix all those things together. Therefore, they all have the same type of unit, so that when you scale them, you would only change square kilometres into square miles or acres or something like that. You also want the distribution of the numbers to have no sort of built-in maximum/minimum, which would skew the number in some way. So if you took people's heights, it probably would not be a good candidate. There is, in effect, a maximum for adults, and there would be a minimum, typically, in any sample. So you want distributions that really have an enormous range.

Furthermore, you do not want the numbers to be assigned in some pattern which skews the story, so phone numbers are no good, because if you drop the 0, my phone code is 1223, and so they tend to begin with 1s.

There is a part of the country where they begin with 2s, in London, and so on. So those sorts of numbers which are orderly in some pre-assigned way are not useful for this purpose. You also want the underlying distribution to be pretty smooth across all the possible numbers, so you do not want someone to not allow any numbers to begin with 5 because they are suspicious of numbers where the first digit begins with 5. There might be someone who picks their lottery numbers because they have some secret prejudice against certain numbers – they never put in 13s or 12s.

Also, a good rule, in real distributions, is that there tend to be more small things than large ones. So, with the lake areas, for example, there are more small lakes than there are large lakes. If you were looking at the sizes of grains of sand and rock on the Earth, there are more small ones than there are large ones. Most important of all, and in some ways linked to some of these things, we want the data to span a large range of values. We want to be able to take the numbers in the data and re-express them as 10 to the power something, so instead of 100, we would say 10^2. We want those powers to span a very large range, certainly a range much bigger than the distance between individual digits. So, if all the data was sort of crammed in around 10 to the power 1 up to 10 to the power 1-and-a-bit, there would not be enough range to explore this probability distribution. You say that the distribution must be broad rather than narrow – it must spread across a very wide range of values.

Here are two examples. Here is a broad distribution with numbers along the bottom expressed in this logarithmic scale. The red area is exploring the numbers where the first digit is 1: so numbers between 1 and 2; the ones from 10 up to 19; 100 to 199 and so on. Each time you see a red area, it is the probability that the first digit is a 1, and the blue, the probability that the first digit 8, has been put in here as another example. It is a bit like when we took the averages earlier on.

Here is another type of distribution for all the numbers which is very narrow. I think this was probably something like the sizes of electoral districts in California which are engineered so that they are all the same size. Thus, in some sense, there is a maximum and they have been deliberately chosen so there are about the same number of voters. That is why it is very narrow. Again, you can see, the red represents the probability of the first digit being 1 and the blue does the same for the 8.

The first distribution is so broad that it covers many multiples of the coloured bands, so you have a chance for the distribution to be reached. The second distribution does not. Secondly, the probability is equal to the area under the curve, and for the first one, the area is nicely proportional to the width along the bottom, but for the second distribution, the width is not a good guide to the area because the width is so small compared with the height, and so the area under this curve is dominated by the height of the curve, which is determined by this very steep, narrow distribution. Whereas, with the first distribution, because it is broad and rather shallow, the width, this probability, is a good guide to what this area will be and the relative probabilities are well indicated by the relative widths of these bands.

So a good example, would be the case for numbers that you were generating from incomes of large numbers of people from populations of countless countries of the world where you have a very large range of sizes; whereas things like human heights, IQ scores, populations in small electoral districts that are well-defined and engineered in size, will have this narrow and a non-Benford distribution.

Here are some pictures of data which are similar or dissimilar to the previous case. So, the top distribution represents population of countries of the world. The first graph shows population against number, while the second graph shows frequency against the log of the population, which is nice and broad. The distribution is much broader than the individual slots here. And so, when you look at the leading digits of this, it's a broad distribution, you expect it to approach Benford, and it does, okay, so the first digit distribution follows Benford.

This next example is of jackpot numbers for the US Powerball game. Again, if you look at the frequency of various first digits arising on different numbers and the range of numbers it has a nice broad distribution. All the possibilities are explored pretty randomly, uniformly, and accordingly, you get a Benford distribution for the first digits. However, here are the little Congressional districts, in California, in the previous distribution. When you first look at the frequency versus population size, it looks a bit odd – it is quite flat - then when you look at this on the logarithmic scale, it is really very narrow indeed. They are all roughly the same size, and so there is not a probability distribution that spans lots of decades on the log scale. There is no room for the probability distribution to be explored, and so, the leading digit distribution does not look like Benford at all. This one at the bottom shows population sizes of cities in California. Again, it is quite good. There is a steep, but reasonably broad, distribution on the logarithmic scale, and things follow Benford fairly well.

We now have an understanding really as to how and when you should expect to find a Benford type distribution. You want a nice broad underlying distribution, no skewing with maxima or minima or special limitations.

What does this tell you about winning lotteries? It is fairly obvious really. Here is a small lottery in America, the Massachusetts Numbers Game. You just bet on a four-digit number, which is generated at random, and all the people that guess the right number share the jackpot. I am not sure whether they had a second round where they then put you in and you picked another number or something like that. You can see that there are going to be lots of winners here – it is not like our lottery. You are not going to win much because you tend to share a lot, and as our children used to say when they came home from nursery, "It's nice to share, Dad." I was usually eating an ice cream or something like that at the time. So, what should your strategy be? Well, you might have in

mind that you do not want to share the jackpot, or you want to share it with as few other punters as possible, so you want to try and pick a number that is less likely to be picked by others. Suppose that all the other entrants pick their numbers from one of these Benford-like collections of numbers drawn from their experience. We assume they do not pick them at random, but from the newspaper or something like that. They do not pick them from birthdays - that would not be good.

So, what would then happen? Well, they are most likely to pick numbers the first digit of which is 1 and we know the probability of the second digit, third digit and so on. So you should pick numbers that are the opposite of the Benford probabilities; you should pick numbers that begin with 8s and 9s because they would have the lowest probability of occurring in these naturally occurring collections of numbers that people may be picking out of their experience.

There was a curious psychological experiment carried out by Theodore Hill, who was a mathematician, some years ago in the US, just looking at what numbers people chose in psychological experiments when they were asked to pick numbers at random, with many digits in them. Curiously, the probability was weighted very much towards people picking numbers that began with small digits. I do not know why that is, but it gives you more confidence that if you pick numbers that begin with large digits, you have an even better chance of winning a prize outright.

As with most things in mathematics, you can have too much of a good thing, and once people find one rule, the first reaction to it is to ask whether it can be generalised and be part of something bigger or better than has more applications. We have seen that Benford's Law is really a probability distribution for an underlying process, where P(x) is proportional to 1/x, so as x gets big, there is a smaller probability, and that is the distribution that has those nice scale invariant and base invariant properties. $\int_d^{d+1} x^{-1}\,dx = [\ln x]_d^{d+1} = \ln(d+1) - \ln(d) = \ln(d+1/d) = \ln(d/d + 1/d) = \ln(1 + 1/d)$ which is this Benford distribution for the first digit being d.

However, you could forget about Benford for the moment and be interested in processes where the probability is not 1/x, but $1/x^a$, particularly in the cases where a does not equal to 1.

$P(x) = C\int_d^{d+1} x^{-a}\,dx = C[x^{1-a}/1-a]_d^{d+1} = C(((d+1)^{1-a} - d^{1-a})/1-a) = (10^{1-a} - 1)^{-1}((d+1)^{1-a} - d^{1-a})$.

We normalise the probability so the integral = 1 and C is the normalisation. If a = 2, for example, as this is now a rather steep distribution, the probability of the first digit as 1 would be 0.56. If you want to know the probability of it being 3, it lowers to 9%, and it's 1% for d = 9. So, you can see the general effect, that when the distribution is steeper than Benford's, there is even more chance of the first digit being 1.

The interest in this is that one application of a Benford-like approach is to think about the prime numbers. So whenever you have something about numbers, some probability theorem, you always wonder what it tells you about prime numbers. The probability distribution for the first digits in the prime numbers does not quite follow Benford's Law. It follows the law $a(N) = 1/[\log N - a]$ where $a = 1.1 \pm 0.05$. The value of a depends on the quantity of numbers looked at. So if you look at all the prime numbers from 1 up to n, the first digit distribution is very well-fitted by this rule here. So, $a = 1/[\log N - 1.1]$ which is just a little offset.

Here is the graph for the distribution for the prime numbers. This tells you that as n becomes very large these first digits are evenly distributed so there is no bias towards any number at all, but the approach to that distribution is very well-ordered. The percentage is on the x axis and n is on the y axis. The solid curve is that mathematical rule, and the dotted line is the data from the prime numbers up to very large n. You can see, it really does fit extremely well with a value of a = 1.1. So the prime numbers do not quite follow Benford's law – that would be a = 1.0 – but they follow a similar type of law, as you approach their uniform probability distribution.

I want to show you how people have rather imaginatively thought that they might use Benford's Law to detect fraud with false accounting and spurious tax returns. So, if you believe that Benford's Law is rather ubiquitous so long as your distribution is broad and is not biased in particular ways, it might give you a way of telling whether people had manipulated or doctored the collections of numbers that they are giving to you as pieces of evidence or their financial records.

This idea occurred first to Mark Nigrini, who was a graduate student at the University of Cincinnati back in 1992, and his PhD thesis was called "The detection of income evasion through an analysis of digital distributions". First of all, he looked at huge numbers of IRS, the American revenue service, files to explore the extent to which Benford's Law was well followed, and, overall, in this huge mass of data, it followed beautifully accurately. So the large mass of income tax returns in the US to which he had access followed Benford's Law for the first digit, and then for the second digit very well.

Then he had access to data which was known to be fraudulent, and this was mixed in with other real data by the people who he was working with, so it was a double-blind experiment, and he correctly identified the false returns by the fact that they did not follow Benford's first digit law. They tended to have rather odd distributions. Here is Benford's Law in the blue with 30% having a 1, and 17% a 2 and so on. Here is the good tax data, or so he believes, which are the green bars, and you see they follow Benford really rather well – it was broad

distribution with huge amounts of data spanning all the range of values. The brown distribution was the fraudulent data – it has loads more 5s and 6s and really sticks out like a sore thumb. Finally, here is the random data, which is pretty much 0.11 all the way across.

So this excited the Chief Financial Investigator of Brooklyn District who gave a press conference and had all sorts of meetings with Nigrini. This is his account of a test with Nigrini's theory with 7 cases of fraud. "We used them as a test of Nigrini's computer program. It correctly spotted all 7 cases as involving probable fraud."

Then something rather interesting happened because, years after this all became rather well-known, the Clintons, after much controversy and criticism about their financial affairs, published their tax return, and so Nigrini thought that he should analyse their tax returns. The 64 million dollar question is: what did he find? It was all okay - the Clintons' tax return data followed Benford's Law rather beautifully - disappointing, but true.