Benford's law, Zipf's law, and the Pareto distribution

3 July, 2009 in expository, math.PR, math.ST, non-technical | Tags: Benford's law, Pareto distribution, randomness, Zipf's law

A remarkable phenomenon in probability theory is that of *universality* – that many seemingly unrelated probability distributions, which ostensibly involve large numbers of unknown parameters, can end up converging to a universal law that may only depend on a small handful of parameters. One of the most famous examples of the universality phenomenon is the central limit theorem; another rich source of examples comes from random matrix theory, which is one of the areas of my own research.

Analogous universality phenomena also show up in *empirical* distributions – the distributions of a statistic X from a large population of "real-world" objects. Examples include Benford's law, Zipf's law, and the Pareto distribution (of which the Pareto principle or *80-20 law* is a special case). These laws govern the asymptotic distribution of many statistics X which

(i) take values as positive numbers;

(ii) range over many different orders of magnitude;

(iiii) arise from a complicated combination of largely independent factors (with different samples of X arising from different independent factors); and

(iv) have not been artificially rounded, truncated, or otherwise constrained in size.

Examples here include the population of countries or cities, the frequency of occurrence of words in a language, the mass of astronomical objects, or the net worth of individuals or corporations. The laws are then as follows:

Benford's law: For $k = 1, \ldots, 9$, the proportion of X whose first digit is k is approximately $\log_{10} \frac{k+1}{k}$. Thus, for instance, X should have a first digit of 1 about 30% of the time, but a first digit of 9 only about 5% of the time.

Zipf's law: The n^{th} largest value of X should obey an approximate power law, i.e. it should be approximately $Cn^{-\alpha}$ for the first few $n = 1, 2, 3, \ldots$ and some parameters $C, \alpha > 0$. In many cases, α is close to 1.

Pareto distribution: The proportion of X with at least m digits (before the decimal point), where m is above the median number of digits, should obey an approximate exponential law, i.e. be approximately of the form $c10^{-m/\alpha}$ for some $c, \alpha > 0$. Again, in many cases α is close to 1.

Benford's law and Pareto distribution are stated here for base 10, which is what we are most familiar with, but the laws hold for any base (after replacing all the occurrences of 10 in the above laws with the new base, of course). The laws tend to break down if the hypotheses (i)-(iv) are dropped. For instance, if the statistic X concentrates around its mean (as opposed to being spread over many orders of magnitude), then the normal distribution tends to be a much better model (as indicated by such results as the central limit theorem). If instead the various samples of the statistics are highly correlated with each other, then other laws can arise (for instance, the eigenvalues of a random matrix, as well as many empirically observed matrices, are correlated to each other, with the behaviour of the largest eigenvalues being governed by laws such as the *Tracy-Widom law* rather than Zipf's law, and the bulk distribution being governed by laws such as the semicircular law rather than the normal or Pareto distributions).

To illustrate these laws, let us take as a data set the populations of 235 countries and regions of the world in 2007 (using the CIA world factbook); I have put the raw data here. This is a relatively small sample (cf. my previous post), but is already enough to discern these laws in action. For instance, here is how the data set tracks with Benford's law (rounded to three significant figures):

k	Countries	Number	Benford prediction
1	Angola, Anguilla, Aruba, Bangladesh, Belgium, Botswana, Brazil, Burkina Faso, Cambodia, Cameroon, Chad, Chile, China, Christmas Island, Cook Islands, Cuba, Czech Republic, Ecuador, Estonia, Gabon, (The) Gambia, Greece, Guam, Guatemala, Guinea-Bissau, India, Japan, Kazakhstan, Kiribati, Malawi, Mali, Mauritius, Mexico, (Federated States of) Micronesia, Nauru, Netherlands, Niger, Nigeria, Niue, Pakistan, Portugal, Russia, Rwanda, Saint Lucia, Saint Vincent and the Grenadines, Senegal, Serbia, Swaziland, Syria, Timor-Leste (East-Timor), Tokelau, Tonga, Trinidad and Tobago, Tunisia, Tuvalu, (U.S.) Virgin Islands, Wallis and Futuna, Zambia, Zimbabwe	59 (25.1%)	71 (30.1%)
2	Armenia, Australia, Barbados, British Virgin Islands, Cote d'Ivoire, French Polynesia, Ghana, Gibraltar, Indonesia, Iraq, Jamaica, (North) Korea, Kosovo, Kuwait, Latvia, Lesotho, Macedonia, Madagascar, Malaysia, Mayotte, Mongolia, Mozambique, Namibia, Nepal, Netherlands Antilles, New Caledonia Norfolk Island, Palau, Peru, Romania, Saint Martin, Samoa, San Marino, Sao Tome and Principe, Saudi Arabia, Slovenia, Sri Lanka, Svalbard, Taiwan, Turks and Caicos Islands, Uzbekistan, Vanuatu, Venezuela, Yemen	44 (18.7%)	41 (17.6%)
3	Afghanistan, Albania, Algeria, (The) Bahamas, Belize, Brunei, Canada, (Rep. of the) Congo, Falkland Islands (Islas Malvinas), Iceland, Kenya, Lebanon, Liberia, Liechtenstein, Lithuania, Maldives, Mauritania, Monaco, Morocco, Oman, (Occupied) Palestinian Territory, Panama, Poland, Puerto Rico, Saint Kitts and Nevis, Uganda, United States of America, Uruguay, Western Sahara	29 (12.3%)	29 (12.5%)
4	Argentina, Bosnia and Herzegovina, Burma (Myanmar), Cape Verde, Cayman Islands, Central African Republic, Colombia, Costa Rica, Croatia, Faroe Islands, Georgia, Ireland, (South) Korea, Luxembourg, Malta, Moldova, New Zealand, Norway, Pitcairn Islands, Singapore, South Africa, Spain, Sudan, Suriname, Tanzania, Ukraine, United Arab Emirates	27 (11.4%)	22 (9.7%)
5	(Macao SAR) China, Cocos Islands, Denmark, Djibouti, Eritrea, Finland, Greenland, Italy, Kyrgyzstan, Montserrat, Nicaragua, Papua New Guinea, Slovakia, Solomon Islands, Togo, Turkmenistan	16 (6.8%)	19 (7.9%)
6	American Samoa, Bermuda, Bhutan, (Dem. Rep. of the) Congo, Equatorial Guinea, France, Guernsey, Iran, Jordan, Laos, Libya, Marshall Islands, Montenegro, Paraguay, Sierra Leone, Thailand, United Kingdom	17 (7.2%)	16 (6.7%)
7	Bahrain, Bulgaria, (Hong Kong SAR) China, Comoros, Cyprus, Dominica, El Salvador, Guyana, Honduras, Israel, (Isle of) Man, Saint Barthelemy, Saint Helena, Saint Pierre and Miquelon, Switzerland, Tajikistan, Turkey	17 (7.2%)	14 (5.8%)
8	Andorra, Antigua and Barbuda, Austria, Azerbaijan, Benin, Burundi, Egypt, Ethiopia, Germany, Haiti, Holy See (Vatican City), Northern Mariana Islands, Qatar, Seychelles, Vietnam	15 (6.4%)	12 (5.1%)
9	Belarus, Bolivia, Dominican Republic, Fiji, Grenada, Guinea, Hungary, Jersey, Philippines, Somalia, Sweden	11 (4.5%)	11 (4.6%)

Here is how the same data tracks Zipf's law for the first twenty values of n, with the parameters $C \approx 1.28 \times 10^9$ and $\alpha \approx 1.03$ (selected by log-linear regression), again rounding to three significant figures:

n	Country	Population	Zipf prediction	Deviation from prediction
1	China	1,330,000,000	1,280,000,000	$+4.1\%$
2	India	1,150,000,000	626,000,000	$+83.5\%$
3	USA	304,000,000	412,000,000	-26.3%
4	Indonesia	238,000,000	307,000,000	-22.5%
5	Brazil	196,000,000	244,000,000	-19.4%

6	Pakistan	173,000,000	202,000,000	-14.4%
7	Bangladesh	154,000,000	172,000,000	-10.9%
8	Nigeria	146,000,000	150,000,000	-2.6%
9	Russia	141,000,000	133,000,000	$+5.8\%$
10	Japan	128,000,000	120,000,000	$+6.7\%$
11	Mexico	110,000,000	108,000,000	$+1.7\%$
12	Philippines	96,100,000	98,900,000	-2.9%
13	Vietnam	86,100,000	91,100,000	-5.4%
14	Ethiopia	82,600,000	84,400,000	-2.1%
15	Germany	82,400,000	78,600,000	$+4.8\%$
16	Egypt	81,700,000	73,500,000	$+11.1\%$
17	Turkey	71,900,000	69,100,000	$+4.1\%$
18	Congo	66,500,000	65,100,000	$+2.2\%$
19	Iran	65,900,000	61,600,000	$+6.9\%$
20	Thailand	65,500,000	58,400,000	$+12.1\%$

As one sees, Zipf’s law is not particularly precise at the extreme edge of the statistics (when n is very small), but becomes reasonably accurate (given the small sample size, and given that we are fitting twenty data points using only two parameters) for moderate sizes of n.

This data set has too few scales in base 10 to illustrate the Pareto distribution effectively – over half of the country populations are either seven or eight digits in that base. But if we instead work in base 2, then country populations range in a decent number of scales (the majority of countries have population between 2^{23} and 2^{32}), and we begin to see the law emerge, where m is now the number of digits in binary, the best-fit parameters are $\alpha \approx 1.18$ and $c \approx 1.7 \times 2^{26}/235$:

m	Countries with $\geq m$ binary digit populations	Number	Pareto prediction
31	China, India	2	1
30	”	2	2
29	“, United States of America	3	5
28	“, Indonesia, Brazil, Pakistan, Bangladesh, Nigeria, Russia	9	8
27	“, Japan, Mexico, Philippines, Vietnam, Ethiopia, Germany, Egypt, Turkey	17	15
26	“, (Dem. Rep. of the) Congo, Iran, Thailand, France, United Kingdom, Italy, South Africa, (South) Korea, Burma (Myanmar), Ukraine, Colombia, Spain, Argentina, Sudan, Tanzania, Poland, Kenya, Morocco, Algeria	36	27
25	“, Canada, Afghanistan, Uganda, Nepal, Peru, Iraq, Saudi Arabia, Uzbekistan, Venezuela, Malaysia, (North) Korea, Ghana, Yemen, Taiwan, Romania, Mozambique, Sri Lanka, Australia, Cote d’Ivoire, Madagascar, Syria, Cameroon	58	49
24	“, Netherlands, Chile, Kazakhstan, Burkina Faso, Cambodia, Malawi, Ecuador, Niger, Guatemala, Senegal, Angola, Mali, Zambia, Cuba, Zimbabwe, Greece, Portugal, Belgium, Tunisia, Czech Republic, Rwanda, Serbia, Chad, Hungary, Guinea, Belarus, Somalia, Dominican Republic, Bolivia, Sweden, Haiti, Burundi, Benin	91	88
23	“, Austria, Azerbaijan, Honduras, Switzerland, Bulgaria, Tajikistan, Israel, El Salvador, (Hong Kong SAR) China, Paraguay, Laos, Sierra Leone, Jordan, Libya, Papua New Guinea, Togo, Nicaragua, Eritrea, Denmark, Slovakia, Kyrgyzstan, Finland, Turkmenistan, Norway, Georgia, United Arab Emirates, Singapore, Bosnia and Herzegovina, Croatia, Central African Republic, Moldova, Costa Rica	123	159

Thus, with each new scale, the number of countries introduced increases by a factor of a little less than 2, on the average. This approximate doubling of countries with each new scale begins to falter at about the population 2^{23}

(i.e. at around 4 million), for the simple reason that one has begun to run out of countries. (Note that the median-population country in this set, Singapore, has a population with 23 binary digits.)

These laws are not merely interesting statistical curiosities; for instance, Benford's law is often used to help detect fraudulent statistics (such as those arising from accounting fraud), as many such statistics are invented by choosing digits at random, and will therefore deviate significantly from Benford's law. (This is nicely discussed in Robert Matthews' New Scientist article "The power of one"; this article can also be found on the web at a number of other places.) In a somewhat analogous spirit, Zipf's law and the Pareto distribution can be used to mathematically test various models of real-world systems (e.g. formation of astronomical objects, accumulation of wealth, population growth of countries, etc.), without necessarily having to fit all the parameters of that model with the actual data.

Being empirically observed phenomena rather than abstract mathematical facts, Benford's law, Zipf's law, and the Pareto distribution cannot be "proved" the same way a mathematical theorem can be proved. However, one can still *support* these laws mathematically in a number of ways, for instance showing how these laws are compatible with each other, and with other plausible hypotheses on the source of the data. In this post I would like to describe a number of ways (both technical and non-technical) in which one can do this; these arguments do not fully explain these laws (in particular, the empirical fact that the exponent α in Zipf's law or the Pareto distribution is often close to 1 is still quite a mysterious phenomenon), and do not always have the same universal range of applicability as these laws seem to have, but I hope that they do demonstrate that these laws are not completely arbitrary, and ought to have a satisfactory basis of mathematical support.

— 1. Scale invariance —

One consistency check that is enjoyed by all of these laws is that of *scale invariance* – they are invariant under rescalings of the data (for instance, by changing the units).

For example, suppose for sake of argument that the country populations X of the world in 2007 obey Benford's law, thus for instance about 30.1% of the countries have population with first digit 1, 17.6% have population with first digit 2, and so forth. Now, imagine that several decades in the future, say in 2067, all of the countries in the world double their population, from X to a new population $\tilde X := 2X$. (This makes the somewhat implausible assumption that growth rates are uniform across all countries; I will talk about what happens when one omits this hypothesis later.) To further simplify the experiment, suppose that no countries are created or dissolved during this time period. What happens to Benford's law when passing from X to $\tilde X$?

The key observation here, of course, is that the first digit of X is linked to the first digit of $\tilde X = 2X$. If, for instance, the first digit of X is 1, then the first digit of $\tilde X$ is either 2 or 3; conversely, if the first digit of $\tilde X$ is 2 or 3, then the first digit of X is 1. As a consequence, the proportion of X's with first digit 1 is equal to the proportion of $\tilde X$'s with first digit 2, plus the proportion of $\tilde X$'s with first digit 3. This is consistent with Benford's law holding for both X and $\tilde X$, since

$$\log_{10}\frac{2}{1} = \log_{10}\frac{3}{2} + \log_{10}\frac{4}{3} (= \log_{10}\frac{4}{2})$$

(or numerically, $30.1\% = 17.6\% + 12.5\%$ after rounding). Indeed one can check the other digit ranges also and that conclude that Benford's law for X is compatible with Benford's law for $\tilde X$; to pick a contrasting example, a uniformly distributed model in which each digit from 1 to 9 is the first digit of X occurs with probability $1/9$ totally fails to be preserved under doubling.

One can be even more precise. Observe (through telescoping series) that Benford's law implies that

$$\mathbb{P}(\alpha 10^n \leq X < \beta 10^n \hbox{ for some integer } n) = \log_{10} \frac{\beta}{\alpha} \qquad (1)$$

for all integers $1 \leq \alpha \leq \beta < 10$, where the left-hand side denotes the proportion of data for which X lies between $\alpha 10^n$ and $\beta 10^n$ for some integer n. Suppose now that we generalise Benford's law to the *continuous Benford's law*, which asserts that (1) is true for all *real* numbers $1 \leq \alpha \leq \beta < 10$. Then it is not hard to show that a statistic X obeys the continuous Benford's law if and only if its dilate $\tilde X = 2X$ does, and similarly with 2 replaced by any other constant growth factor. (This is easiest seen by observing that (1) is equivalent to asserting that the fractional part of $\log_{10} X$ is uniformly distributed.) In fact, the continuous Benford law is the *only* distribution for the quantities on the left-hand side of (1) with this scale-invariance property; this fact is a special case of the general fact that Haar measures are unique (see e.g. these lecture notes).

It is also easy to see that Zipf's law and the Pareto distribution also enjoy this sort of scale-invariance property, as long as one generalises the Pareto distribution

$$\mathbb{P}(X \geq 10^m) = c 10^{-m/\alpha} \qquad (2)$$

from integer m to real m, just as with Benford's law. Once one does that, one can phrase the Pareto distribution law independently of any base as

$$\mathbb{P}(X \geq x) = c x^{-1/\alpha} \qquad (3)$$

for any x much larger than the median value of X, at which point the scale-invariance is easily seen.

One may object that the above thought-experiment was too idealised, because it assumed uniform growth rates for all the statistics at once. What happens if there are non-uniform growth rates? To keep the computations simple, let us consider the following toy model, where we take the same 2007 population statistics X as before, and assume that half of the countries (the "high-growth" countries) will experience a population doubling by 2067, while the other half (the "zero-growth" countries) will keep their population constant, thus the 2067 population statistic $\tilde X$ is equal to $2X$ half the time and X half the time. (We will assume that our sample sizes are large enough that the law of large numbers kicks in, and we will therefore ignore issues such as what happens to this "half the time" if the number of samples is odd.) Furthermore, we make the plausible but crucial assumption that the event that a country is a high-growth or a zero-growth country is *independent* of the first digit of the 2007 population; thus, for instance, a country whose population begins with 3 is assumed to be just as likely to be high-growth as one whose population begins with 7.

Now let's have a look again at the proportion of countries whose 2067 population $\tilde X$ begins with either 2 or 3. There are exactly two ways in which a country can fall into this category: either it is a zero-growth country whose 2007 population X also began with either 2 or 3, or it was a high-growth country whose population in 2007 began with 1. Since all countries have a probability $1/2$ of being high-growth regardless of the first digit of their population, we conclude the identity

$$\mathbb{P}(\tilde X \hbox{ has first digit } 2, 3) = \frac{1}{2} \mathbb{P}(X \hbox{ has first digit } 2,3) \qquad (4)$$

$$+ \frac{1}{2} \mathbb{P}(X \hbox{ has first digit } 1)$$

which is once again compatible with Benford's law for $\tilde X$ since

$$\log_{10} \frac{4}{2} = \frac{1}{2} \log_{10} \frac{4}{2} + \frac{1}{2} \log \frac{2}{1}.$$

More generally, it is not hard to show that if X obeys the continuous Benford's law (1), and one multiplies X by some positive multiplier Y which is independent of the first digit of X (and, *a fortiori*, is independent of the fractional part of $\log_{10} X$), one obtains another quantity $\tilde{X} = XY$ which also obeys the continuous Benford's law. (Indeed, we have already seen this to be the case when Y is a deterministic constant, and the case when Y is random then follows simply by conditioning Y to be fixed.)

In particular, we see an absorptive property of Benford's law: if X obeys Benford's law, and Y is any positive statistic independent of X, then the product $\tilde{X} = XY$ also obeys Benford's law – *even if* Y *did not obey this law*. Thus, if a statistic is the product of many independent factors, then it only requires a single factor to obey Benford's law in order for the whole product to obey the law also. For instance, the population of a country is the product of its area and its population density. Assuming that the population density of a country is independent of the size of that country (which is not a completely reasonable assumption, but let us take it for the sake of argument), then we see that Benford's law for the population would follow if just one of the area or population density obeyed this law. It is also clear that Benford's law is the only distribution with this absorptive property (if there was another law with this property, what would happen if one multiplied a statistic with that law with an independent statistic with Benford's law?). Thus we begin to get a glimpse as to why Benford's law is universal for quantities which are the product of many separate factors, in a manner that no other law could be.

As an example: for any given number N, the uniform distribution from 1 to N does not obey Benford's law; for instance, if one picks a random number from 1 to $999,999$ then each digit from 1 to 9 appears as the first digit with an equal probability of $1/9$ each. However, if N is not fixed, but instead obeys Benford's law, then a random number selected from 1 to N also obeys Benford's law (ignoring for now the distinction between continuous and discrete distributions), as it can be viewed as the product of N with an independent random number selected from between 0 and 1.

Actually, one can say something even stronger than the absorption property. Suppose that the continuous Benford's law (1) for a statistic X did not hold exactly, but instead held with some accuracy $\varepsilon > 0$, thus

$$\log_{10} \frac{\beta}{\alpha} - \varepsilon \leq \mathbb{P}(\alpha 10^n \leq X < \beta 10^n \text{ for some integer } n) \leq \log_{10} \frac{\beta}{\alpha} + \varepsilon \qquad (5)$$

for all $1 \leq \alpha \leq \beta < 10$. Then it is not hard to see that any dilated statistic, such as $\tilde{X} = 2X$, or more generally $\tilde{X} = XY$ for any fixed deterministic Y, also obeys (5) with exactly the same accuracy ε. But now suppose one uses a variable multiplier; for instance, suppose one uses the model discussed earlier in which $\tilde{X}$ is equal to $2X$ half the time and X half the time. Then the relationship between the distribution of the first digit of $\tilde{X}$ and the first digit of X is given by formulae such as (4). Now, in the right-hand side of (4), each of the two terms $\mathbb{P}(X \text{ has first digit } 2,3)$ and $\mathbb{P}(X \text{ has first digit } 1)$ differs from the Benford's law predictions of $\log_{10} \frac{4}{2}$ and $\log_{10} \frac{2}{1}$ respectively by at most ε. Since the left-hand side of (4) is the average of these two terms, it also differs from the Benford law prediction by at most ε. But the averaging opens up an opportunity for cancelling; for instance, an overestimate of $+\varepsilon$ for $\mathbb{P}(X \text{ has first digit } 2,3)$ could cancel an underestimate of $-\varepsilon$ for $\mathbb{P}(X \text{ has first digit } 1)$ to produce a spot-on prediction for $\tilde{X}$. Thus we see that variable multipliers (or variable growth rates) not only preserve Benford's law, but in fact *stabilise* it by averaging out the errors. In fact, if one started with a distribution which did not initially obey Benford's law, and then started applying some variable (and independent) growth rates to the various samples in the distribution, then under reasonable assumptions one can show that the resulting distribution will converge to Benford's law over time. This helps explain the universality of Benford's law for statistics such as populations, for which the independent variable growth law is not so unreasonable (at least, until the population hits some maximum capacity threshold).

Note that the independence property is crucial; if for instance population growth always slowed down for some inexplicable reason to a crawl whenever the first digit of the population was 6, then there would be a noticeable deviation from Benford's law, particularly in digits 6 and 7, due to this growth bottleneck. But this is not a particularly plausible scenario (being somewhat analogous to [Maxwell's demon](http://en.wikipedia.org/wiki/Maxwell%27s_demon) in thermodynamics).

The above analysis can also be carried over to some extent to the Pareto distribution and Zipf's law; if a statistic X obeys these laws approximately, then after multiplying by an independent variable Y, the product $\tilde X = XY$ will obey the same laws with equal or higher accuracy, so long as Y is small compared to the number of scales that X typically ranges over. (One needs a restriction such as this because the Pareto distribution and Zipf's law must break down below the median.) These laws are also stable under other multiplicative processes, for instance if some fraction of the samples in X spontaneously split into two smaller pieces, or conversely if two samples in X spontaneously merge into one; as before, the key is that the occurrence of these events should be independent of the actual size of the objects being split. If one considers a generalisation of the Pareto or Zipf law in which the exponent α is not fixed, but varies with n or k, then the effect of these sorts of multiplicative changes is to blur and average together the various values of α, thus "flattening" the α curve over time and making the distribution approach Zipf's law and/or the Pareto distribution. This helps explain why α eventually becomes constant; however, I do not have a good explanation as to why α is often close to 1.

— 2. Compatibility between laws —

Another mathematical line of support for Benford's law, Zipf's law, and the Pareto distribution are that the laws are highly compatible with each other. For instance, Zipf's law and the Pareto distribution are formally equivalent: if there are N samples of X, then applying [(3)](#pareto) with x equal to the n^{th} largest value X_n of X gives

$$\frac{n}{N} = \mathbb{P}(X \geq X_n) = cX_n^{-1/\alpha}$$

which implies Zipf's law $X_n = Cn^{-\alpha}$ with $C := (Nc)^\alpha$. Conversely one can deduce the Pareto distribution from Zipf's law. These deductions are only formal in nature, because the Pareto distribution can only hold exactly for continuous distributions, whereas Zipf's law only makes sense for discrete distributions, but one can generate more rigorous variants of these deductions without much difficulty.

In some literature, Zipf's law is applied primarily near the extreme edge of the distribution (e.g. the top 0.1% of the sample space), whereas the Pareto distribution in regions closer to the bulk (e.g. between the top 0.1% and and top 50%). But this is mostly a difference of degree rather than of kind, though in some cases (such as with the example of the 2007 country populations data set) the exponent α for the Pareto distribtion in the bulk can differ slightly from the exponent for Zipf's law at the extreme edge.

The relationship between Zipf's law or the Pareto distribution and Benford's law is more subtle. For instance Benford's law predicts that the proportion of X with initial digit 1 should equal the proportion of X with initial digit 2 or 3. But if one formally uses the Pareto distribution [(3)](#pareto) to compare those X between 10^m and 2×10^m, and those X between 2×10^m and 4×10^m, it seems that the former is larger by a factor of $2^{1/\alpha}$, which upon summing by m appears inconsistent with Benford's law (unless α is extremely large). A similar inconsistency is revealed if one uses Zipf's law instead.

However, the fallacy here is that the Pareto distribution (or Zipf's law) does not apply on the entire range of X, but only on the upper tail region when X is significantly higher than the median; it is a law for the *outliers* of X only. In contrast, Benford's law concerns the behaviour of *typical* values of X; the behaviour of the top 0.1% is of negligible significance to Benford's law, though it is of prime importance for Zipf's law and the Pareto distribution. Thus the two laws describe different components of the distribution and thus complement each other. Roughly speaking, Benford's law asserts that the bulk distribution of $\log_{10} X$ is locally uniform at unit scales, while the Pareto distribution (or Zipf's law) asserts that the tail distribution of $\log_{10} X$ decays exponentially. Note that Benford's law only describes the fine-scale behaviour of the bulk distribution; the coarse-scale distribution can be a variety of distributions (e.g. log-gaussian).