FREQUENCIES
The Journal of Size Law Applications
Aug 2001
Vol 1 no 1
Published by
Ekaros Analytical Inc.

War, Politics & Customer Loyalty: Forecasting Using Benford's Law

Q: What do these have in common?

- Months a renter stays at one address
- Years a purchaser stays with a supplier
- Minutes spent by visitors to a website
- Time for inventory to cycle
- The term of a political party in power
- The length of a war

A: They all follow Benford's Law!

So do hundreds of other time variables, from the time required to sell a house, to the time required to complete an engineering project. And this simple fact could have tremendous impact on decision-makers, if they knew and understood it. The forecasting and analysis opportunities provided by Benford's Law are tremendous. If you're thinking of making an investment in customer relationship management software, or selling property, or running for political office—or almost any activity involving management of time—you already have a large incentive to investigate forecasting using Benford's Law.

For example, take a moment to consider all the addresses you have lived at, and how long you stayed at each one. You almost certainly had more short stays than long ones. The longest might have been more than 15 years, but the average is likely to be under five. If you included vacation stays and summer jobs, the average would be still lower.

This tendency for small items to outnumber large ones is astoundingly common in nature; think of sand, pebbles, and boulders on a beach.

The core of this technique is therefore no more than a common-sense observation we have all made at some point: "small" outnumbers "large". The same principle has turned up in the work of Benford, Pareto, Zipf, Mandelbrot, and others. But when we apply it to time measurements, surprising things happen.

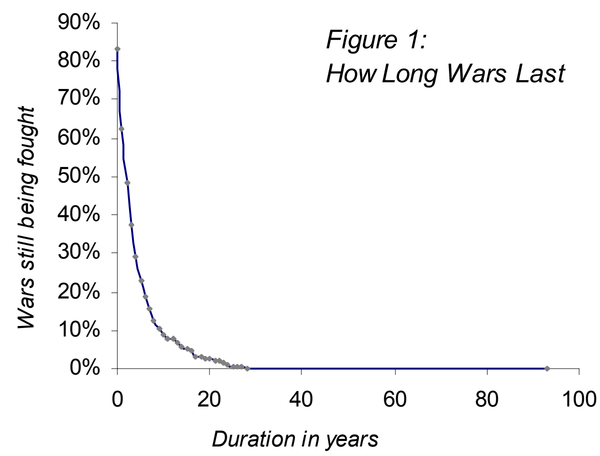


Figure 1: How Long Wars Last

This graph is based on 428 wars from 500 BC to the present. The curve represents wars still underway at each calendar year end. The highest point is 83 percent, because 17 percent of the wars began and ended in the same Year 0.

The critical point about the war-lengths graph is that while it declines very steeply at first, it soon levels off in the classic logarithmic "hockey-stick" shape. Most wars end after only 1 to 3 years; unfortunately, those wars that exceed 3 years tend to last much longer than the average. The lengths of wars are distributed according to Benford's Law.

A distribution, not just digits

In 1938, engineer and physicist Frank Benford published an article entitled "The Law of Anomalous Numbers". He showed that the initial digits of many measurements from nature followed a logarithmic distribution, such that the relative likelihood of occurrence P(dd) = log(1+1/dd) for any digit combination 'dd'. For example:

$$P(1)=\log(1+1/1)=0.301$$

meaning that a first digit '1' will occur about 30.1 percent of the time in a large sample of measurements. P(9)=log(1+1/9)=0.046, or 4.6 percent. Benford collected data such as baseball statistics, molecular weights, and the drainage areas of rivers, and found that as his collection grew in size and diversity, its conformity to the law become almost perfect. Even such odd items as the street addresses listed in *American Men of Science* tended to conform.

This digit law came as an immense surprise to most readers, and still does today. What we expect is for each digit to have a more or less equal chance of occurring; but this emphatically is not the case. Because we know what distribution to expect, it is possible to detect errors and fraud in data sets by measuring deviations from this law.

In almost every article on Benford's Law during the past 63 years, the focus has been on expected digit frequencies. Most discussions of the Law take for granted that Benford came up with a law *about digits*, and that this is the most remarkable aspect of his discovery. However, Benford could better be said to have discovered a law of distribution—what is sometimes called a size law.

Size laws have had a long and colorful history in science. Other investigators have made discoveries paralleling Benford's, including engineer and economist Vilfredo Pareto, who sometime in the 1890s coined an income distribution law of the form:

$$\text{Log } N = \log A + m \log x$$

where N is the number of income earners who receive incomes higher than x, and A and m are constants. In simplified terms, 80% of the wealth is owned by 20% of the population. This same proportion showed up in Pareto's studies of machinery breakdowns, where a small number of causes and individual machines were responsible for the vast majority of problems. The well-known '80-20 rule', so commonly cited in business and engineering, is the result of Pareto's research into size laws.

As another example of a size law, Harvard linguist George Kingsley Zipf formulated a ranking principle in the 1940s, pertaining to the frequency of occurrence for words in a large manuscript. He found that the relative frequency of a word could be approximated by its rank, such that the 2nd-ranked word appeared one-half as often as the most common one, and the 3rd-ranked word appeared 1/3rd as often, and so on. Zipf's work has gotten renewed attention in recent decades because he has been cited by Benoit Mandelbrot, the leading figure in fractal mathematics. (Fractal mathematics, and related developments in chaos theory, are yet another application of size laws.)

In fact, it was astronomer Simon Newcomb who actually published the earliest observations concerning a logarithmic order in nature and its consequences for digit frequencies. His work preceded Benford's by nearly 60 years, appearing in 1881. Unfortunately, Newcomb's letter on the subject went largely unread, whereas Benford's attracted the attention of mathematicians and physicists.

To summarize an immense body of work in just a few words: there is an underlying order in nature that makes Benford's Law work—and that same underlying order leads to the income distributions and machine breakdown patterns studied by Pareto, the linguistic patterns studied by Zipf, and many of the fractal patterns discovered by Mandelbrot. This consistent logarithmic ordering is found in money, populations, physical measurements, and frequencies of occurrence. It also applies to *time*. This is what allows us to create forecasts.

Frequencies editorial advisor Mark J. Nigrini published summaries of several corporate databases in his pioneering text *Digital Analysis Using Benford's Law*. These databases typically contain hundreds of thousands, or even tens of millions, of individual records. In general, they contain several hundred purchases of around $50 each, for every purchase in the neighborhood of $25,000—because $25,000 is hundreds of times larger than $50. This is size law behavior: an ordered relationship or distribution pattern in the magnitudes of things.

If you increase the size of something by N times, it becomes about 1/Nth as likely to show up. Thus as Pareto found, if you count the number of individuals who earn $10,000 per year, and then compare them with the number earning $100,000, those earning $10,000 are close to ten times as numerous.

Here is one essential but not very obvious corollary: for two successive earnings ranges that differ only by a common factor k, the total number of individuals will be roughly comparable. For example, if k=2, there will tend to be as many individuals (or slightly more) earning from $50,000 to $100,000 as there are earning $100,000 to $200,000.

Size laws have already proven hugely useful in many contexts. An insurance actuary trying to estimate the relative likelihood of a disaster involving claims for $100 to $150 million could work from known figures for past claims of $10 to $15 million, and extrapolate to arrive at a reasonable answer. The eventual human toll from an epidemic will also follow this pattern: if 1,000 are presently sick, the odds of the final number being in the range 1,000 to 9,999 are slightly greater than the odds of it being between 10,000 and 99,999.

In forecasting, history matters

In forecasting time values using the logarithmic law, we are attempting to exploit what we know about this distribution curve in general, to narrow down what the final number will be for some particular case.

As usual, the devil is in the details. To get from this principle to practical forecasting requires that we understand several nuances of the logarithmic curve as it applies in real life.

· First, the ideal curve is a limiting case, never completely reached. Although the ideal curve continues to infinity, real data invariably has an highest "x" value that is not exceeded . . . obviously, there has to be a largest value in any particular data set, whether we are dealing with incomes, accident rates, or how long a tenant stays in an apartment. Furthermore, the ideal curve extends to infinitely small values, whereas real data do not. Real curves diverge from the ideal logarithmic curve more and more as they approach these two extremes.

· Second, keep in mind that the change in probability is proportional to the *ratio* of successive "x" values, not the change in *absolute value* of "x", so the true logarithmic probability curve does not decline as quickly as in most other distributions used in forecasting. This is sometimes called "log-log" behavior, whereas the more commonly used geometric curve is "log-linear".

For example, if we are trying to evaluate an ongoing war at the end of its first calendar year, we know that this particular war has so far beaten odds of 17 percent by not ending the same year. We can take the original odds table and divide its entries by the overall chance remaining (100-17=83 percent) to estimate the new odds for each year still ahead. However, if a full year goes by and the war has still not ended, the odds change again; the war now must end in Year 2 or greater. Each year that goes by reduces the remaining possibilities and shifts the probability distribution.

Let's return to the graph shown previously, but applying a different scale for the y-axis and some hypotheti-

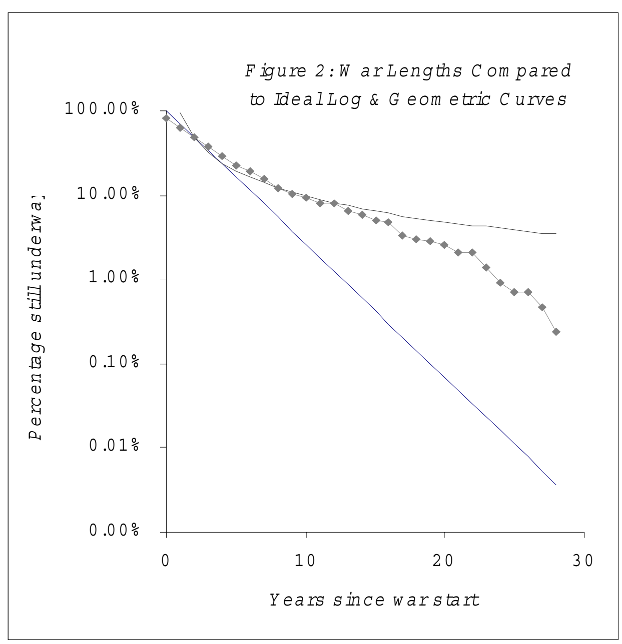


Figure 2: War Lengths Compared to Ideal Log & Geometric Curves

cal boundaries. We can see at a glance how the result falls inside the range. We can imagine many concrete reasons for this pattern.

The top line is the log-log curve, that is, the "ideal" Benford curve. The bottom line is the log-linear curve, in which the chance of the war ending is constant each year. You can see how using a logarithmic scale for the y-axis makes this log-linear or geometric survival curve into a straight diagonal line. The actual curve is clearly bracketed by the ideal log-log and log-linear curves, meaning that real wars fall short of perfectly ideal Benford behavior, but nevertheless that the chances of the war ending are not the same each year.

You'll find mathematical details in a separate article entitled "Getting Started with Benford Forecasting," available at www.ekaros.ca. To sum up: in real situations, a declining quantity over time will very often follow a curve that falls *below* the ideal and infinite log-normal curve, but well *above* a comparable time-independent geometric curve. There will always be a highest value, but it will not be strictly predictable from the total quantities involved. It will have to be observed in practice instead. As we approach that point, the declining probability of occurrence will follow the ratio rule . . . but very near the highest value, the ratio rule won't work as expected.

The wars are plotted by calendar year, so that if a war began and ended in the same calendar year (which we'll call Year 0), its duration is given as zero. Almost 17 percent of all wars in fact did this. Then 25.2 percent of the wars still underway on January 1st of the next calendar year ended in *that* year, Year 1. Notice the percentage is much higher than for Year 0 because each war now has a full year, not a part-year, in which to stop.

As of January 1st in the third calendar year, some 62.2 percent of all wars were still underway, and as a proportion of these, 22.5 percent ended on or before December 31st. This is a substantial drop from the 25.2 that ended in Year 1. This trend continues, year by year. By Year 10, the proportion of wars stopping per year is down to 13.3 percent. By Year 20, it has fallen to 8.3 percent. There are some slight variations due to the fairly small sample size, but the trend is unmistakable.

Here is a tragic lesson of history. The longer a war has continued, the longer it *will* continue. The hope for peace grows measurably smaller with each season that fighting persists. The pattern is not very surprising: the more cynical among us may say they knew it all along. But put it in mathematical terms, and suddenly it seems like a completely new fact.

Here is a tragic lesson of history. The longer a war has gone on, the longer it will go on.

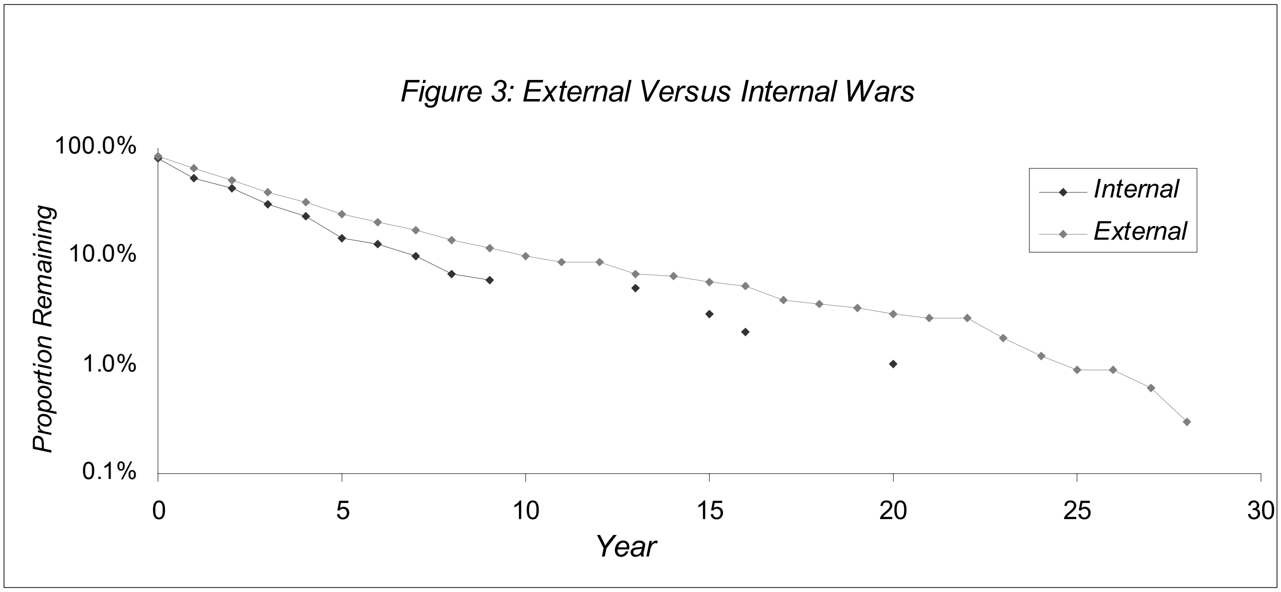
Do these numbers actually conform to Benford's original digit-frequency law? Yes, quite closely. For simplicity when dealing with sketchy historical records, I have rounded this entire set of dates to the nearest year. Such rounding does interfere with the way the Law works. But if we were to examine the exact date of each war's beginning and end, and so break up the rounding, we would find it doesn't matter if we measure time in weeks, months, or years. The digit frequencies conform equally well. Also, just as Benford found that conformity to his formula improved as he included more and more diverse "outlaw" numbers like baseball statistics and house numbers, here the fit gets better when we include civil wars and revolutions, brushfire wars and world wars, ancient and modern.

Scale invariance = shorter wars?

If two different sets of wars follow substantially different survival curves, how are they adhering to one law?

This is partly because of the ironic effect of scale invariance.

One absolute requirement if Benford's digit frequency law is to operate, is that when we measure some value, it does not matter what units we use. Benford's Law applies equally well if we are counting in dollars or yen, feet or inches, hogsheads or fortnights. So when we multiply all the numbers in a distribution by some constant, to convert from one unit to another, there should be no change in the relative frequencies of initial digits '1', '2', '3', and so on.



This is an incredible, yet totally logical consequence. Frank Benford said that "outlaw data without known relationship" tend to follow a logarithmic distribution. The lengths of wars are a kind of "outlaw data" and they follow a logarithmic distribution. So when we say that the hope for peace grows smaller with each season that a war continues, that is actually a corollary, a direct implication, of saying that the digit frequencies for the lengths of wars obey Benford's Law. All we needed to do was to look at the distribution curve itself, rather than the digit frequencies, to realize that one implies the other.

Next we will consider one of the dozens of questions the war-lengths graph brings to mind. Do different categories of war follow different distribution curves? For example, it might be that civil wars and revolutions last longer, or else end more quickly, than wars between sovereign nation-states.

As the graph shows, internal wars are typically shorter than international conflicts. The decline from year to year is consistently steeper. For this data set, the means are 4.56 and 3.15 years respectively, meaning internal wars average one third shorter.

continued from p. 4

As mathematician Roger Pinkham observed in 1961, the logarithmic Benford distribution has the unique property of being scale-invariant and base-invariant. No other distribution meets this requirement. Pinkham's argument greatly strengthened the argument that Benford's Law was more than a minor curiosity, and reflected a deeper order in nature. The initial digits of our measurements conform to a consistent law because in a sense, they must do so . . . or the universe would not work properly.

However, when we come to plot the actual values of the items in a distribution, this uncanny consistency of relative digit frequencies translates into a tremendous flexibility in absolute sizes.

Scale invariance implies scale indifference. Though past writers on Benford had no reason to mention it, being able to use different units of measurement also implies that we can *modify* our data, and actually *change* the outcomes we are measuring, yet still conform to the law!

If all the values on the war-lengths table were reduced by 25 percent, the resulting curve would still satisfy Benford's Law to a very good approximation, and it would still fall somewhere between a typical geometric distribution and the ideal log-log distribution. But if we were to have done this in reality, the world would have suffered from 25 percent less war!

The key in this case is the lower proportion of *long* wars. The small cumulative difference in probability during each of the first 10 years ensures that very few internal wars last into their second decade. The final four data points on the internal wars graph are deliberately shown as separate points, because they represent the four longest internal wars in the sample.

The difference between internal and external wars is so significant that it even shows up when the data are broken down month-by-month. Starting at January 1st of Year 1, 35.4 percent of the internal wars underway will end that year—which means that between 4 and 5 percent of the declining total will end each month. For external wars, the corresponding figure is barely above 2 percent.

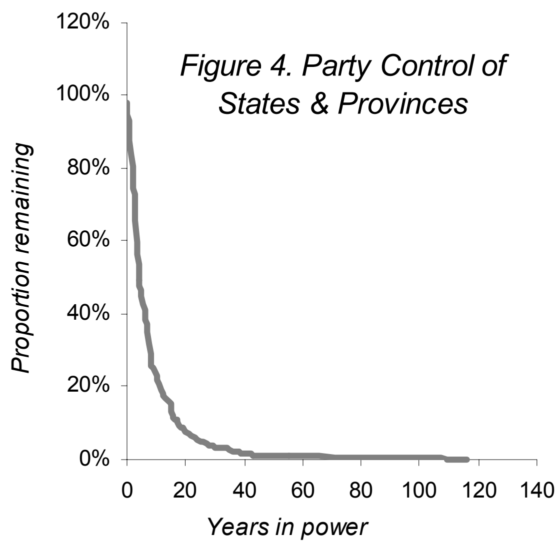


Figure 4. Party Control of States & Provinces

At present, I cannot propose any simple way to reduce the lengths of wars, or the casualties they cause—which also follow Benford's Law. We are clearly still in the stage of understanding and forecasting that phenomenon, not doing away with it. But in other areas, such as customer relationship management (CRM), or call center management, or even politics, the challenges are not so daunting and the payoff can be very significant.

First, we observe that the same pattern is evident in political party control of the fifty U.S. state governorships and the Canadian provincial and federal governments. Here I might warn any aspiring challenger: Beware the long-serving incumbent!

This graph is shown with linear axes that again emphasize the steepness of the initial drop. A logarithmic y-axis would produce a graph similar to Figure 2.

This pattern differs from the war data for several reasons. The most important is that while a war can potentially end on any given day, party control is normally only challenged once every two to four years. There are exceptions: occasionally an early election may be held due to a governor resigning or dying in office, or if a party holds a weak parliamentary majority and loses a vote of confidence. When all these possibilities are taken into account, the resulting graph from year to year is somewhat irregular, with many terms ending at 2, 4, or 6 years, and fewer at 1, 3, or 5. The slope of the line during the first decade is also not quite as steep as in the war data curve.

However, the pattern is still unmistakable. By the end of Year 4, only 49 percent of officeholders are still in

Thought Experiment

Our world would be very different if the trend was geometric or "log-linear" rather than "log-log," so that wars stood a steady 30 percent chance of ending each year. No more than one war in fifty thousand would last a full thirty years.

One study puts the total number of wars human beings have fought since 500 BC at around 140,000. In a log-linear world, events like the Ethiopian-Eritrean civil war (1961-91) would be so unlikely as to have occurred no more than two or three times in 2,500 years of recorded history. Only one war in *3,000 trillion* would last a century.

Consider the Hundred Years' War between England and France, which encompassed the famous battles of Crecy, Poitiers, and Agincourt. It already has considerable significance in our understanding of medieval Europe: for example, it shows the merits of longbows versus armored horsemen. Now it has a new significance for what it proves about the nature of war in general.

power. If the odds of ousting an incumbent remained at that level, then unbroken Democratic control of just one U.S. state for more than a century would have required the Democrats to overcome odds of about 30 million to 1. Yet in fact between the mid-19th and late 20th century they held four states for that long, while the Republican Party similarly controlled Vermont.

The practical implications for political strategists are huge. The first and largest is the absolute necessity to mount a challenge in lost jurisdictions as soon as possible. To show how large the effect is, consider this table, derived from the party control graph, of the average remaining span of control for incumbents who hold on to a given year.

The reward for surviving from Year 2 to Year 4 is not just the two years of time already served, but another 3.9 years, on average, that the incumbent party can expect to see added to its term. Thus, a party that wins a new governorship in 2002 can expect to hold on to about 2012—but if that party then successfully defends its office in 2004, it can expect, on average, to stay in power until 2018!

Table 1. Party Control in States & Provinces Over Time

End of Year	Percentage Remaining	Average Time Remaining
0	97.8	8.5 y
2	77.1	10.6 y
4	49.2	14.5 y
6	39.7	16.6 y
8	27.0	20.7 y
10	22.8	22.7 y
12	17.8	25.9 y
14	15.9	27.3 y
16	11.4	31.9 y
18	9.5	34.8 y
20	7.6	38.6 y
22	6.6	41.2 y
24	5.2	45.6 y
26	4.8	47.3 y
28	4.0	51.2 y

We can easily quantify the changing stakes imposed on us by Benford's Law, and that makes it possible to alter our strategy. If our donors will not pay for two full campaigns, then (all else being equal) the race for a jurisdiction that has been out of our hands for one term must get priority over the jurisdiction that was lost fifteen years ago. But notice: if we won fifteen years ago, and have held on since then, our future stake is now proportionately much larger. It is perhaps better to spend our money to protect a strong chance at thirty additional years in power, than to pour most of it into contesting a doubtful ten.

Here, as with the war data, we can ask what various subsets of the data look like, and how they might differ. (See Figure 5 on page 8.)

The first most obvious distinction is between the Democrats and the Republicans. The two parties have historically performed equally well in the early years. However, the Democrats have shown a consistent ability to hold on well past their 40th year in power. The Republican record trails off drastically at that point. In only one case out of 320, were the Republicans able to keep power for more than 40 years. The average Republican term in power was 9 years, while the Democrats held on for 11.5—28 percent better.

In Canada, a similar performance gap exists between the long-lived Liberals and their historical rivals, the Conservatives (who some decades ago renamed themselves the Progressive Conservatives). There is one crucial difference, however: the gap is narrowing in the U.S., but widening in Canada. Many southern states were Democratic strongholds from the Civil War onward, but finally fell to Republican challenges in the 1980s and 1990s.

By contrast in Canada, the PCs were nearly annihilated in the early 1990s, losing official party status at the federal level and then narrowly regaining it. Presently the party holds just one province, and ranks fifth in Canada's fragmented Parliament.

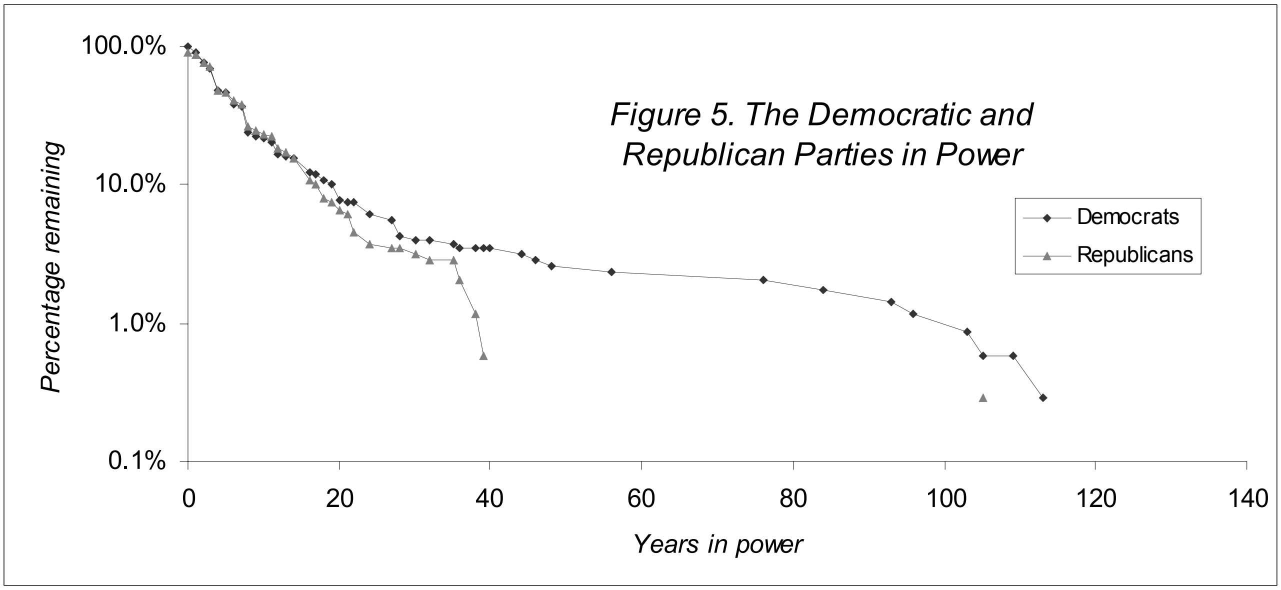


Figure 5. The Democratic and Republican Parties in Power

Data like this is very provocative. It leads to a hundred additional questions. Does the Democratic-Republican gap widen or narrow when the economy is booming or shrinking? Are wars over religion longer or shorter than wars for secular purposes? And so on.

The potential for business should be obvious as well. I suspect that the thoughts of many readers will already be racing ahead at this point. "What does my customer-retention curve look like? And what about compared with my competitors? How long are my engineering projects taking, and how long should they take? Which of my key business measures follow the law, and how quickly can I get the data together to find out?"

The mystery of customer disloyalty solved

Customer relationship management, or CRM, is a very hot topic. Estimates of ongoing investment in CRM systems worldwide are in the billions of dollars.

One frequently cited but poorly understood statistic says that increasing customer retention by just 5 percent can double net profit. The underlying argument, which is the essence of the CRM "revolution," is that a customer's total lifetime value to the firm is what should motivate business decisions, and that it is generally easier to find new ways to appeal to existing customers, than to find new customers.

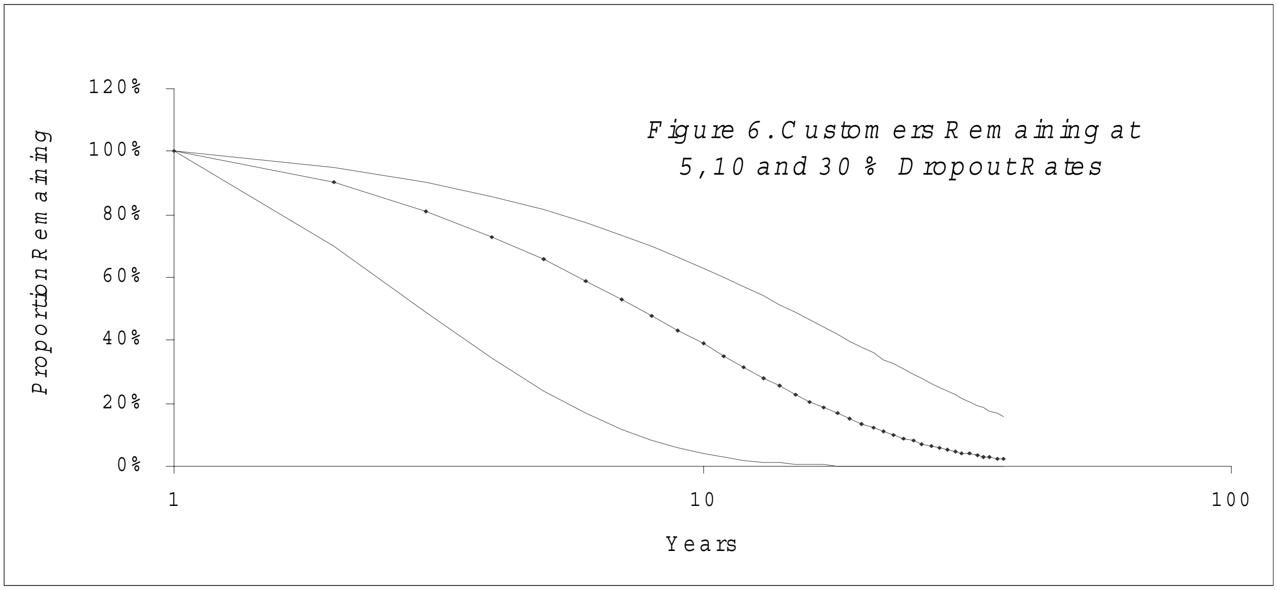
There is a strong presumption, not only among sellers of CRM systems, but also among observers, that once proper systems are in place, customer behavior will be dramatically different. It is taken for granted that today's methods of handling customers are deeply flawed. Thomas Stewart’s tone in *Intellectual Capital* (1994) even contains a hint of outrage:

> Only mismanagement of customer capital can explain why U.S. companies on average lose half their customers in five years, or why—despite obvious improvement in the quality of manufactured goods, negligible price increases, and unceasing rhetoric about treating the customer right—customer satisfaction is actually declining in the United States.

In fact, we can explain the strong tendency of customers to defect, and why defection rates have risen in recent years, without indicting America's corporations for gross mismanagement. To do so, however, requires careful attention to Benford's Law.

The "5 percent = doubled profits" statistic is based on a hypothetical case of the log-linear or geometric depreciation curve. If 10 percent of your customers consistently drop out each year, then the expected average customer lifetime is equal to (1.00-0.10)/0.10, or about 9 years. Reduce your dropout percentage to 5, and the average lifetime rises to about 19 years. Doubled lifetime revenue from each customer means doubled profit.

In one sense, this kind of calculation is elementary. Reducing the dropout percentage to 5 is the same in



this case as reducing it by half—so it's not surprising that when customer dropouts fall by half, profits are free to eventually double. However, in another sense, the result is highly misleading.

If your dropout rate is more like 35 percent (and many businesses would be delighted with 65 percent repeat sales from year to year), then your eventual revenue gain from reducing dropouts to 30 percent will be a lot smaller. It will be about 25 percent, not 100 percent. Meanwhile the number of additional customers induced to stay is exactly the same, and the work of hanging onto them is the same as well.

In Figure 6, three lines are shown, representing different customer dropout rates. The middle line reflects dropouts of 10 percent per year. The upper line represents dropouts of 5 percent, and the lower line represents a rate of 30 percent, which is much more commonly observed.

Dropout rates are like compound interest, in reverse. By year 5, a dropout rate of 30 percent will have eliminated three-quarters of your customers, whereas with a rate of only 10 percent, you will still retain two-thirds of them. By year 40 or so, the number of customers remaining on the 5 percent curve is nearly *ten times* the total for 10 percent.

A steep dropout curve robs a CRM investment of much of its value. To see this clearly, imagine you operate at a 35 percent dropout rate, and by investing in CRM measures, induce a hundred extra customers to stay—5 percent of a total pool of 2,000 customers. Next year, even if you maintain your more attractive offer, 30 out of that 100 will stop buying. On average, your "rescued" customers will only stay for two or three years of additional business. Practicing CRM in a high-turnover environment is therefore like building a break-water out of sand. The hostile business environment seems to wash away your gains as fast as you make them.

Another point often overlooked: even this modest return will take more than two years to start showing up, because it amounts to adding an additional year to the *far end* of the typical customer lifetime. For this example, the expected first-year gain amounts to 5/65ths of your present revenue, or about 7 to 8 percent, before any special costs for servicing these additional customers are considered. These numbers also assume that you can change your prevailing dropout rate in the first place—something that requires creative insights and skill in execution that many businesses just won't have.

Many businesses, seeing their customer dropouts hovering year after year at about 30 to 35 percent, are likely to conclude that lifetime-value approaches are nice in theory, but less worthwhile in practice. One may ask, "What good is it to spend millions on a CRM strategy to alter customer behavior, when virtually our entire customer base gets replaced every two or three years? Particularly when our competitors have exactly the same problem, and nobody has any ideas for solving it?"

What does Benford's Law say about all this? First, that all else being equal, the rate at which customers defect will decline dramatically with time. If customers in their first year drop out at 35 percent per year, by their tenth year, the defection rate may fall to 10 to 15 percent. In another decade it will drop by half again. *On average, there will be many more long-term survivors than the short-term loss rate seems to indicate.*

Yes, customer turnover has increased—not because the products are bad, but because new customers tend to leave, and new products make everyone a new customer.

However, the law also says that attempts to secure consistent long-term relationships from the start will generally not work. There is a period of "infant mortality" in relationships that is inevitable.

There are clear differences from company to company and industry to industry, as shown in a much-discussed study by Frederick Reichheld of Bain & Co. Reichheld found the following potential for profit improvement in different industries, assuming a 5 percent gain in retention:

Industry	*Profit Improvement*
Software	35 %
Credit cards	75 %
Advertising	95 %

These numbers depend less on fundamental differences between industries than they do on simple differences in turnover. The software industry typically re-invents itself every 18 months or less, and continually has to begin again with new products and new customers. The credit card business is comparatively stable, with customers often keeping a particular card from college through midlife or later—but the continual flow of new cardholders tends to bring averages down. Major advertising accounts are even more stable, as they tend to belong to large corporations that have been around for decades.

Companies with average turnover of 35 percent or higher per year tend to emphasize capturing new business as fast as possible. The perceived high dropout rate drives management in the direction of seeking maximum market share, rather than maximum long-term profit per customer. The first purchase the customer makes is often the only one, so it has to carry the entire financial load. A policy of seeking immediate increases in volume will remain attractive compared with subtler notions based on long-term returns from CRM.

This can reach absurd extremes. The dot-coms spent tens or even hundreds of millions to create a state-of-the-art Web presence, then threw in Super Bowl-style advertising, to arrive at costs for each new customer acquired that were double those of their bricks-and-mortar competitors.

These companies let a frenzied bull market and "New Economy" hype pull their attention away from fundamentals. But the pattern is not new; the harsh realities of poor customer retention have contributed even more to this trend. How can you hold fickle customers who are determined to go elsewhere? Perhaps by making yourself the only game in town!

It is widely thought to be better to spend a billion dollars and wind up as Number One in the market, able to increase volume at will, than to spend a hundred million, run second, and get shut into a perpetually unprofitable niche. And this may in fact be true—if the product is new, and if the customers are new.

Thus we see more clearly the irony of what Thomas Stewart complained about. Yes, measured in terms of retention, customer satisfaction has declined in this era of rising product quality and radically new products and services. Why? Not because of mismanagement—

but because the products are genuinely new, the relationships are new, and history matters!

Paradoxically, this can be the reality for some companies even though long-term relationships may make up 90 percent of all business at others. It entirely depends on where your company is in relation to its product life-cycle, and its customer history. Are you at the beginning of the life cycle with both—a startup with brand-new customers? Are you a mature business with long-term contracts? Or possibly you have a mature business that is breaking into a new market. Once you know the answers to these questions, you can establish a coherent CRM strategy and use Benford-based forecasting correctly.

The key measurement, in all cases, is customer retention. Advocates of CRM, seeking to avoid setting unrealistic expectations about turnover, and to sell their systems to the widest possible market, will often describe the purpose of CRM as something other than boosting retention—to understand the customer, to improve the buying experience, to plan better products, to reduce transaction costs, and so on. If seen in these terms, CRM can make sense even if it does not boost retention.

However, without retention, there is no relationship. To paraphrase a recent book title, customer satisfaction isn't actually worth that much. Customer *loyalty* is worth everything. The critical insight that Benford's Law provides, is that it offers us a reliable measure of how much customer loyalty is possible at a given moment in the relationship.

Calculating customer survival curves

Although many writers on CRM have hailed the customer-lifetime-value approach as revolutionary, the method is far from new. Savvy mail-order operators and subscription magazines used it as far back as the 1920s, and it was probably not new even then. Julian L. Simon first wrote in 1965 that estimating what a customer is ultimately worth is "the most important calculation a mail-order merchant makes," because it is the foundation of profitability. He found it "sad and amazing" that many companies operated in complete ignorance of their own repeat purchase rates. (Source: *How to Start & Operate a Mail-Order Business*, Fifth Edition, 1993.) Yet today, business managers who can correctly calculate a customer's lifetime value are still rare.

For those who investigate with sufficient attention to detail (as Simon urged), the pattern will become clear, whether the investigator knows Benford's Law or not. The fact is that in the long run, virtually every business has two broad classes of customer, and experiences high turnover in one, plus low turnover in the other, at the same time. Wherever Benford's Law holds, the distribution of dropouts will be high in the early years, and will become proportionately smaller with time. History matters!

This table of customer retention values was derived from several real sources. I am keeping the sources anonymous by producing a composite curve.

A small percentage change in retention among customers who have only been customers for a few years will not produce major gains. Turnover at the beginning of the history curve is always high. However, a similar percentage change among the long-term survivors will almost certainly pay big dividends.

Table 2. Customer Retention and Projected Time Remaining

End of Year	Percentage Remaining	Average Time Remaining
0	81.5	3.7 y
1	63.4	4.0 y
2	40.7	4.6 y
3	25.2	5.8 y
4	17.1	7.1 y
5	13.9	7.5 y
6	11.9	7.6 y
7	10.6	7.4 y
8	9.3	7.3 y
9	7.7	7.6 y
10	6.8	7.6 y
11	5.5	8.1 y
12	5.0	8.3 y

What is more, if a supplier focuses on lowering turnover for the long-term minority, then sometime in the company's second decade, that minority will be transformed into a majority. Customer relationship management aimed at securing greater loyalty will work for virtually any business, provided it recognizes the inevitability of "infant mortality" among new customers and focuses on identifying long-term survivors.

Consider this example. You must decide whether to send your best salespeople to pitch to customers you have only had for two years, or those you have had for ten or more. What is at stake in each case, and what can you expect to happen?

A brand-new customer with no history will stay on into the next calendar year about 81.5 % of the time, and can be expected to stay about 3.7 more years on average. At the end of Year 0, each such customer therefore represents about 63.4/81.5 x 3.7 = 2.9 years of additional business.

The goal of CRM is to help you invest the right amount in each customer. Forecasting using Benford's Law provides a measure of how much loyalty is likely at a given point in the relationship.

For a customer who has lasted through Year 1, the odds of getting to the end of Year 2 fall to 40.7/63.4 = 64 %. The expected remaining life is still 2.6 years. The two-year customer is barely distinguishable from a brand-new sale in terms of total future value-but don't forget that for a complete comparison, you must add back the additional year of business you have already had. The total of 3.6 years compares favorably with the 2.9 years you expected a year ago.

In the case of a customer who reached the end of Year 4, you would now be looking at 5.8 years of projected additional business, double your stake at Year 2. Add back the two years that have elapsed, and the new total of 7.8 compares very well with 3.6.

By Year 8, the projected payoff has risen to 6.1 years. Your estimated total reward for holding onto a customer during Years 4-8 is therefore 4+6.1 = 10.1 years of sales!

In this combined data set, the drop-off was quite steep, and no relationship lasted longer than 35 years. As a result, the average remaining customer lifetime at Year 8 came out to be only 7 to 8 years. If we had calculated the results using a data set that extended back fifty years or more, like the political longevity curve, the payoff would likely have doubled again between Year 4 and Year 8, to about 12 years.

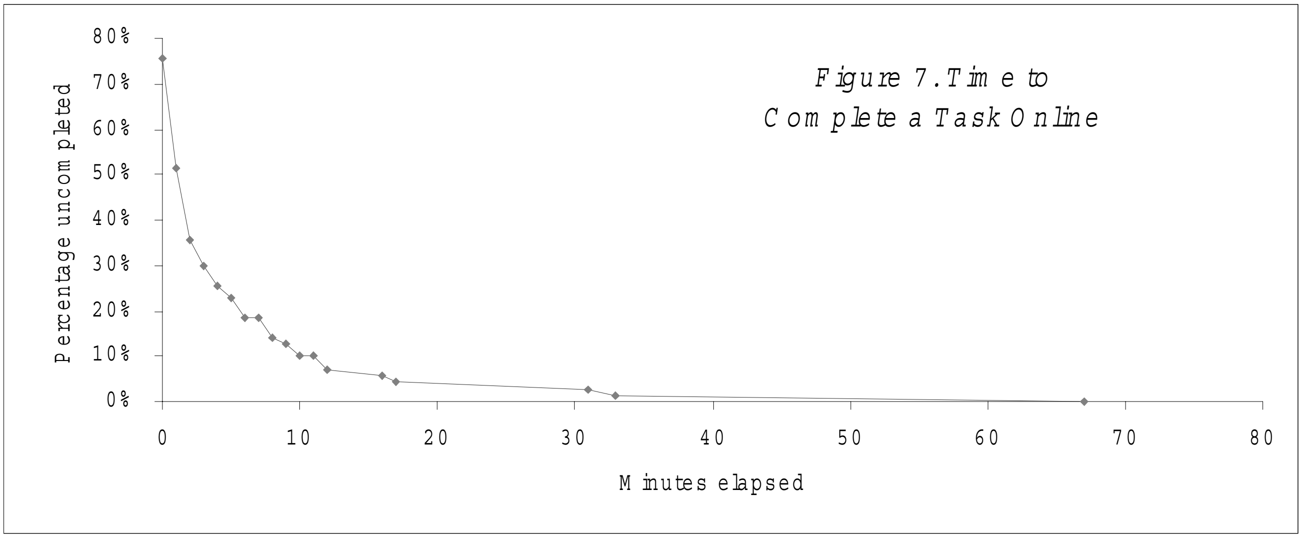
This same pattern should hold for any industry in which lasting customer relationships are key. Up to a point, each doubling in the length of the relationship leads to a similar doubling in the payoff still to come.

Here we must deal with a separate technical puzzle relating to the odds of success. Suppose we concede that Benford's Law does operate in customer-supplier relationships. So what? Is it in fact better to focus your best sales talent and your largest investment on long-term customers, or short-term ones? Should you offer bigger discounts to win new business, or hold on to existing business?

It may seem as though long-term customers are "in the bank," given that the odds of success in getting the next order can approach 90 percent. In a very competitive industry, the odds of your very first sale to a given prospect might be only 5 percent or 10 percent, and first-year retention might be lower than 50 percent. So to avoid letting rivals gain market share, a company may feel the need to put its best team into pursuit of new business, to discount most heavily when going head-to-head with competitors, and to treat long-term business as a "cash cow," a place to break in newly hired junior sales staff.

This amounts to treating the Benford curve as a kind of cosmic reward for relationship longevity—after a certain point, the reasoning goes, you don't have to try as hard. You can focus on the next battle.

This strategy is potentially flawed on several counts. First, it puts the highest cost on the market segment least able to bear it. The cost of a sales call is the same whether the odds of success are 5 percent or 95 percent. And as the dot-com frenzy proved yet again, added



advertising or discounts that double your cost of sales won't necessarily double your odds of success.

Second, what happens when your competitors decide to follow the same logic, and commit their best sales team and lowest discounts to win over your "cash cows"? Granted, your competitor may not hang on to them, and most likely won't obtain the same volume of business you would have-but that's not very comforting when twenty years of future orders vanish.

The point of this analysis and of CRM in general is to invest the right amount into winning business—to invest no more and no less than what that customer is worth. Taking a long-term customer for granted is about as reasonable as taking your spouse or partner of twenty years for granted.

What else can Benford forecasting do? A single article doesn't leave room to list all the ways in which this method can be used. In fact, it doesn't even come close to covering the territory, guaranteeing that *Frequencies* will return to this topic. But before I close this essay, I have chosen some provocative possibilities to focus on from call centers, healthcare, and real estate. In each case, whether the professionals in question are doctors or realtors or technical support staff, the pattern is already widely recognized in an informal, anecdotal sense. The critical insight needed here, is that in each case we are dealing with one very specific and consistent law, on which quantitative forecasts can safely be based.

Call center & website management

This example is drawn from a website where users select from a variety of downloadable educational materials. On one particular web page, two articles are offered. One is a brief introduction, the other is a much denser and longer treatment of the same topic.

Some users will elect to read only the introduction. A smaller group moves directly to the long version. The majority of users download the easy one, read it, then download the other. The time in minutes between the first download and the second follows Benford's Law.

This same pattern holds for the amount of time spent on websites in general. It also holds for call centers, where service personnel deal with customers "live"—again, only provided that the problems to be solved are sufficiently varied.

This raises some interesting possibilities for creating Benford forecasts in real time. Most call centers pay close attention to wait times—how long a customer must stay on hold before speaking to a live operator. Computing correct wait times, minute by minute, and then communicating them to waiting callers, is a very important task.

It should be very easy to modify the call-routing software to track how long each individual call has been "live" with a service person, and use this information to generate a more accurate forecast of the call's final length. The same information may also be helpful in managing staffing levels during the day.

Possibly the most important lesson for call center operators is a negative one: time limits *will not work*. Consider this story from *Fortune* magazine, about PC manufacturer Gateway Computers:

> . . . one policy put a time limit on customer-service calls; reps who spent more than 13 minutes talking to a customer didn't get their monthly bonuses. As a result, workers began doing just about anything to get customers off the phone: pretending the line wasn't working, hanging up, or often—at great expense—sending them new parts or computers. Not surprisingly, Gateway's customer satisfaction rates, once the best in the industry, fell below average.

Gateway quickly reversed its policy on time limits, as well as other initiatives relating to commissions and pricing, and at the time of writing (April 30, 2001), customer satisfaction and sales had rebounded.

The critical point is that the shape of the distribution curve is not a matter of poor operator training, problem customers, or any one variable. To change the average call time or average length of visit on a website is certainly possible—but in the absence of rules arbitrarily ending every call or visit at a set limit, the shape of the curve will remain.

Real estate sales and rental tenancy

The length of time that properties remain on the market, or that tenants remain in a particular location, follows the log-log Benford pattern. For any organization managing significant numbers of properties—even a seller with one property—estimating the week-to-week likelihood of a vacancy or sale can be well worthwhile. Interest costs, maintenance and cleaning, and many other costs, can be estimated with greater confidence. So can the yield of rental income.

For example, take two apartment buildings with 40 tenants apiece. All else being equal, if the average tenant in Building A has been there five years, and the average tenant in building B has been there only two, Building B can expect to clean and re-rent perhaps one suite per month, while Building A will deal with closer to one-third that number of new vacancies.

New buildings will often contend with abnormally high turnover rates for their first five to ten years, even if the buildings are similar in every way to their neighbors. The problem is not the rents, the location, the quality of building management or the clientele, but something more subtle: the fact that people who have just moved are generally the most likely to move again soon. This is a major expense that owners may not anticipate or account for properly in their planning.

In fact, mathematically knowledgeable managers can easily make the mistake of assuming the opposite. By thinking in terms of the more well-known Gaussian or bell curve, one would assume that people who have just moved in are less likely to leave, while those who have stayed several years are "about due" for a move. This might perhaps be true if the tenants were college students or transferred military personnel. However, the broader the sample, the more Benford-like the resulting distribution of tenancies.

Patient-stay management for health care

A large general hospital or healthcare system typically treats a wide enough variety of illnesses for Benford's Law to apply to the lengths of patient stays. Occupancy forecasts based on the particular complaints of the patients, plus doctors' reports, prescription records, and so on, are not only complex and uncertain, but labor-intensive and therefore challenging to turn out on a day-by-day basis. By contrast, any hospital should be able to report in minutes on who is occupying what bed, and how many days they have been there—which is virtually all that is needed to create a precise Benford forecast.

The same method can be applied to sick-leave numbers by the Human Resources departments of large organizations. Separate distributions can be computed for different wards or divisions or patient categories, wherever the differences are consistent and measurable. The savings in terms of better day-to-day planning are likely to be significant, for example in making the decision when to hire temporary replacements. Most of us (at least here in Canada) know someone who was scheduled for surgery or other overnight treatment, but then faced delays when patients expected to be discharged were not.

One attractive aspect is that a Benford forecast does not require confidential and personal information to be reviewed by yet another layer of analysts or clerks. Decisions can be made with a high degree of impartiality.

Some additional applications for Benford-based forecasting

The average survival time of a new business startup, in months or years. The often-quoted statistic that half of all businesses fail in five years, simply restates Benford's Law.

Time to complete a painting. While most Renaissance and modern-style oil paintings are finished in a few weeks or months, a substantial number take years. Cases exist of paintings that stood half-finished in an artist's studio for 30 to 50 years, but were completed and sold.

Weeks or months that a delinquent account will remain unpaid (very helpful from the point of view of computing and adjusting a bad-debts reserve).

Days, weeks, or months for an engineering design project to be completed (depending on the kind of work). Months or years for a major construction project to be completed.

Weeks for a teenage runaway to remain unaccounted for. Of 47,000 teenagers who go missing in Canada each year, 95 percent return home or are located within 12 months.

Time to failure for some structures and some kinds of equipment. This topic deserves an article in itself, as electronic component failures often follow a failure curve determined by their physics, not the Benford curve. On the other hand, consider Roman bridges still standing today.

Hours or days for a wilderness search effort to find a missing person. This kind of estimate can yield even more useful information if correlated with the number of searchers.

Days, weeks or months to turn a prospect into a sale (depending on the product).

The law also applies to real estate sales. Multiple listing services already pay close attention to average turnover times. When the average rises or falls significantly across a region, realtors are quick to draw conclusions about their own prospects. However, such services normally do not display how long a property has been for sale, precisely because the information is so meaningful to potential buyers.

Ironically, while Benford-based forecasting would neatly sidestep several ethical and privacy issues for hospitals, it might actually cause some added conflict between real estate sellers (who would prefer not to have time-on-market discussed), and buyers (who would benefit from knowing which properties have had the most difficulty finding a buyer). In the short term, private use of Benford by sellers and their agents is more likely.

Conclusion

In almost any practical field, we will find a rough, anecdotal understanding already in place among professionals, that the longer certain kinds of process or relationship have lasted, the longer they will likely continue. This may be termed "the good old 80-20 rule," or may be described in more specific and quantitative terms. For example, the movie industry already understands quite well that most movies close after a few weeks, and very few run for six months in one location. Any receivables manager understands that long delays in payment are generally indicative of longer delays to come.

This correlation can represent good news, as with supplier-customer relationships. In other cases it is taken as bad news, as in the lengths of wars. (I will leave it to the reader to decide if the incumbent's advantage in politics is good or bad.) Either way, to systematically exploit these various patterns requires a simple but rigorous mathematical approach, and above all, an appreciation that the observed distribution of outcomes in this field or that field are all *reflections of a single underlying law, and not unique.*

Individuals from any single profession may feel reluctant to invest time and effort to investigate survival curves or forecasting techniques unique to their industry or field. Knowing that they are working from a well-grounded theory, and can compare their results with those in other fields, will help remove that reluctance.

As Julian Simon's writing on mail-order business shows, we have had some of the mathematical tools for decades, but they have not led to widespread understanding. Yet as today's gigantic investments in CRM show, the demand for better forecasting and assessment has never been greater. Benford's Law can offer us dramatic progress in filling that need. □

Size Law Pioneer: Simon Newcomb

An ordinary person can recognize the extraordinary. But it takes an extraordinary person to demonstrate important truths in things that are quite ordinary. Such a person was Simon Newcomb (1835-1909), the most famous American astronomer of his time. Among Newcomb's many accomplishments was his discovery of the initial-digits law now called Benford's Law. Newcomb's observations preceded Benford's by nearly 60 years; but it was Benford's paper that led to further critical inquiry and widespread acceptance of the law.

Newcomb's name is hardly a household word today, as it was in the nineteenth century. He was probably the greatest scientist or mathematician to come out of Canada in that era.

> ***"Many men stumble across the truth, but most manage to pick themselves up and continue as if nothing had happened."***
>
> ***—Winston Churchill***

Born in Canada of New England descent, Simon Newcomb spent his early years in Nova Scotia, Prince Edward Island and New Brunswick, where his father served as a rural schoolteacher. The father's notions of natural education ran counter to the accepted norms of his day, so he found it necessary to move often in search of new employment. Simon's family knew little security in these years. Simon displayed mathematical aptitude at a young age; having spent several hours a day on mathematics from age five, he had learned to extract cube roots by the age of seven. However, he gained little formal education.

Ironically, his early experiences with science were all discoloured by superstition. He studied phrenology, the aberrant theory that human personality is formed by the shape of the skull. He served as an apprentice to a quack doctor, who was later jailed for malpractice. Perhaps these experiences taught Simon to re-examine the confusion that commonly passes as knowledge, and not to accept as fact things that are inexact or groundless.

In Newcomb's autobiography, he describes his youth in a chapter entitled "The World of Cold and Darkness". His "World of Sweetness and Light" began at age 21 with his employment as a computer (computers were people in that age) at the Nautical Almanac Office in Cambridge, Massachusetts, and his concurrent immersion into mathematical astronomy. He soon overcame his lack of formal education, and from that day on, Newcomb lived, breathed, ate and slept numbers.

From Harvard University, he came into contact with great scientists and thinkers of his day. His astronomical travels took him on eclipse expeditions to northern Canada, and on data-collection visits to the astronomical centres of Europe. In time, he became director of the Nautical Almanac Office, which became part of the U.S. Naval Observatory in Washington. He became a professor of mathematics and astronomy at Johns Hopkins University. He wrote profusely on mathematics, economics, and other subjects, and carried on correspondence with the great minds of his day. His influence guided the construction of several great observatories. He measured the speed of light in space, and his monumental works on the motions of solar system bodies remained standard until the mid-twentieth century. The system of astronomical constants he devised retains important even today.

Perhaps his most enduring accomplishment was at an 1896 conference between the United States, Great Britain, France and Germany, to promote international cooperation among astronomers. In cooperation with A.M.W. Downing, superintendent of the British Nautical Almanac Office, he secured a general agreement on the constants that were to be used by all ephemerides—constants derived for the most part from his own work. At a similar conference held in 1950, it was the unanimous decision of the delegates that the system of 1896 was still superior to any other available, so that Newcomb's system continued in use until late in the 20th century.

Formal recognition of Newcomb's accomplishments came in many forms, too numerous to allow more than a brief mention. He was founder and first president of the American Astronomical Society, first president of the American Society for Psychical Research, president of the American Mathematical Society, president of the American Association for the Advancement of Science, president of the Philosophical Society of Washington, and vice-president of the National Academy of Sciences. He received the highest astronomical awards from scientific societies in the United States, Britain, the Netherlands and Germany. An award called the Simon Newcomb Award of the Royal Astronomical Society of Canada continues to honor him up to the present day.

We are left with very little information on how he discovered what is known today as Benford's Law. Newcomb's two-page "Note on the Frequency of Use of the Different Digits in Natural Numbers" appeared in the *American Journal of Mathematics* in 1881, while he was its editor. It seems to have generated no interest. As with his astronomical notions of "dark stars" (astronomers are still wrestling with that idea), and his monetary theories, he was just too far ahead of his time.

Newcomb is also credited with a second law later named for another man, Fisher's Law. This "quantity theory of money" later proved fundamental to the work of Milton Friedman and the monetarists.

In contrast with Benford's later and much lengthier treatment, the paper did not include any observational data such as numbers from almanacs or newspapers. Possibly Newcomb, so familiar with data tables himself, did not feel it was necessary to provide those details. Or possibly he edited his observations out for lack of space. His paper did, however, provide the same equation that later appeared in Benford's paper: $P(d) = \log(1+1/d)$.

It is perhaps not surprising that Newcomb, whose interests were extremely far-ranging, is also credited with originating a *second* widely used law attributed to another man, Fisher's Law. This is better known as the Quantity Theory of Money, which says that:

$$MV = PT$$

where M is the quantity of money in circulation, V is its velocity (or the rate at which transactions occur), P is the price level, and T is the volume of transactions. This theory, introduced by Newcomb in 1885, was revived by Irving Fisher in 1911. The theory has had enormous influence on economics. For example, the central thesis of Milton Friedman and Anna Schwartz in their *Monetary History of the United States* (1963) was to show that T/V was a stable quantity, and thus that an increase in the supply M would result in rising prices (P).

Newcomb combined the qualities of the abstract mathematician with a very practical turn of mind. His work tended to result in empirical approximations that have nonetheless stood up extremely well over time. He seemed to have a very astute sense of how much precision was needed to establish a law, and spared no effort to reach the needed result. As he once observed: "Ten places of Pi are sufficient to give the circumference of the Earth to a fraction of an inch, and thirty decimal places would give the circumference of the visible Universe to a quantity imperceptible to the most powerful microscope." Advances in telescopes and microscopes during the intervening century have tempered the literal accuracy of this statement somewhat, but the notion stands. How much do we need to know, to rely on our models as enduring laws? Few men were a better judge of "how much" than Simon Newcomb. □

Notable books by Simon Newcomb

ABC of Finance (1877)
Principles of Political Economy (1886)
Reminiscences of an Astronomer (1903), Newcomb's autobiography

Editorial

Scaling the 'Iceberg'

Thank you for picking up our first issue.

Our title, *Frequencies: The Journal of Size Law Applications*, neatly summarizes our primary subject matter. We plan to focus on the numerous practical applications and benefits of size laws, more specifically the logarithmic or "heavy-tailed" distribution. This will mean venturing into many parallel fields of study, based on the works of Pareto, Zipf, Benford, Mandelbrot, and others.

Why size laws? And why now?

In one sense, such a focus is long overdue. As the table at right shows, the first explorations of logarithmic distributions in nature by Pareto and Newcomb are now more than a century old.

We are overdue in a more recent sense as well. From two bibliographies by Mark Nigrini and Wentian Li, I counted 143 papers or books prior to 1990. Since 1990, there have been 137. Interest has grown with each new practical application discovered. The field of possibilities has become so vast that no researcher can expect to stay fully informed on them all.

It is no coincidence that the major figures in size law theory have invariably focused on more than one area of application, or otherwise stressed the generality of their findings. Mandelbrot began his career by writing about communications statistics, as an extension of Zipf's work, in 1951. He moved on to financial analysis in the 1960s, and only came to his famous discoveries in fractal geometry in the 1970s. Pareto examined failure rates and income distributions. Zipf studied languages, city sizes, and other topics.

In summing up their findings, researchers have tended to agree that the most interesting features of Zipf's Law, Pareto's Law, Benford's Law, and size laws in general are yet to be discovered. In effect, for the past century we seem to have been exploring the most accessible one-tenth of an immense iceberg.

The practical value and intellectual impact of size laws remain largely unknown to the public. An ironic illustration of this fact comes from columnist Kevin Maney in *USA Today,* writing about Benford's Law:

> Now, you'd think that those numbers, which are basically assembled randomly, would be spread out randomly. For instance, there'd be just as many numbers beginning with nine or four or one. But that's where you'd be wrong. Some unseen and unknown universal force—possibly similar to the cosmic force that impels all preteen girls to like the same pop star at the same moment—bunches these kinds of random numbers into very predictable patterns.

Size Law Applications 1881-2001
digit-frequency law (Newcomb 1881, Benford independently 1938)
machinery failure rates and 80-20 rule (Pareto, 1897)
income distributions (Pareto, 1897)
word-frequency and word-length models in language (Zipf ,1932)
modeling city population dynamics (Zipf , 1949)
portfolio management & firm sizes (Mandelbrot , 1963)
estimating rounding errors in computers (Knuth, 1968)
bibliometric ranking for retrieval systems (Fairthorne 1969)
ecological population statistics (B. M. Hill, 1970)
detection of false data by auditors (Nigrini, Carswell and others, 1988-89, from a suggestion by Raimi, 1976)
fractal geometry (Mandelbrot, 1977)
modeling website traffic (Taqqu and others, 1994)
rank-ordering of earthquake statistics (Sornette and others, 1996)
distinguishing "real" digital images from computer-generated ones (DeKok, 1999)
forecasting time-series relationships (Brooks, 2001)

Size laws do indeed explain why millions of teenagers suddenly buy, for example, Britney Spears CDs. More precisely, they provide a predictive model, an empirical rule of thumb, that unless some kind of powerful constraint is applied to prevent it, the distribution of CD sales will tend to be logarithmic. Thusthere will tend to be hundreds or thousands of artists selling fewer than 50,000 copies apiece, versus a handful selling in the millions or tens of millions.

This high ratio of "also-rans" to winners is not the result of a conspiracy by music industry executives, nor some sort of tragic flaw in capitalism. It is an expression of a much deeper natural order. I would argue that essentially the same "invisible hand" dictates that the Amazon carries 20 percent of the world's fresh water, and that a handful of Internet sites carry a high proportion of its total traffic.

Is there really one law?

Skeptical readers might protest that there is as yet no proof that Zipf's Law, Pareto's Law, and Benford's Law are reflections of *one* unifying principle. Perhaps operating on the idea of an underlying unity is premature; one could even argue that size laws fall short of being "laws" at all, being more in the nature of empirical approximations.

Certainly nothing has been published to date demonstrating a fundamental unity. But it is worth remembering what Ralph Raimi observed about Benford's Law, back in 1976: those who object to treating that law as a genuine law could also have objected on similar grounds to Newton's laws of force and motion.

Skeptical challenges are both welcome and encouraged. But I believe that this is exactly the way a fundamental law becomes recognized as being a law. Gradually, over decades, diverse evidence accumulates and confidence grows in the strength of the "description" the law provides.

'Forecasting Using Benford's Law' and 'ontics'

The forecasting technique I discuss in this issue's main essay rests on close to a decade of part-time research. So far as I know (and Mark Nigrini concurs in this), it has not previously been proposed anywhere.

'Ontics' is a term I have coined from the Greek *ontos*, meaning 'entity'. I am proposing it as an omnibus term for what seems to be an emerging new branch of science, focused on the distribution of entities and the consequences thereof. Essentially, ontics is the study of "heavy-tailed" or logarithmic size laws. For more on this point, see www.ontics.com.

Our editorial philosophy

We plan to publish to two main audiences: (1) those who are interested in size laws for practical business or professional reasons, and (2) those who focus on the science, theory, history, and broader implications. In fact, these are not really distinct groups. Because the field is growing and changing so rapidly, to properly apply size law theory requires close attention to new developments.

Judging by the list at left, the audience is likely to be very diverse, and given the expectations in this Internet-driven era regarding charging for information, we are taking a two-tiered approach.

Fee-for-service aspects. For certain items, including *Frequencies* itself, and proprietary Ekaros products such as the forecasting software now in development, we will charge what we think the market will bear. Ekaros welcomes proposals and inquiries on the commercial side.

The information clearinghouse. The Ekaros website, host of *Frequencies*, will facilitate as broad an exchange of ideas as possible, open to everyone. We also plan to accept and distribute research papers, popular articles, reviews and letters.

This two-tiered approach has become increasingly common in recent years. Scientific and historical information that belongs in the public domain will remain available to everyone. Information with commercial applications will involve fees proportionate to its value.

Our goal is to follow the size law story wherever it leads. As our story develops, we intend to host seminars, online discussion groups, and other ventures to bring interested readers and writers together. Ekaros will also provide professional services, training, and software, for those seeking to actively exploit size law applications. We hope you'll join us.

Welcome to *Frequencies*!

Dean Brooks

Contributors wanted

We are actively soliciting new material. If you want to write for *Frequencies*, contact us at:

frequencies@ekaros.ca

How to subscribe

Frequencies is published quarterly in Vancouver, Canada. A two-year subscription is $275.00 US for eight issues, overseas mailing charges not included. The unused portion of your subscription will be refunded if for any reason you are not satisfied.

Please be sure to provide your full name, mailing address, e-mail and telephone number. Send your cheque in U.S. funds or international money order to:

Frequencies
c/o Ekaros Analytical Inc.
125A - 1030 Denman St., #408
Vancouver, B.C. V6G 2M6

Anyone needing to receive an invoice before making payment should mark their subscription request accordingly. E-mailed subscription requests should be directed to:

frequencies@ekaros.ca

***Frequencies* staff**

Editor and publisher Dean M. Brooks is an engineer and technical writer with 15 years' experience in defense, aerospace, and software. He was the editor for Mark Nigrini's groundbreaking *Digital Analysis Using Benford's Law,* published by Global Audit Publications in May 2000.

Editorial advisor Mark J. Nigrini has taught at St. Mary's University, Halifax, and at Southern Methodist University in Dallas. He has become known as the world's leading investigator in the field of digital analysis, which uses Benford's Law and related principles to find anomalies in data.

Contributing writers in this issue

Chris Aikman is a technical writer and amateur astronomer living in Vancouver, Canada.

In This Issue

Coming Next Issue

Variance & Ratio Analysis

by David G. Coderre

Auditors rely on the practice of ratio analysis to detect abnormal or fraudulent transactions. This technique relies on consistent ranked size distributions in data sets. Learn about the fundamentals of ratio analysis, including case studies from the author of *Fraud Detection: Using Data Analysis to Detect Fraud*.

Zipf Ranking, Benford's Law, and Ratio Analysis

by Dean Brooks

Although Zipf's ranking law closely resembles Benford's Law, it is subtly different. The values generated by Zipf's system do not follow Benford's Law. Auditors rely both on ranking-based ratio analysis, and Benford-based digital analysis. Find out how Zipf's Law and Benford's Law are related, and what practical impact this has on detection of anomalies.

Available at www.ekaros.ca/frequencies

Getting Started with Benford Forecasting

by Dean Brooks

The mathematical supplement to this issue's feature article will show you how to perform a basic Benford forecast.

The Psychology of Number Invention: How Prophets of Doom Make Predictions Conforming to Benford's Law

by Dean Brooks

A fundamental principle of digital analysis is that invented numbers tend not to conform to Benford's Law. However, in some exceptional circumstances, they do. Find out how frauds in the paranormal succeed in meeting this test even though most business frauds cannot—and what stronger detection techniques are available to beat their methods.