Zipf's Law

HAVE YOU EVER NOTICED THAT A LIST OF COMPANIES ordered by revenue often resembles a list of cities ordered by population? That is, the largest company or city would usually be significantly bigger than the second-largest company or city, and so on. Rarely would one see a market in which two or three companies had the same market share, or a country with two or three of the largest cities being roughly the same size.

These types of observations are generally considered examples of Zipf's Law. George Zipf was a linguist and Harvard University professor in the first half of the last century. He noted that the second most common word in the English language ("of") appears at approximately half the rate of the most common word ("the"). The third most common word ("to") appears at approximately one-third the rate of the most common word.

In general, a phenomenon is said to follow Zipf's Law if the frequency of occurrence of the event (*P*), as a function of its rank (*i*), is a power-law function $P_i \sim 1/i^a$ with the exponent *a* being near 1.00.

Of course, Zipf's Law is really not a law at all. It's merely a simple mathematical model that appears to describe some human behavior.

Incidentally, researchers have discovered that almost all languages follow Zipf's Law. This turns out not to be surprising since it's been shown that even random sequences of 26 letters and a space will also follow Zipf's Law. A word is considered a random sequence of letters separated by spaces on each side.

This occurs because short random words (a space separates the random letter sequences into words) are more common than longer words.

Second Place

While George Zipf may have been describing a property of random numbers more than a human preference, Zipf's Law does appear to accurately model a number of human phenomena.

So far, we've mentioned three different situations that could be modeled using Zipf's Law: revenue of companies, population of cities, and frequency of words. However, Zipf's Law has been used to describe a variety of incidents such as:

Web site traffic, sand particle sizes, meteor impacts on the moon, areas burnt in forest fires, earthquakes, genetic factors in cancer, individual incomes, and number of posts by poster on a web forum.

Metro areas of the United States appear to follow Zipf's Law to some degree. However, the No. 2 metro area, Los Angeles, is significantly bigger than the Zipf model would predict. Perhaps the geographical distance between Los Angeles and New York has allowed Los Angeles to grow beyond the typical second-ranked population center in a country (see Fig. 1).

How well does Zipf's Law work in insurance? Figure 2 shows the application of Zipf's Law for the Top 10 groups for all auto lines in 2001.

What does this mean for market share dynamics? Probably not much if you're not the No. 1 or No. 2 player in a market. But a No. 2-ranked company might strategize on growing to be the same size as the No. 1 player. If the market share dynamics follow Zipf's Law, the currently ranked No. 1 and No. 2 companies wouldn't be similar sizes. So this strategy is likely to fail; either the No. 2 player will stay No. 2, or the growth of the No. 2 player will take market share from the player No. 1. The market share rank of the players will reverse.

So what explains the wide variety of human behavior that apparently is under the control of such a wide law? Does Zipf's Law occur because of some hidden human psychology that manifests itself into organizing by power law? Or is it inevitable that societies will organize by power law?

William J. Reed of the University of Victoria makes a convincing case that it's the latter. One component of Zipf's Law that's not a random variable is the component of time. For example, in the market share of auto insurance companies, there's a time (around 1900) when no auto insurance was written. During the next 100 years, companies began writing auto insurance. In general, the rate of growth (or decline) of these companies is an exponential process.

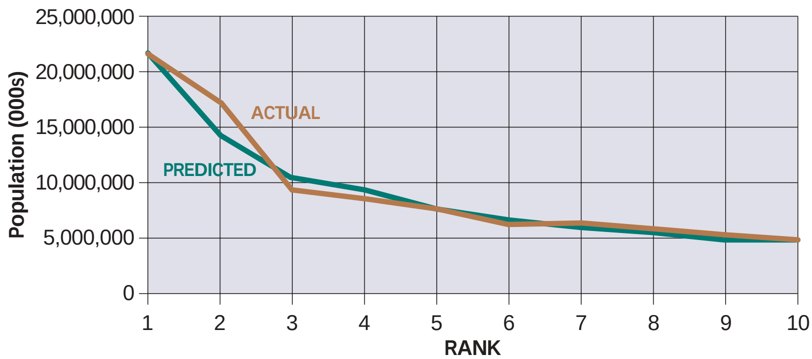
Now, 100 years later, we're examining the largest of all companies who have written auto insurance (the tail of the distribution). While the distribution of premium was for one year only, there is likely a strong relationship between the sizes of firms in the current year and their size in the

LEE BOWRON IS MEMBER AND MANAGER OF KERPER AND BOWRON, LLC, IN BIRMINGHAM, ALA.

FIGURE 1

RANK	METRO AREA	ACTUAL	PREDICTED
1	New York	21,600,000	21,600,000
2	Los Angeles	16,800,000	13,765,255
3	Chicago	9,400,000	10,576,085
4	Washington	7,850,000	8,772,326
5	San Francisco	7,250,000	7,587,929
6	Philadelphia	6,300,000	6,739,931
7	Boston	5,900,000	6,097,335
8	Detroit	5,850,000	5,590,431
9	Dallas	5,500,000	5,178,406
10	Houston	4,900,000	4,835,638

Zipf's Law—Top 10 Metro Areas (Alpha = .65)



Source: www.citypopulation.de

prior year. This is true for income and city size as well. The best indicator for this year's population for a city is probably last year's population.

It can be shown that the tail of this exponential distribution will follow a power-law distribution. Therefore, any rank ordering of these processes at a certain time will tend to follow Zipf's Law.

Even some natural worlds may be subject to the same time constraints. For example, forest fires might be considered to be exponential-growth processes until stopped at random by rain or lack of fuel.

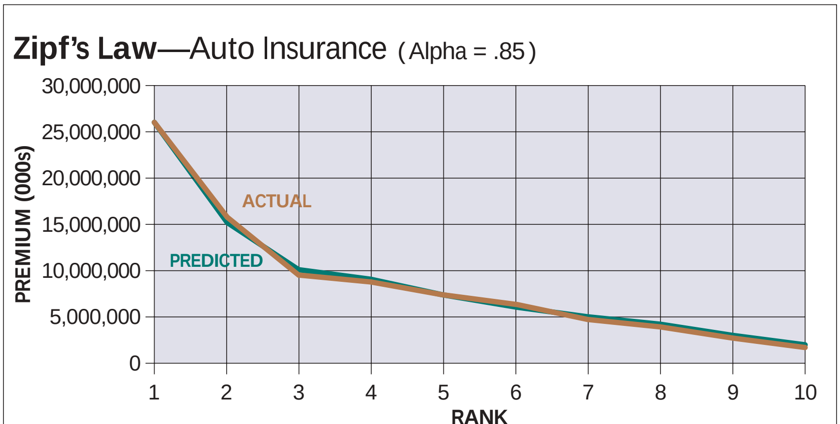
Net worth and income distribution of households have attracted political interest in the United States and other countries. The disparate distribution of income is perceived by some as bad public policy. Indeed, the United States has a high degree of disparity of net worth and income. This is often debated in equity and taxation issues.

The inequity in wealth, however, may be an inevitable function of the exponential process that drives net worth and income. Since net worth and income can be thought of as exponential processes that grow (or shrink) over time, they follow a Zipf's Law-type of distribution.

The economist Paul Krugman quantifies the well-known "the rich are getting

FIGURE 2

RANK	METRO AREA	ACTUAL	PREDICTED	RANK	METRO AREA	ACTUAL	PREDICTED
2	State Farm	25,572,683	25,572,683	6	Berk./Hath.	6,199,297	5,428,445
2	Allstate	15,200,422	14,040,590	7	USAA	4,522,861	4,750,797
3	Zurich/Farmers	9,065,206	9,887,043	8	Travelers/Citigrp.	3,946,760	4,232,563
4	Progressive	7,222,716	7,708,935	9	Liberty Mutual	3,701,427	3,822,579
5	Nationwide	6,743,054	6,355,756	10	Amer. Intern.	3,262,745	3,489,605



Source: A.M. Best

richer" argument by noting that the top 20 percent have grown wealthier than the bottom 80 percent. In addition, the top 5 percent have grown faster than the next 15 percent (the 80th to 95th percentile). The top 1 percent has grown faster than the 95th to 99th percentile.

Any political solution to income distribution inequality will need to consider that income inequality is actually the norm and not an exception. But there are exceptions to the rule.

One that comes to mind covers Coke™ and Pepsi™. Both these companies have similar market shares in the soft-drink industry. But it's also interesting that the "Cola War" is considered one of the great business competitive battles of all time. Indeed, this might be because this is one area of business where there is no clear leader; where Zipf's Law doesn't hold.

Most actuaries generally consider exponential processes elementary. Loss trends and compound interest don't require complicated theories or mathematics.

But exponential processes may lend themselves to more interesting results. The fundamental ordering of societies in areas such as the relative size of our cities, the relative size of business firms, and the distribution of national wealth may be the result of the simple exponential process that governs their change. ●