1
THE LOG-NORMAL DISTRIBUTION

S. Sinha

TOPICS TO BE COVERED IN THIS CHAPTER:

- The Log-normal distribution: another long-tailed distribution
- The ubiquity of log-normal distribution
- Is it Power-Law or Log-normal ? Why the two are often confused
- Testing empirical data for statistical features
- Maximum likelihood estimation, Jacknife and Bootstrap
- Hypothesis testing and statistical significance
- Other statistical tests: Kolmogorov-Smirnov tests etc
- Why log-normal appears so frequently in economics ?
- Gibrat's Law of Proportionate Effect: the multiplicative random walk
- *Bio-box on R Gibrat*
- The Yule-Simon model
- Extreme value distributions
- *Bio-box on E J Gumbel and W Weibull*

Appropriate quotation here

Econophysics. Sinha, Chatterjee, Chakraborti and Chakrabarti

ISBN: 3-527-XXXXX-X

1.1
The Log-normal distribution

1.2
The Law of Proportionate Effect

Robert Gibrat's doctoral thesis on economic inequality (1931) pioneered the use of the lognormal form in studying distributions arising in the economic domain, such as that of income. According to Gibrat, the income of an individual (or a firm) can be thought of as arising from a combination of large number of independent processes, each of which may have evolved over a long time. We will consider processes which evolve in discrete time, i.e., $t = 1, 2, 3, ...$ At a given time t, the change in a variable of interest, x (say), is a random fraction of a function of its value at the previous instant:

$$x_t - x_{t-1} = \epsilon_t f(x_{t-1}). \quad (1.1)$$

The random variables ϵ are the output of an uncorrelated stochastic process and do not depend on x. The *law of proportionate effect* of Gibrat corresponds to the case when $f(x) = x$, so that the change in the variable is a random fraction of its value at the previous instant. Taking logarithm of the variable, we can express the its value at any instant n, as:

$$\ln x_n = \ln x_0 + \Sigma_{t=1}^{n} \epsilon_t, \quad (1.2)$$

where x_0 is the value of the variable at some initial time $t = 0$. Then, using the central limit theorem, we can see that ln x will follow a normal distribution, thereby making x log-normally distributed.

The probability distribution function of the lognormal distribution is

$$p(x) = \frac{1}{x\sqrt{2\pi\sigma}} exp[-(1/2\sigma^2)(\ln x - \mu)^2], \quad (1.3)$$

where $x > 0$, and the mean and variance of the distribution are given by $E(x)$ = exp($\mu + \frac{1}{2}\sigma^2$) and var ($x$) = (exp($\sigma^2$) - 1) exp($2\mu + \sigma^2$).

==

Robert Pierre Louis Gibrat (1904-1980)

Robert Gibrat was educated at the Ecole Polytechnique in Paris and specialized in mining engineering at Ecole des Mines, before beginning his career as a technical consultant. During 1927-31 he was a professor at St Etienne Ecole des Mines. His 1931 doctoral thesis (at Lyon) on economic inequality was a historical study of the lognormal distribution and its use in modeling, using a single parameter, the distributions of wealth, industrial concentrations, city populations, family statistics, etc. It also contained his formulation of the "law of proportionate effect", Gibrat's principal contribution to economics. Although using the term "law" (like Pareto), Gibrat confessed the essentially statistical character of the principle. Gibrat was made a Fellow of the Economteric Society in 1948 primarily for the developments arising from this contribution. Afterwards Gibrat returned to Paris and in 1936 became a Professor at the Ecole des Mines. In later life, the study of nuclear energy and the harnessing of hydrolelectric energy of tides and rivers occupied Gibrat. During the Nazi occupation of France, Gibrat occupied high government positions in the puppet French government, for which he spect a year in prison after the liberation of France as a collaborator. In prison, Gibrat developed the theory for a new type of hydraulic power plant. After release, he resumed his engineering consulting, both on tidal energy and the development of French atomic policy. Among the many honors he received, Gibrat was a Knight of the Legion of Honour and was president of the Scientific Committee of EURATOM (European Atomic Energy Community) in 1962. Gibrat had a great passion for languages, not only European ones such as English, German, Italian, Spanish, Norwegian and Finnish, but also Japanese and Chinese, and just before his death had started a systematic study of Sanskrit.

==

1.3 Extreme Value Distributions

On October 19, 1987, the Dow Jones Industrial Average dropped by 29.2%, the worst day of trading until that time in at least a century. Based on a theory of normally distributed fluctuations, the probability of such a cataclysmic event happening was less than one in 10^{50}. Such extreme events are rare, but occur much more often than conventional theories of markets would lead us to belive. These events are difficult to predict, but must be taken into account into any proposal for assessing the risk of market investmets. Is it possible to arrive at a theory that better approximates their observed frequency ? The fact that the distribution of fluctuations (e.g., of individual stock price or of the index for the entire market) actually seem to follow a power law at short time-scales should make us expect that large events are indeed more common than expected from a Gaussian distribution. But other schools of thought hold that large crashes are special, and they need to be explained using other distributions. This is where the theory of extreme value distribution comes in. Their application in the financial context not only include major crashes in the stock market, but also collapses of large banks and firms, often following each other in a cascading effect.

The probability of extreme events depends to a large extent on how the pdf $p(x)$ decays to zero as $|x| \to \infty$. The rate of decay necessarily has to be estimated from the empirical data. As (almost by definition) extreme events are rare, this estimation is not a simple task, as even very large quantities of data will give us only limited information about the true probability of a very high return (positive or negative) in the market (say). In this context we can use the methods of extreme value statistics to give us a more reliable estimate of the risk involved.

Let us consider the maximum value X for a set of idenpendent random variables $x_1, x_2, \ldots, x_n$, which are obtained from identical distribution. From the perspective of financial markets, one can think of X as $|r_{min}|$, the highest negative return. Only three distributions can be the asymptotic limit of the distribution for X, which are collectively referred to as *extreme value distributions*. The cumulative distribution functions for the three are as follows:

Gumbel distribution: $P(X \leq x) = exp(-e^{(x-\mu)/\sigma})$,

Frechet distribution: $P(X \leq x) = exp(-[(x-\mu)/\sigma]^{-\xi})$ if $x \geq \mu$, $= 0$, otherwise.

Weibull distribution: $P(X \leq x) = exp(-[(\mu - x)/\sigma]^{-\xi})$ if $x \leq \mu$, $= 0$, otherwise.

Thus, the Frechet and Weibull distributions are concentrated on the positive and negative real numbers, while Gumbel distributed variables can have any real number value. Note that, Frechet and Weibull are related to the Gumbel form by a log-transformation of the variable, i.e., $x' \to \log(\mu - x)$ or $\log(x - \mu)$. Also, note that the complementary cumulative distribution function for Weibull is the well-known stretched exponential function $e^{-}(x/\lambda)^k$, where λ and k are often referred to as scale and shape parameters of the distribution, respectively.

Measuring risk is a vital criterion for most present-day financial instruments. For example, the Black-Sholes formula for option pricing uses a single input for empirical market data that is a measure of the risk, measured as standard deviation of price fluctuations. Indeed, in calculating the risk for a given market portfolio, often the weighted combination of the standard deviations of all stock prices compirising it is used. Thus, it is assumed that the standard deviations contains all relevant information about risk. However, it is not an intuitive measure of risk, which is thought of in terms of amount of money lost rather in terms of units of deviations from a mean. More importantly, the fluctuations occurring below the expected return may not have the same likelihood as those above, wehereas using the standard deviation as a measure of risk assumes symmetric deviations. Therefore, another measure of risk was deemed necessary, one that would estimate the loss associated with a given small probability of deviation occurrence, i.e., the possibility of losing a certain amount of money over a given holding period of an investment. In other words, higher risk will imply a higher loss at the given probability. This was the purpose that the quantity Value at Risk (VaR) was designed to serve. When the returns are symmetric about the mean, the information conveyed by VaR is exactly the same as standard deviation except for a scaling factor.

====== The general approaches to VaR computation have fallen into three classes called parametric, historical simulation, and Monte Carlo. Parametric VaR is most closely tied to MPT, as the VaR is expressed as a multiple of the standard deviation of the portfolio's return. Historical simulation expresses the distribution of portfolio returns as a bar chart or histogram of hypothetical returns. Each hypothetical return is calculated as that which would be earned on today's portfolio if a day in the history of market rates and prices were to repeat itself. The VaR then is read from this histogram. Monte Carlo also expresses returns as a histogram of hypothetical returns. In this case the hypothetical returns are obtained by choosing at random from a given distribution of price and rate changes estimated with historical data. Each of these approaches have strengths and weaknesses.

The parametric approach has as its principal virtue speed in computation. The quality of the VaR estimate degrades with portfolios of nonlinear instruments. Departures from normality in the portfolio return distribution also represent a problem for the parametric approach. Historical simulation (my personal favorite) is free from distributional assumptions, but requires the portfolio be revalued once for every day in the historical sample period. Because the histogram from which the VaR is estimated is calculated using actual historical market price changes, the range of portfolio value changes possible is limited. Monte Carlo VaR is not limited by price changes observed in the sample period, because revaluations are based on sampling from an estimated distribution of price changes. Monte Carlo usually involves many more repricings of the portfolio than historical simulation and is therefore the most expensive and time consuming approach. =======

The Gumbel distribution has been shown to be useful for describing certain aspects of Asian stock markets

A. Da Silva and V. De Melo Mendes, Value-at-risk and extreme returns in Asian stock markets. Int. J. Business 8 (2003) 17-40

==

BIO-BOX ON E J GUMBEL AND W WEIBULL

Emil J. Gumbel (1891-1966) a German mathematician and political writer. He graduated from the University of Munich shortly before the outbreak of the First World War. He was Professor of Mathematical Statistics at the University of Heidelberg. Following the murder of a friend he investigated several political murders and published his findings in Four Years of Political Murder in 1922. He spent much of his life studying the statistics of extreme values, (rare events). Born in Germany he resisted the secret rearmament of Germany after World War I and left. He did much of his research in the USA at NBS, now NIST. He and Waloddi Weibull did a sabbatical together at Columbia University and became good friends. Gumbel was fascinated by the fact that the Weibull distribution and Extreme Value Type III minimum are the same distribution. Extreme Value Type I is called the "Gumbel" distribution. It is employed for predicting maximum and minimum values, flood levels, wind gusts, the size of inclusions in metal. The Weibull and the Gumbel minimum are related like the normal and the log normal through a logarithmic transformation. One of his greatest contributions was to prove that if a part or component had multiple failure modes and our interest was in the first failure, Type III minimum, the Weibull, is the appropriate distribution. This idea is called the "weakest link in the chain" theory and is the reason that the Weibull is the world's most popular distribution for life data analysis. (For more details see www.bobabernethy.combios_stats.htm) Wrote "Statistics of Extremes" in 1958.

One of the most important figures in the German pacifist movement, Gumbel was the leading chronicler of the numerous political murders that were common in the Weimar Republic. He uncovered the secret rearmament of Germany after the First World War and identified groups that carried out acts of terror, including the National Socialists. As a professor of statistics in the University of Heidelberg, he became the focal point for a prolonged cotroversy with the largely anti-Republican professors and the pro-Nazi students. Forced into exile, he became an important figure in the German political and iltellectual exiled community in France and later, the United States.

Gumbel's doctoral dissertation was an exploration of methods to determine the population of a nation or region between censuses. This set the tone for much of his research in the 1920s and early 1930s which mainly concerned population statistics and life expectancy. During the early 1920s he also published a few scholarly articles on statistical problems in physics. Gumbel studied with several leading statisticians, including Georg von Mayr in Munich, and carried out correspondence with Karl Pearson, in England. His years in exile in France were marked by modest professional success. At the University of Lyon, where he was between 1934-1940 he progressed from lecturer to *Maitre des Recherches* and was on track to become a director of research had not WWII intervened. In the late 1930s Gumbel shifted his research focus from social questions to the statistical analysis and prediction of natural and physical phenomena, He developed the theory of extreme values to answer questions like: "How frequently can a flood (or earthquake or rainstorm or wind) of a given magnitude be expected to occur ?" From the late 1930s to the early 1950s, he worked exclusively on calculating flood flows (for determining the required strength of dams) and the breaking strength of metals. During the German occupation of France, Gumbel had to flee to Portugal and sailed from there to United States. In USA, Gumbel initially spent four years at the New School. During the war, he also worked for several government agencies, inclduing the Office of Strategic Services (the precursor of CIA) and the US Department of Interior's Geological Survey. From 1945 to 1952, except for a brief period as a professor at Brooklyn College, Gumbel did not have any full-time employment. In 1953, he was appointed adjunct professor in engineering in Columbia University, where he continued to work for the rest of his life. The crowning work of Gumbel's career was "Statistics of Extremes". The product of a decade of labor, it was first published in 1958 and immediately won high regard in the international scientific community.

E. H. Waloddi Weibull (1887-1979) was a Swedish physicist and engineer, who applied the ideas of extreme values to the study of strength of materials and in the process came up with one of the extreme value distributions that now carries his name. He joined the Swedish Coast Guard in 1904 as a midshipman. Weibull moved up the ranks eventually becoming a Major in 1940. While in the coast guard he took courses at the Royal Institute of Technology, Stockholm. He graduated in 1924 and became a full professor. Weibull obtained his doctorate from the University of Uppsala in 1932. He worked as consulting engineer for various Swedish and German industries. His first scientific paper was on the propagation of shock waves generated by an explosion (1914). As part of this interest, he took part in voyages to Mediterranean Sea, the Caribbean abd the Pacific Ocean aboard a reserch ship to develop a technique for determining the type and thickness of ocean bed sediments by using data obtained from explosive charges. In 1939, he published his first paper on the connection between statistics of extreme values and the phenomenon of rupture in solids. E J Gumbel later showed that the distribution proposed by Weibull was identical to one of the three extreme value distributions. Further, he proved that if the failure of a system depends on the probability of the failure of any one of its multiple parts (the concept that a chain is only as strong as its weakest link), then the time to first failure is best modeled by this distribution. In 1941, Weibull became a research professor in Technical Physics at the Royal Institute of Technology in Stockholm. In later life, he continued to publish on many aspects of strength of materials and fatigue. He died on October 12, 1979 in Annecy, France.

==

References