Short communication

Scalable functions used for empirical forecasting

Abstract

Taleb in his book the Black Swan introduces a concept of scalable functions. Here it is shown that the only scalable functions are power-law functions and they can be treated as one and the same. Moreover, the analytical problems of these functions are discussed.

Keywords - distribution functions, gaussian, scalability

Extreme events, ’outliers’, have tiny occurrence, but they will happen every now and then. Nassim Taleb calls these outliers Black Swans[1]. In his words, Black Swans are 1) Outliers, 2) Carry extreme impact, 3) Humans concoct explanations after the fact, making them explainable and predictable (a posteriori). The name derives from the fact that scientific theories were constructed that swans necessarily have to be white ... until a black swan was encountered, an unlikely event, but happening nonetheless. These unlikely events happen in many systems and are not well described by common analytical techniques and forecasting models. (Empirical) forecasters, the primary readers of this journal, doing forecasting based solely on past events, go for proven useful models. ”Forecasting is based on history. Many supply chain and financial forecasting models are based on bell curves, with the tails of the curve representing highly unlikely events that can be ’safely’ ignored, at least for forecasting purposes.”[2].

While the book of Taleb is a milestone in empirical forecasting – basically holding a mirror in front of us, showing how we are often fooling ourselves into thinking we have everything under control – he introduces some concepts in a non-scientific way. In particular, he introduces the concept of ’scalable functions’ and scalability in general to explain some phenomena.

These functions are compared to the more often used non-scalable functions, that are mentioned to be ’Gauss’, ’normal’, or ’bell-curve’, which are all synonyms and very well studied and documented (See Ref. [3] for a historical overview). The bell curve is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{1}$$

with σ and μ parameters. Generally, they are form the family of exponential functions like $f(x) = a^{-x^b}$ (with the bell curve having $a = e$, and $b = 2$). The most important aspect, in view of empirical analysis and forecasting is that they have ’short tails’, meaning that ’outliers’ – values far away from the median μ – have a very tiny probability of occurrence. That makes them docile and the favorite tool of analysts/forecasters. Contrasting them are the 'long tail' functions, where the outliers also have a small frequency of occurring, but not small enough to keep you on safe grounds. For these distributions every now and then an outlier will occur that will mess up everything, like the collapse of the financial system. This according to Taleb, who summarizes his strong criticism on (financial) forecaster's use of bell curves by the phrase "fooled by the reductionist need for simple functions". Fendler and Muzaffar make a similar statement in their aforementioned historical overview, "Our purpose is to criticize [...] the belief that a normal-curve distribution is a representation of real things in nature"[3]. A viewpoint that is shared by many people, for instance Dudley-Marning[4], Goertzel and Fashing[5]. The latter cite Bradley's informative phrase ".. the experimenters fancy that [the bell curve] is a theorem in mathematics and the mathematicians that it is an experimental fact". Basically, most forecasting is done using docile functions like the Gauss curve and its derivatives that are based on the assumption that the system under study is a (multiple)-random-events process, i.e., has *probability* character, like flipping coins or throwing dice. Assumptions of Gauss (normally-distributed) and Markov (history-independent)[6] processes are used everywhere (Ex. risk evaluation[7], chaotic algorithms[8], (physical or data) traffic[9, 10], Etc.). (Indeed, often Monte-Carlo simulations are made of the system). This not because of having theoretically proven that it is correct, but because it is a simple and very powerful analytical tool. Even if not always 100% correct, it is better than using unusable tools. Most literature in forecasting journals is based on this. (For sheer abundance, no references supplied here). (Empirical) forecasting is often based on past events that are assumed to be the results of random, probabilistic processes.

While not going into that controversial discussion here, an important question arises: What are scalable functions exactly? It would be nice to know what are the limitations to the class of scalable functions. Are there scalable functions that do not have a long tail, or vice versa? Scalable functions that will not mess up our analysis? Why do scalable functions mess up our analysis? In this short communication we address this issue.

Scalability is defined by Taleb as the fractal property of a function $f(x)$ that any multiplication of x to ax introduces a constant factor in the function value, independent of x[1];, i.e., they 'look' the same everywhere. In an equation,

$$\frac{f(ax)}{f(x)} = \frac{f(ax')}{f(x')} = C(a). \tag{2}$$

For instance, an earthquake of Richter (logarithmic) magnitude m (amplitude $x = 10^m$) has $10^{3.04}$ times more occurrence than a ten-times stronger earthquake of magnitude $m+1$, independent of m; $C(a) = a^{-3.04}$. It is easy to show that the Gaussian is not scalable. (The ratio above is $C(x,a) = \exp[-x^2(a^2-1)]$, not independent of x). Which functions are? The above equation can be solved to find the set of functions that are scalable. In the first step, noting that the fraction is constant, we take the logarithm on the left side, divide it by $\ln(a)$ and letting this factor go to zero,

$$\lim_{\ln(\mathrm{a})\to 0} \frac{\ln[f(ax)] - \ln[f(x)]}{\ln(a)} = c. \tag{3}$$

($c = \ln(C)$). The left side here is the definition of a derivative of the function on a log-log scale ($\ln(ax) - \ln(x) = \ln(a)$). Therefore, the function in a log-log plot is a line, with constant derivative,

$$\begin{aligned}
\frac{\mathrm{d}\ln[f(x)]}{\mathrm{d}\ln(x)} &= c, \\
\frac{\mathrm{d}\ln[f(x)]/\mathrm{d}x}{\mathrm{d}\ln(x)/\mathrm{d}x} &= c, \\
\frac{\frac{\mathrm{d}\ln[f(x)]}{\mathrm{d}(f(x))} \cdot \frac{\mathrm{d}(f(x))}{\mathrm{d}x}}{1/x} &= c, \\
\frac{\frac{1}{f(x)} \cdot \frac{\mathrm{d}f(x)}{\mathrm{d}x}}{1/x} &= c, \\
\frac{\mathrm{d}f(x)}{\mathrm{d}x} &= c\frac{f(x)}{x}.
\end{aligned} \tag{4}$$

This is a differential equation which has as unique solutions the power-law functions,

$$f(x) = f_1 x^{-\alpha}, \tag{5}$$

with f_1 a constant being the function value at $x = 1$ and α a constant ($\alpha = -c$) denoting the power (for instance, earthquakes follow power $\alpha = 3.04$[11]) – the lower the number, the more the system is pestered by outliers[1]. (See Figure 1 for an example with $\alpha = 0.5$). Thus, we conclude that 'scalable', 'fractal' and 'power-law' are all one and the same thing; the only functions that are scalable are power-law functions.

Here the work of Newman should be mentioned who made an excellent summary of the power-law functions[11]. Basically, power-law functions are very difficult to work with; they can behave like Gaussian functions (thus sometimes making us believe the data are following the bell curve), especially for small numbers of samples, but they lack the parameter σ and drawing conclusions on basis of this parameter can thus not be done.

Why are the scalable functions so difficult to work with? The function of probabilities, for instance for the lottery, is by definition integrable (and integrated to unity; *somebody* has to win). Not so for the frequency-of-occurrence (empirical analysis) functions of many natural phenomena. A power-law function (Eq. 5) cannot be integrated over all possible values of x (zero to infinity), since it results in infinity, for any value of α, either on the zero-side or on the infinite tail of x. (Often, the function is 'truncated' by setting arbitrary lower or higher limits to x, for instance the strongest earthquake ever measured, until now, to force it to be usable, integrable, because that is so much needed. Yet, in some cases the truncation is natural, like setting the minimal word length to 1 letter, where word-length and frequency-of-use follow a power 2.20). It can thus also not be scaled to result in unity when integrated (Interestingly, or confusingly, scalable functions cannot be scaled), and these functions cannot represent probability-density functions. 'Probabilities' are not

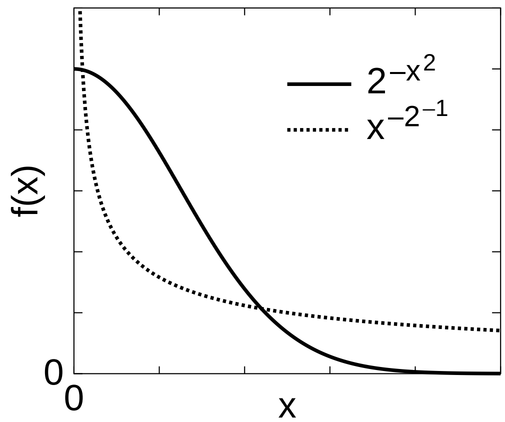

Fig. 1: Two types of functions (half shown, only positive x), non-scalable (like Gaussian, or the similar 2^{-x^2} shown with solid line), and scalable, power-law (like the shown $1/\sqrt{x}$, dashed line). The former is docile and integratable, the latter not (the shown example because of the long tail, not going fast enough to zero; for other exponents, problems can occur at $x = 0$)

defined; only real-events statistics remain. As, such, the average expected magnitude of a future event,

$$<x>= \frac{\int_0^\infty x f(x) \mathrm{d}x}{\int_0^\infty f(x) \mathrm{d}x} \tag{6}$$

cannot be calculated on basis of empirically-found parameters (f_1 and α). In other words, the average magnitude is not defined for such phenomena Ex. earthquakes; One cannot make a statement about the predicted average earthquake next year. The past has an average, the future not (yet). There are some once-in-a-while-occurring high-magnitude outliers – Black Swans – that mess-up the analysis.

Finally, interesting to point out, the scalability can also be defined as the property that the derivative, normalized by the scaled function, is constant

$$\frac{\mathrm{d}f(x)/\mathrm{d}x}{f(x)/x} = c. \tag{7}$$

This is just saying the same as Taleb with other words, and was part of the demonstration of Eq.

(4).

In conclusion, we have shown here that the only functions that are 'scalable' ('fractal') as defined by Taleb are power-law functions of Eq. (5). Moreover, we have shown why it is that they give problems in the analysis, which basically stems from the fact that they are not-integratable (Eq. (6)) and thus they are not normalizable into a probability function.

References

[1] Taleb NN. 2011. The Black Swan: The Impact of the Highly Improbable. Allen Lane. ISBN 978-1846145650.

[2] http://logisticsviewpoints.com/2010/08/04/turkeys-forecasting-and-supply-chain-risk-management/ visited September 2013.

[3] Fendler L, Muzaffar I. 2008. The history of the bell curve: sorting and the idea of normal. *Educational Theory* **58**: 63.

[4] Dudley-Marling C, Gurn A. 2010. The Myth of the Normal Curve. *Series: Disability Studies in Education Volume 11*, Peter Lang International Science Publishers. ISBN 978-1-4331-0729-0.

[5] Goertzel T, Fashing J. 1981. The Myth of the Normal Curve: A Theoretical Critique and Examination of its Role in Teaching and Research. *Humanity and Society* **5**: 14-31.

[6] Ibe OC. 2013. *Markov Processes for Stochastic Modeling.* DOI: http://dx.doi.org/10.1016/B978-0-12-407795-9.00003-7

[7] Wong YK, Rad AB. 1998. Gauss-Markov models for forecasting and risk evaluation. *Proceedings Energy Management and Power Delivery 1998.*

[8] Hong W-C , Dong Y, Zhang WY, Chen L-Y, Panigrahi BK. 2013. Cyclic electric load forecasting by seasonal SVR with chaotic genetic algorithm. *Electrical Power Energy Systems* **44**: 604.

[9] Goerlandt F, Kujala P. 2011. Traffic simulation based ship collision probability modeling. *Reliability Engineering and System Safety* **96**: 91107.

[10] Krithikaivasan B, Deka1 K, Medhi D. 2004. Adaptive Bandwidth Provisioning Envelope based on Discrete Temporal Network Measurements. *IEEE INFOCOM 2004*.

[11] Newman MEJ, Power laws, Pareto distributions and Zipfs law. *Contemporary Physics* **46**: 323-351.