Black Swans and the Domains of Statistics

Nassim Nicholas Taleb

1. INTRODUCTION

The Black Swan: The Impact of the Highly Improbable (hence *TBS*) is only critical of statistics, statisticians, or users of statistics in a very narrow (but consequential) set of circumstances. It was written by a veteran practitioner of uncertainty whose profession (a mixture of quantitative research, derivatives pricing, and risk management) estimates and deals with exposures to higher order statistical properties. Derivatives depend on some nonlinear function of random variables (often square or cubes) and are therefore extremely sensitive to estimation errors of the higher moments of probability distributions. This is the closest to *applied statistician* one can possibly get. Furthermore, *TBS* notes the astonishing success of statistics as an engine of scientific knowledge in (1) some well-charted domains such as measurement errors, gambling theory, thermodynamics, and quantum mechanics (these fall under the designation of "mild randomness"), or (2) some applications in which our vulnerability to errors is small. Indeed, statistics has been very successful in "low moment" applications such as "significance testing" for problems based on probability, not expectation or higher moments. In psychological experiments, for instance, the outlier counts as a single observation, and does not cause a high impact beyond its frequency.

TBS is critical of some statistics in the following areas:

1. The unrigorous use of statistics, and reliance on probability in *domains* where the current methods can lead us to make consequential mistakes (the "high impact") where, on logical grounds, we need to force ourselves to be suspicious of inference about low probabilities.

2. The psychological effects of statistical numbers in lowering risk consciousness and the suspension of healthy skepticism—in spite of the unreliability of the numbers produced about low-probability events.

3. Finally *TBS* is critical of the use of commoditized metrics such as "standard deviation," "Sharpe ratio," "mean-variance," and so on in fat-tailed domains where these terms have little practical meaning, and where reliance by the untrained has been significant, unchecked and, alas, consequential.

Let me summarize the aims of *TBS*. What one of the reviewers calls "philosophy" (a term that generally alludes to the sterile character of some of the pursuits in philosophy departments), owing perhaps to the lack of quantitative measures in *TBS*, I tend to call "risk management." That is, practical wisdom and translation of knowledge into responsible decision making. Again, for a practitioner "philosophy" is, literally, "wisdom," not empty talk. As put directly in *TBS*, it is about how "not to be a sucker." My aim of the book is "how to avoid being the turkey." It cannot get more practical (and less "philosophical" in the academic sense) than that.

Accordingly, *TBS* is meant to provide a roadmap for *dealing* with tail events by exposing areas where our knowledge can be deemed fragile, and where tail events can have extreme impacts. It presents methods to avoid such events by not venturing into areas where our knowledge is not rigorous. In other words, it offers a way to live safely in a world we do not quite understand. It does not get into the trap of offering another precise model to replace another precise model; rather it tells you where we should have the courage to say "I don't know," or "I know less."

2. CONFIDENCE ABOUT SMALL PROBABILITIES

I will next outline the "inverse problem" of the real world. Life is not an artificial laboratory in which we are supplied with probabilities. Nor is it an urn (alas) as in elementary statistics textbooks. Nor is it a casino where the state authorities monitor and enforce some probabilistic transparency (i.e., try to eliminate the uncertainty about the probabilities). Empirical estimation of probabilities poses a problem in domains with unbounded or near-unbounded payoffs. (*I am not assuming, which is key, that an upper or lower bound does not exist, only that we do not know where it is*.)

Suppose that you are deriving probabilities of future occurrences from the data, assuming (in the "rosy" case) that the past is representative of the future. An event can be a market crash, a banking crisis, a loss for an insurance company, a riot, people affected in an epidemic, an act of terrorism, and so on. The severity of the event here will be inversely proportional to its expected frequency: the so-called 10-year flood will be more frequent than the 100-year flood, and the 100-year flood will be more devastating. In these events, we are not sampling from a problem-style closed urn of known composition and impacts. We don't even know if there is a 200-year flood, and what impact it may have. We are now subjected to the classical problem of induction: making bold claims about the unknown based on assumed properties of the known. So (1) the smaller the probability, the larger we need the sample size to be in order to make inferences, and the smaller the probability, the higher the relative error in estimating this probability. (2) Yet in these domains, *the smaller the probability, the more consequential* the impact of the absolute probability error on the moments of the distribution.

Estimation errors for tail probabilities are very important when their large impact is considered. The pair probability *times* impact is a rectangle that gets thinner as probabilities becomes smaller, but its area can become more stochastic if the probabilities do not drop too quickly as the impact becomes larger. This is clearly intractable. It can be solved on paper, of course, by assuming a priori a certain class of distributions. Indeed the choice

Nassim Nicholas Taleb is a veteran derivatives trader and researcher, London Business School and Empirica Laboratory Limited (E-mail: *gamma@fooledbyrandomness.com*).

of distributions with characteristic scale—that is, what Mandelbrot defined as "mild randomness," more on that later—appears to conveniently push such problems under the rug.

3. SELF-REFERENCE

This problem has been seemingly dealt away with the use of "off-the-shelf" probability distributions. But distributions are self-referential. Do we have enough data? If the distribution is, say, the traditional Gaussian, then yes, we may be able to say that we have sufficient data—for instance, the Gaussian *itself* tells us how much data we need. But if the distribution is not from such a well-bred family, then we may not have enough data. But how do we know which distribution we have on our hands? Well, *from the data itself.*

So we can state the problem of self-reference of statistical distributions in the following way. If (1) one needs data to obtain a probability distribution to gauge knowledge about the future behavior of the distribution from its past results, and if, at the same time, (2) one needs a probability distribution to gauge data sufficiency and whether or not it is predictive outside its sample, then we are facing a severe regress loop. We do not know what weight to put on additional data. And unlike many problems of regress, this one can have severe consequences when we talk about risk management.

4. NOT ANY FAT TAILS WOULD DO

Although they do not share some aspects of the style of the message, the four discussants appear to agree with *TBS* about the role of outliers and their primacy over the ordinary in determining the statistical properties. The discussants advocate the following: robust statistics, stochastic volatility or GARCH, or Extreme Value Theory. These approaches either do not solve the problem of confidence about small probability, or they create new ones: many of these are tools, not solutions. Robust statistics are certainly more natural tools (Goldstein and Taleb 2007), but I fail to see how robust statistics will produce more information about the probability of events that are not in the sample of the past realizations (see Freedman and Stark 2003). Moreover, there is a major methodological difference between our standpoints: I do not believe in using *any* distribution that naively produces *some* extreme event (or calibrates one from past data). From an operational (and risk management) standpoint, not any fat tails would do.

The central idea of *TBS* concerns the all-too-common logical confusion of *absence of evidence* with *evidence of absence*, associated with the error of confirmation. It tries to avert this logical error in the interpretation of statistical information. As it is impossible to make precise statements about unseen events, those that lie outside the sample set, we need to make the richest possible scenarios about them. For this *TBS* uses, on both logical and empirical grounds, the classification made by Mandelbrot (1963) between two classes of probability distributions: those that have "true fat tails" and others that do not. I had difficulty understanding why the statistical literature has neglected for so long the Mandelbrotian classification.

True fat-tailed distributions have a scale-free or fractal property that I can simplify as follows: for X large enough, (i.e., "in the tails"), $P[X > nx]/P[X > x]$ depends on n, not on x. In financial securities, say, where X is a monthly return, there is no reason for $P[X > 20\%]/P[X > 10\%]$ to be different from $P[X > 15\%]/P[X > 7.5\%]$. This self-similarity at all scales generates power-law, or Paretian, tails; that is, above a crossover point, $P[X > x] = Kx^{-\alpha}$. (Note that the same properties hold for $P[X < x]$ in the negative domain.)

The standard Poisson and stochastic volatility models are not scale-invariant. There is a known value of x beyond which these distributions become thin-tailed—when in reality we do not know what the upper bound is. Further, the Poisson lends itself to in-sample overfitting: you can always use a Poisson jump to fit, in past samples, the largest realization of a fractal fat-tailed process. But it would fail out of sample. For instance, before the 23% drop in the stock market crash of 1987, the worst previous in-sample move was close to 10%. Calibrating a Poisson jump of 10% would not have prepared the risk manager for the ensuing large drop. On the other hand, for someone using the framework of Mandelbrot (1963), the crash of 1987 would not have been surprising—nor would hundreds of large moves we've had in currencies and stocks (*TBS* presents an overview of the literature on dozens of empirical tests across socioeconomic random variables).

Unless there are logical reasons to assume "Mediocristan," or mild randomness, *TBS* advocates using a fractal distribution for the tails *as a default*, which is the opposite of what I've seen practiced. Why? There is a logical asymmetry: a true fat-tailed distribution can camouflage as thin-tailed in small samples; the opposite is not true. If I see a "20-sigma" event, I can be convinced that the data are not Gaussian. If I see no such deviation I cannot make statements that the tails are necessarily thin—in fat-tailed distributions, nothing eventful takes place most of the time. The burden of proof is not on a fat-tailed distribution.

Decision makers are mostly concerned about the cost of mistakes, rather than exact knowledge about the statistical properties. We are dealing with plenty of invisibles, so I do not use power-law tails as a way to estimate precise probabilities—since the parameter α is not easily computed—rather as an aid to make decisions. How?

First, we use power laws as risk-management tools; they allow us to quantify sensitivity to left- and right-tail measurement errors and rank situations based on the *full* effect of the unseen. We can effectively get information about our vulnerability to the tails by varying the power-law exponent α and looking at the effect on the moments or the shortfall (expected losses in excess of some threshold). This is a fully structured stress testing, as the tail exponent α decreases, all possible states of the world are encompassed. And skepticism about the tails can lead to action and allow ranking situations based on the fragility of knowledge; as these errors are less consequential in some areas than others. I explain as follows. If your left tail is "organically" truncated (i.e., the state of the world is not possible or cannot affect you), then you may not worry about negative low-probability events and look forward to positive ones. In a business that benefits from the rare event (bounded left-tail exposure, unbounded right one), rare events that the past did not reveal are almost certainly going to be good for you. When you look at past biotech revenues, for example, you do not see the superblockbuster in them, and owing to the potential for a cure for a disease, there is a small probability that the sales in that industry may turn out to be far larger than what was revealed from past data. This is illuminated by thickening the right tail: varying the α to gauge the effect of the unseen.

On the other hand, consider businesses negatively exposed to rare events (bounded right tails). The track record you see is likely to overestimate the properties—and any thickening of the left tail lowers your expectation. *TBS* discusses the 1982 blowup of banks that lost a century of profits in a single episode: on the eve of the episode, they appeared to the naïve observer to be more profitable than they seemed.

The second reason I advocate the "true fat tails" method of Mandelbrot (1963) in finance and economics is, as I said, empirical. As we saw with the crash of 1987, events have remained consistent with statistics since then—unlike other methods (Poisson or stochastic volatility) that failed us out of sample. But methods allowing for "wild randomness" are not popular in economics and the disciplines that rely on times series analyses because they do away with the measure called "variance," embedded in the consciousness, and so necessary for many applications.

5. CONCLUSION

To conclude, I am exposing the fragility of knowledge about the tails of the distributions in domains where errors can be consequential. I discuss my operational reasons to select scalable laws, that is, "true fat tails" as default distributions and as tools to minimize exposure to such consequential errors. It is only in these cases of lessened tail dependence that statistics are safe—and that is where its strength lies.

Finally I would like to thank the discussants and *The American Statistician* for their open-mindedness and for giving me the opportunity to explain myself. This makes me extremely proud to be an applied statistician.

REFERENCES

Freedman, D. A., and Stark, P. B. (2003), "What is the Probability of an Earthquake?" in *Earthquake Science and Seismic Risk Reduction*, NATO Science Series IV: Earth and Environmental Sciences, vol. 32, eds. F. Mulargia and R. J. Geller, Dordrecht, The Netherlands: Kluwer.

Goldstein, D. G., and Taleb, N. N. (in press), "We Don't Quite Know What We are Talking About When We Talk About Volatility," *Journal of Portfolio Management*.

Mandelbrot, B., (1963), "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394–419.

——— (1997), *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*, New York: Springer-Verlag.

——— (2001), "Scaling in Financial Prices," *Quantitative Finance*, 1, 113–123, 124–130, 427–440, and 641–649.

Taleb, N. N. (1997), *Dynamic Hedging: Managing Vanilla and Exotic Options*, New York: Wiley.

——— (2007), *The Black Swan: The Impact of the Highly Improbable*, New York: Random House and London: Penguin.