Probability, Risk, and Extremes

Nassim Nicholas Taleb
Tandon School of Engineering, New York University
in *Extremes (Darwin College Lectures)*, Cambridge University Press,
Duncan Needham and Julius Weitzdorfer, eds.

This chapter is about classes of statistical distributions that deliver extreme events, and how we should deal with them for both statistical inference and decision making. It draws on the author's multi-volume series, *Incerto* [1] and associated technical research program, which is about how to live in a real world with a structure of uncertainty that is too complicated for us.

The *Incerto* tries to connects five different fields related to tail probabilities and extremes: mathematics, philosophy, social science, contract theory, decision theory, and the real world. (If you wonder why contract theory, the answer is at the end of this discussion: option theory is based on the notion of contingent and probabilistic contracts designed to modify and share classes of exposures in the tails of the distribution; in a way option theory is mathematical contract theory).

The main idea behind the project is that while there is a lot of uncertainty and opacity about the world, and incompleteness of information and understanding, there is little, if any, uncertainty about what actions should be taken based on such incompleteness, in a given situation.

I. On the Difference Between Thin and Fat Tails

We begin with the notion of fat tails and how it relates to extremes using the two imaginary domains of Mediocristan (thin tails) and Extremistan (fat tails). In Mediocristan, no observation can really change the statistical properties. In Extremistan, the tails (the rare events) play a disproportionately large role in determining the properties.

Let us randomly select two people in Mediocristan with a (very unlikely) combined height of 4.1 metres a tail event. According to the Gaussian distribution (or its siblings), the most likely combination of the two heights is 2.05 metres and 2.05 metres.

Simply, the probability of exceeding 3 sigmas is 0.00135. The probability of exceeding 6 sigmas, twice as much, is $9.86 * 10^{-10}$. The probability of two 3-sigma events occurring is $1.8 * 10^{-6}$. Therefore the probability of two 3-sigma events occurring is considerably higher than the probability of one single 6-sigma event. This is using a class of distribution that is not fat-tailed. Figure I below shows that as we extend the ratio from the probability of two 3-sigma events divided by the probability of a 6-sigma event, to the probability of two 4-sigma events divided by the probability of an 8-sigma event, i.e., the further we go into the tail, we see that a large deviation can only occur via a combination (a sum) of a large number of intermediate deviations: the right side of Figure I. In other words, for something bad to happen, it needs to come from a series of very unlikely events, not a single one. This is the logic of Mediocristan.

Let us now move to Extremistan, where a Pareto distribution prevails (among many), and randomly select two people with combined wealth of £36 million. The most likely combination is not £18 million and £18 million. It is approximately £35,999,000 and £1,000. This highlights the crisp distinction between the two domains; for the class of subexponential distributions, ruin is more likely to come from a single extreme event than from a series of bad episodes. This logic underpins classical risk theory as outlined by Lundberg early in the 20^{th} Century[2] and formalized by Cramer[3], but forgotten by economists in recent times. This indicates that insurance can only work in Medocristan; you should never write an uncapped insurance contract if there is a risk of catastrophe. The point is called the catastrophe principle.

As I said earlier, with fat tail distributions, extreme events away from the centre of the distribution play a very large role. Black swans are not more frequent, they are more consequential. The fattest tail distribution has just one very large extreme deviation, rather than many departures form the norm. Figure 3 shows that if we take a distribution like the Gaussian and start fattening it, then the number of departures away from one standard deviation drops. The probability of an event staying within one standard deviation of the mean is 68 per cent. As we the tails fatten, to mimic what happens in financial markets for example, the probability of an event staying within one standard deviation of the mean rises to between 75 and 95 per cent. When we fatten the tails we have higher peaks, smaller shoulders, and higher incidence of very large deviation.

II. A (More Advanced) Categorization and Its Consequences

Let us now provide a taxonomy of fat tails. There are three types of fat tails, as shown in Figure 4, based on mathematical properties. First there are entry level fat tails. This is any distribution with fatter tails than the Gaussian i.e. with more observations within one sigma and with kurtosis (a function of the fourth central moment) higher than three. Second, there are subexponential distributions satisfying our thought experiment

This lecture was given at Darwin College on January 27 2017, as part of the Darwin College Series on Extremes. The author extends the warmest thanks to D.J. Needham [and ... who patiently and accurately transcribed the ideas into a coherent text]. The author is also grateful towards Ole Peters who corrected some mistakes.

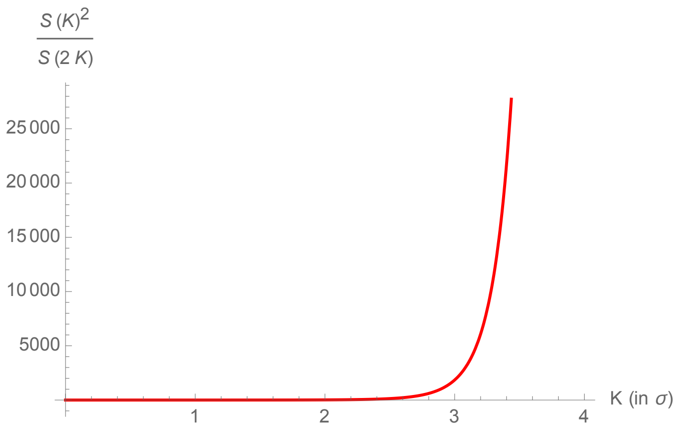

Fig. 1. Ratio of two occurrences of K vs one of $2K$ for a Gaussian distribution. The larger the K, that is, the more we are in the tails, the more likely the event is likely to come from 2 independent realizations of K (hence $P(K)^2$, and the less from a single event of magnitude $2K$
.

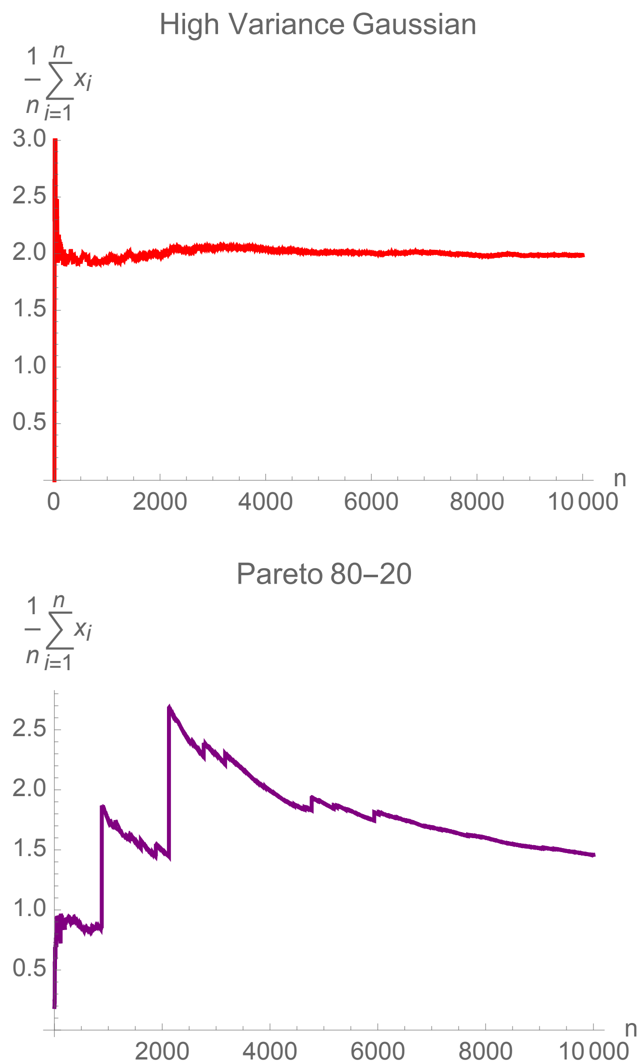

Fig. 2. The law of large numbers, that is how long it takes for the sample mean to stabilize, works much more slowly in Extremistan (here a Pareto distribution with 1.13 tail exponent, corresponding to the "Pareto 80-20"

earlier. Unless they enter the class of power laws, these are not really fat tails because they do not have monstrous impacts from rare events. Level three, what is called by a variety of names, the power law, or slowly varying class, or "Pareto tails" class correspond to real fat tails.

Working from the bottom left of Figure 4, we have the degenerate distribution where there is only one possible outcome i.e. no randomness and no variation. Then, above it, there is the Bernoulli distribution which has two possible outcomes. Then above it there are the two Gaussians. There is the natural Gaussian (with support on minus and plus infinity), and Gaussians that are reached by adding random walks (with compact support, sort of). These are completely different animals since one can deliver infinity and the other cannot (except asymptotically). Then above the Gaussians there is the subexponential class. Its members all have moments, but the subexponential class includes the lognormal, which is one of the strangest things on earth because sometimes it cheats and moves up to the top of the diagram. At low variance, it is thin-tailed, at high variance, it behaves like the very fat tailed.

Membership in the subexponential class satisfies the Cramer condition of possibility of insurance (losses are more likely to come from many events than a single one), as we illustrated in Figure I. More technically it means that the expectation of the exponential of the random variable exists.[1]

Once we leave the yellow zone, where the law of large numbers largely works, then we encounter convergence problems. Here we have what are called power laws, such as Pareto laws. And then there is one called Supercubic, then there is Levy-Stable. From here there is no variance. Further up, there is no mean. Then there is a distribution right at the top, which I call the Fuhgetaboudit. If you see something in that category, you go home and you dont talk about it. In the category before last, below the top (using the parameter α, which indicates the "shape" of the tails, for $\alpha < 2$ but not $\alpha \leq 1$), there is no variance, but there is the mean absolute deviation as indicator of dispersion. And recall the Cramer condition: it applies up to the second Gaussian which means you can do insurance.

The traditional statisticians approach to fat tails has been to assume a different distribution but keep doing business as usual, using same metrics, tests, and statements of significance. But this is not how it really works and they fall into logical inconsistencies. Once we leave the yellow zone, for which statistical techniques were designed, things no longer work as planned. Here are some consequences of moving out of the yellow zone:

1) The law of large numbers, when it works, works too slowly in the real world (this is more shocking than you think as it cancels most statistical estimators). See Figure 2.
2) The mean of the distribution will not correspond to the sample mean. In fact, there is no fat tailed distribution in which the mean can be properly estimated directly from the sample mean, unless we have orders of magnitude more data than we do (people in finance still do not understand this).
3) Standard deviations and variance are not useable. They fail out of sample.

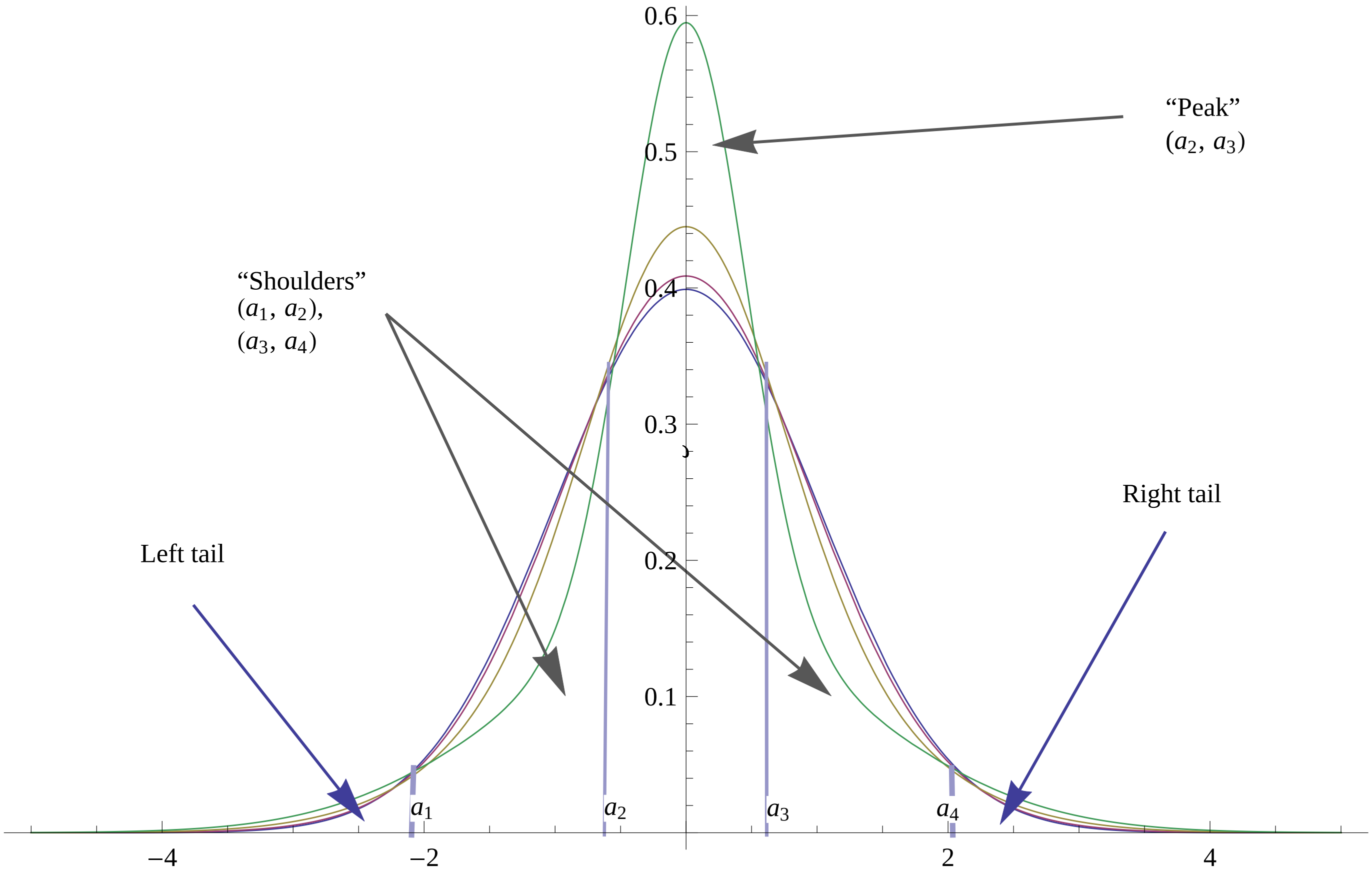

Fig. 3. Where do the tails start? Fatter and fatter fails through perturbation of the scale parameter σ for a Gaussian, made more stochastic (instead of being fixed). Some parts of the probability distribution gain in density, others lose. Intermediate events are less likely, tails events and moderate deviations are more likely. We can spot the crossovers a_1 through a_4. The "tails" proper start at a_4 on the right and $a_1$on the left.

4) Beta, Sharpe Ratio and other common financial metrics are uninformative.
5) Robust statistics is not robust at all.
6) The so-called "empirical distribution" is not empirical (as it misrepresents the expected payoffs in the tails).
7) Linear regression doesn’t work.
8) Maximum likelihood methods work for parameters (good news). We can have plug in estimators in *some* situations.
9) The gap between dis-confirmatory and confirmatory empiricism is wider than in situations covered by common statistics i.e. difference between absence of evidence and evidence of absence becomes larger.
10) Principal component analysis is likely to produce false factors.
11) Methods of moments fail to work. Higher moments are uninformative or do not exist.
12) There is no such thing as a typical large deviation: conditional on having a large move, such move is not defined.
13) The Gini coefficient ceases to be additive. It becomes super-additive. The Gini gives an illusion of large concentrations of wealth. (In other words, inequality in a continent, say Europe, can be higher than the average inequality of its members) [5].

Let us illustrate one of the problem of thin-tailed thinking with a real world example. People quote so-called "empirical" data to tell us we are foolish to worry about ebola when only two Americans died of ebola in 2016. We are told that we should worry more about deaths from diabetes or people tangled in their bedsheets. Let us think about it in terms of tails. But, if we were to read in the newspaper that 2 billion people have died suddenly, it is far more likely that they died of ebola than smoking or diabetes or tangled in their bedsheets? This is rule number one. "Thou shalt not compare a multiplicative fat-tailed process in Extremistan in the subexponential class to a thin-tailed process that has Chernov bounds from Mediocristan". This is simply because of the catastrophe principle we saw earlier, illustrated in Figure I. It is naïve empiricism to compare these processes, to suggest that we worry too much about ebola and too little about diabetes. In fact it is the other way round. We worry too much about diabetes and too little about ebola and other multiplicative effects. This is an error of reasoning that comes from not understanding fat tails –sadly it is more and more common.

Let us now discuss the law of large numbers which is the basis of much of statistics. The law of large numbers tells us that as we add observations the mean becomes more stable, the rate being the square of n. Figure 2 shows that it takes

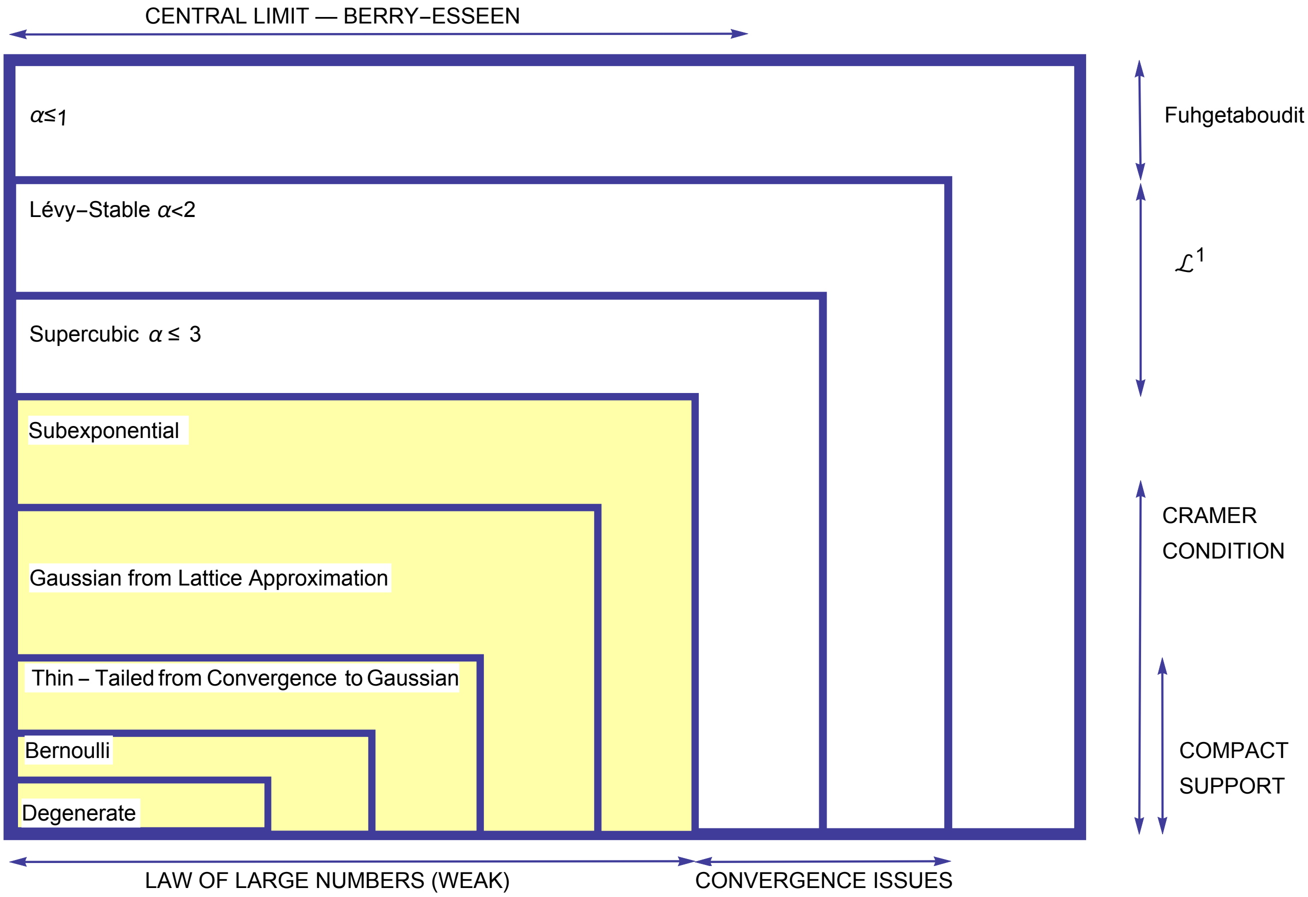

Fig. 4. The tableau of Fat tails, along the various classifications for convergence purposes (i.e., convergence to the law of large numbers, etc.) and gravity of inferential problems. Power Laws are in white, the rest in yellow. See Embrechts et al [4].

many more observations under a fat-tailed distribution (on the right hand side) for the mean to stabilize.

The "equivalence" is not straightforward.

One of the best known statistical phenomena is Paretos 80/20 e.g. twenty per cent of Italians own 80 per cent of the land. Table [**?**] shows that while it takes 30 observations in the Gaussian to stabilize the mean up to a given level, it takes 10^{11} observations in the Pareto to bring the sample error down by the same amount (assuming the mean exists).

Despite this being trivial to compute, few people compute it. You cannot make claims about the stability of the sample mean with a fat tailed distribution. There are other ways to do this, but not from observations on the sample mean.

III. Epistemology and Inferential Asymmetry

Let us now examine the epistemological consequences. Figure 5 illustrates the Masquerade Problem (or Central Asymmetry in Inference). On the left is a degenerate random variable taking seemingly constant values with a histogram producing a Dirac stick.

We have known at least since Sextus Empiricus that we cannot rule out degeneracy but there are situations in which we can rule out non-degeneracy. If I see a distribution that has no randomness, I cannot say it is not random. That is, we cannot say there are no black swans. Let us now add one observation. I can now see it is random, and I can rule out degeneracy. I can say it is not not random. On the right hand side we have seen a black swan, therefore the statement that there are no black swans is wrong. This is the negative empiricism that underpins Western science. As we gather information, we can rule things out. The distribution on the right can hide as the distribution on the left, but the distribution on the right cannot hide as the distribution on the left (check). This gives us a very easy way to deal with randomness. Figure 6 generalizes the problem to how we can eliminate distributions.

If we see a 20 sigma event, we can rule out that the distribution is thin-tailed. If we see no large deviation, we can not rule out that it is not fat tailed unless we understand the process very well. This is how we can rank distributions. If we reconsider Figure 4 we can start seeing deviation and

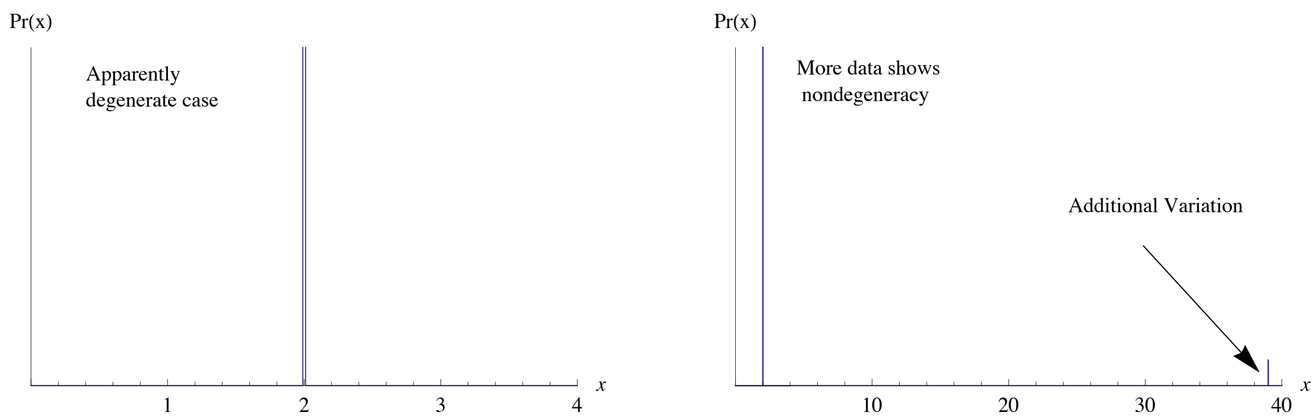

Fig. 5. **The Masquerade Problem (or Central Asymmetry in Inference)**. To the left, a degenerate random variable taking seemingly constant values, with a histogram producing a Dirac stick. One cannot rule out nondegeneracy. But the right plot exhibits more than one realization. Here one can rule out degeneracy. This central asymmetry can be generalized and put some rigor into statements like "failure to reject" as the notion of what is rejected needs to be refined. We produce rules in Chapter **??**.

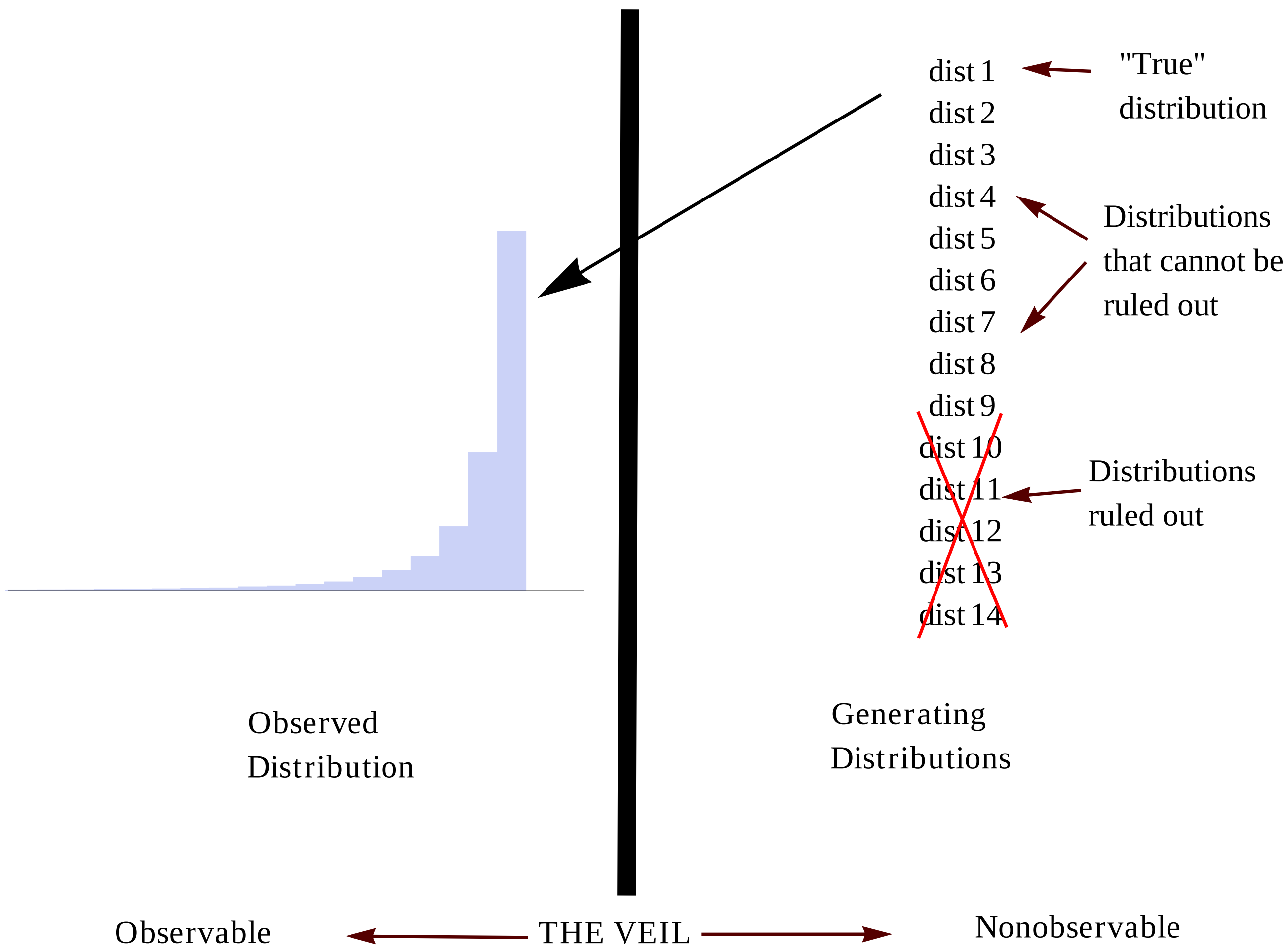

Fig. 6. **"The probabilistic veil"**. Taleb and Pilpel [6] cover the point from an epistemological standpoint with the "veil" thought experiment by which an observer is supplied with data (generated by someone with "perfect statistical information", that is, producing it from a generator of time series). The observer, not knowing the generating process, and basing his information on data *and data only*, would have to come up with an estimate of the statistical properties (probabilities, mean, variance, value-at-risk, etc.). Clearly, the observer having incomplete information about the generator, and no reliable theory about what the data corresponds to, will always make mistakes, but these mistakes have a certain pattern. This is the central problem of risk management.

ruling out progressively from the bottom. These are based on how they can deliver tail events. Ranking distributions becomes very simple because if someone tells you there is a ten-sigma event, it is much more likely that they have the wrong distribution than it is that you really have ten-sigma event. Likewise, as we saw, fat tailed distributions do not deliver a lot of deviation from the mean. But once in a while you get a big deviation. So we can now rule out what is not Mediocristan. We can rule out where we are not we can rule out Mediocristan. I can say this distribution is fat tailed by elimination. But I can not certify that it is thin tailed. This is the black swan problem.

IV. Primer on Power Laws

Let us now discuss the intuition behind the Pareto Law. It is simply defined as: say X is a random variable. For x sufficently large, the probability of exceeding $2x$ divided by the probability of exceeding x is no different from the probability of exceeding $4x$ divided by the probability of exceeding $2x$, and so forth. This property is called "scalability".[2]

So if we have a Pareto (or Pareto-style) distribution, the ratio of people with £16 million compared to £8 million is the same as the ratio of people with £2 million and £1 million. There is a constant inequality. This distribution has no characteristic scale which makes it very easy to understand. Although this distribution often has no mean and no standard deviation we still understand it. But because it has no mean we have to ditch the statistical textbooks and do something more solid, more rigorous.

A Pareto distribution has no higher moments: moments either do not exist or become statistically more and more unstable. So next we move on to a problem with economics and econometrics. In 2009 I took 55 years of data and looked at how much of the kurtosis (a function of the fourth moment) came from the largest observation –see Table III. For a Gaussian the maximum contribution over the same time span should be around .008 $\pm$ .0028. For the S&P 500 it was about

TABLE I
CORRESPONDING n_α, OR HOW MANY FOR EQUIVALENT α-STABLE DISTRIBUTION. THE GAUSSIAN CASE IS THE $\alpha = 2$. FOR THE CASE WITH EQUIVALENT TAILS TO THE 80/20 ONE NEEDS 10^{11} MORE DATA THAN THE GAUSSIAN.

α	n_α Symmetric	$n_\alpha^{\beta=\pm\frac{1}{2}}$ Skewed	$n_\alpha^{\beta=\pm 1}$ One-tailed
1	Fughedaboudit	-	-
$\frac{9}{8}$	6.09×10^{12}	2.8×10^{13}	1.86×10^{14}
$\frac{5}{4}$	574,634	895,952	1.88×10^{6}
$\frac{11}{8}$	5,027	6,002	8,632
$\frac{3}{2}$	567	613	737
$\frac{13}{8}$	165	171	186
$\frac{7}{4}$	75	77	79
$\frac{15}{8}$	44	44	44
2	30.	30	30

TABLE II
AN EXAMPLE OF A POWER LAW

Richer than 1 million	1 in 62.5
Richer than 2 million	1 in 250
Richer than 4 million	1 in 1,000
Richer than 8 million	1 in 4,000
Richer than 16 million	1 in 16,000
Richer than 32 million	1 in ?

80 per cent. This tells us that we dont know anything about kurtosis. Its sample error is huge; or it may not exist so the measurement is heavily sample dependent. If we dont know anything about the fourth moment, we know nothing about the stability of the second moment. It means we are not in a class of distribution that allows us to work with the variance, even if it exists. This is finance.

For silver futures, in 46 years 94 per cent of the kurtosis came from one observation. We cannot use standard statistical methods with financial data. GARCH (a method popular in academia) does not work because we are dealing with squares. The variance of the squares is analogous to the fourth moment. We do not know the variance. But we can work very easily with Pareto distributions. They give us less information, but nevertheless, it is more rigorous if the data are uncapped or if there are any open variables.

Table III, for financial data, debunks all the college textbooks we are currently using. A lot of econometrics that deals with squares goes out of the window. This explains why economists cannot forecast what is going on they are using the wrong methods. It will work within the sample, but it will not work outside the sample. If we say that variance (or kurtosis) is infinite, we are not going to observe anything that is infinite within a sample.

TABLE III
KURTOSIS FROM A SINGLE OBSERVATION FOR FINANCIAL DATA $\frac{\text{MAX}\left(X_{t-\Delta \text{T}i}^4\right)_{i=0}^n}{\sum_{i=0}^n X_{t-\Delta \text{T}i}^4}$

Security	**Max Q**	**Years.**
Silver	0.94	46.
SP500	0.79	56.
CrudeOil	0.79	26.
Short Sterling	0.75	17.
Heating Oil	0.74	31.
Nikkei	0.72	23.
FTSE	0.54	25.
JGB	0.48	24.
Eurodollar Depo 1M	0.31	19.
Sugar	0.3	48.
Yen	0.27	38.
Bovespa	0.27	16.
Eurodollar Depo 3M	0.25	28.
CT	0.25	48.
DAX	0.2	18.

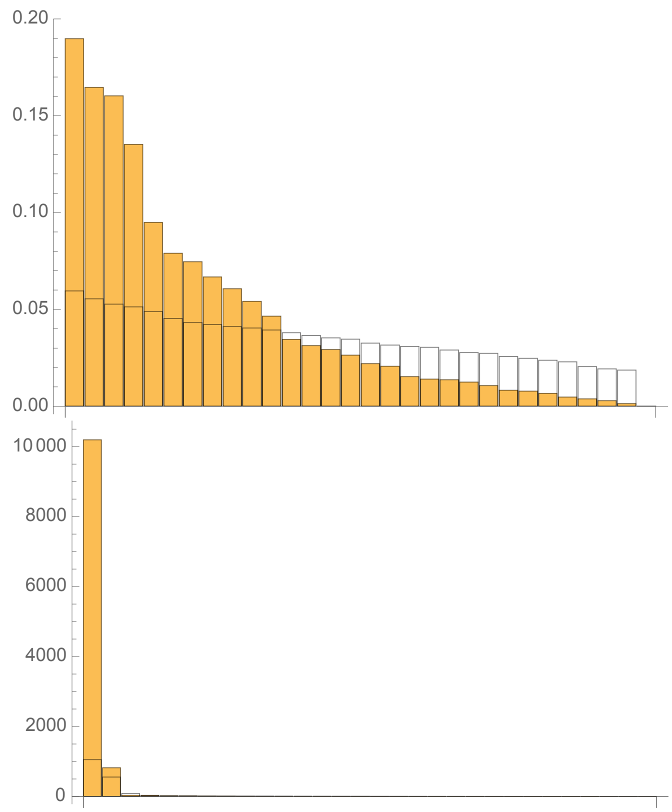

Fig. 7. A Monte Carlo experiment that shows how spurious correlations and covariances are more acute under fat tails. Principal Components ranked by variance for 30 Gaussian uncorrelated variables, n=100 (above) and 1000 data points, and principal Components ranked by variance for 30 Stable Distributed (with tail $\alpha = \frac{3}{2}$, symmetry $\beta = 1$, centrality $\mu = 0$, scale $\sigma = 1$) (below). Both are "uncorrelated" identically distributed variables, n=100 and 1000 data points.

Principal component analysis (Figure 7) is a dimension reduction method for big data and it works beautifully with thin tails. But if there is not enough data there is an illusion of a structure. As we increase the data (the n variables), the structure becomes flat. In the simulation, the data that has absolutely no structure. We have zero correlation on the matrix. For a fat tailed distribution (the lower section), we need a lot more data for the spurious correlation to wash out i.e. dimension reduction does not work with fat tails.

V. Where are the hidden properties?

The following summarizes everything that I wrote in the Black Swan. Distributions can be one-tailed (left or right) or two-tailed. If the distribution has a fat tail it can be fat tailed one tail or it can be fat tailed two tails. And if is fat tailed-one tail, it can be fat tailed left tail or fat tailed right tail.

See Figure 8: if it is fat tailed and we look at the sample mean, we observe fewer tail events. The common mistake is to think that we can naively derive the mean in the presence of one-tailed distributions. But there are unseen rare events and with time these will fill in. But by definition, they are low probability events. The trick is to estimate the distribution and then derive the mean. This is called plug-in estimation, see Table IV. It is not done by observing the sample mean which is biased with fat-tailed distributions. This is why, outside a crisis, the banks seem to make large profits. Then once in a while they lose everything and more and have to be bailed out by the taxpayer. The way we handle this is by differentiating the true mean (which I call "shadow") from the realized mean, as in the Tableau in Table IV.

We can also do that for the Gini coefficient to estimate the "shadow" one rather than the naively observed one.

This is what I mean when I say that the "empirical" distribution is not "empirical".

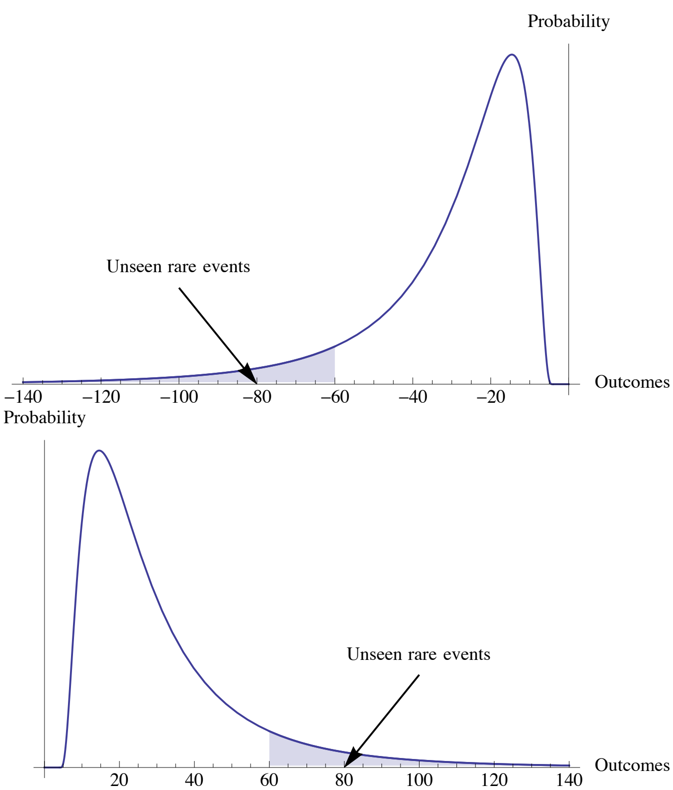

Fig. 8. The biases in perception of risks. Bottom: Inverse Turkey Problem The unseen rare event is positive. When you look at a positively skewed (antifragile) time series and make inferences about the unseen, you miss the good stuff an underestimate the benefits. Top: the opposite problem. The filled area corresponds to what we do not tend to see in small samples, from insufficiency of points. Interestingly the shaded area increases with model error.

TABLE IV
Shadow mean, sample mean and their ratio for different minimum thresholds. In bold the values for the 145k threshold. Rescaled data. From Cirillo and Taleb [7]

Thresh.$\times 10^3$	**Shadow**$\times 10^7$	**Sample**$\times 10^7$	**Ratio**
50	1.9511	1.2753	1.5299
100	2.3709	1.5171	1.5628
145	**3.0735**	**1.7710**	**1.7354**
300	3.6766	2.2639	1.6240
500	4.7659	2.8776	1.6561
600	5.5573	3.2034	1.7348

Once we have figured out the distribution, we can estimate the statistical mean. This works much better than observing the sample mean. For a Pareto distribution, for instance, 98% of observations are below the mean. There is a bias in the mean. But once we know we have a Pareto distribution, we should ignore the sample mean and look elsewhere.

Note that the field of Extreme Value Theory [**?**] [4] [8] focuses on tail properties, not the mean or statistical inference.

VI. Ruin and Path Dependence

Let us finish with path dependence and time probability. Our grandmothers understand fat tails. These are not so scary; we figured out how to survive by making rational decisions based on deep statistical properties.

Path dependence is as follows. If I iron my shirts and then wash them, I get vastly different results compared to when I wash my shirts and then iron them. My first work, *Dynamic Hedging* [9], was about how traders avoid the "absorbing barrier" since once you are bust, you can no longer continue: anything that will eventually go bust will lose *all* past profits.

The physicists Ole Peters and Murray Gell-Mann [10] shed new light on this point, and revolutionized decision theory showing that a key belief since the development of applied probability theory in economics was wrong. They pointed out that all economics textbooks make this mistake; the only exception are by information theorists such as Kelly and Thorp.

Let us explain ensemble probabilities.

Assume that 100 of us, randomly selected, go to a casino and gamble. If the 28^{th} person is ruined, this has no impact on the 29^{th} gambler. So we can compute the casinos return using the law of large numbers by taking the returns of the 100 people who gambled. If we do this two or three times, then we get a good estimate of what the casinos edge is. The problem comes when ensemble probability is applied to us as individuals. It does not work because if one of us goes to the casino and on day 28 is ruined, there is no day 29. This is why Cramer showed insurance could not work outside what he called the Cramer condition, which excludes possible ruin from single shocks. Likewise, no individual investor will achieve the alpha return on the market because no single investor has infinite pockets (or, as Ole Peters has observed, is running his life across branching parallel universes). We can only get the return on the market under strict conditions.

Time probability and ensemble probability are not the same. This only works if the risk takers has an allocation policy

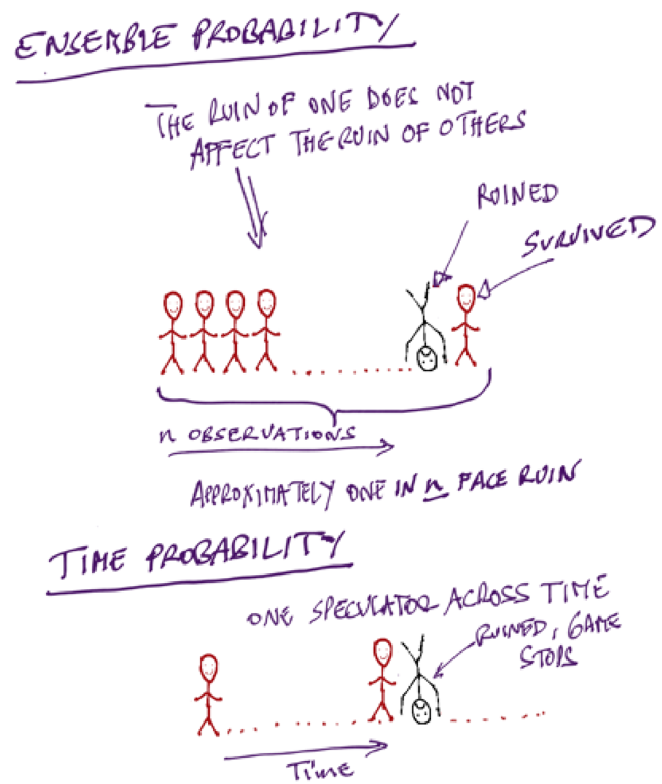

Fig. 9. Ensemble probability vs. time probability. The treatment by option traders is done via the absorbing barrier. I have traditionally treated this in *Dynamic Hedging* and *Antifragile* as the conflation between X (a random variable) and $f(X)$ a function of said r.v., which may include an absorbing state.

compatible with the Kelly criterion[11],[12] using logs. Peters wrote three papers on time probability (one with Gell-Mann) and showed that a lot of paradoxes disappeared.

Let us see how we can work with these and what is wrong with the literature. If we visibly incur a tiny risk of ruin, but have a frequent exposure, it will go to probability one over time. If we ride a motorcycle we have a small risk of ruin, but if we ride that motorcycle a lot then we will reduce our life expectancy. The way to measure this is:

Only focus on the reduction of life expectancy of the unit assuming repeated exposure at a certain density or frequency.

Behavioral finance so far makes conclusion from statics not dynamics, hence misses the picture. It applies trade-offs out of context and develops the consensus that people irrationally overestimate tail risk (hence need to be "nudged" into taking more of these exposures). But the catastrophic event is an absorbing barrier. No risky exposure can be analyzed in isolation: risks accumulate. If we ride a motorcycle, smoke, fly our own propeller plane, and join the mafia, these risks add up to a near-certain premature death. Tail risks are not a renewable resource.

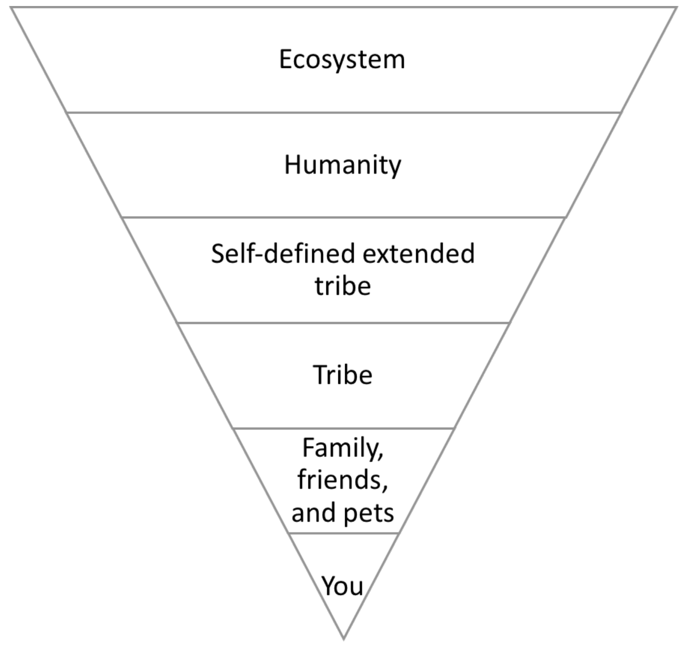

Fig. 10. A hierarchy for survival. Higher entities have a longer life expectancy, hence tail risk matters more for these.

Every risk taker who survived understands this. Warren Buffett understands this. Goldman Sachs understands this. They do not want small risks, they want zero risk because that is the difference between the firm surviving and not surviving over twenty, thirty, one hundred years. This attitude to tail risk can explain that Goldman Sachs is 149 years old –it ran as partnership with unlimited liability for approximately the first 130 years, but was bailed out once in 2009, after it became a bank. This is not in the decision theory literature but we (people with skin in the game) practice it every day. We take a unit, look at how long a life we wish it to have and see by how much the life expectancy is reduced by *repeated* exposure.

The psychological literature focuses on one-single episode exposures and narrowly defined cost-benefit analyses. Some analyses label people as paranoid for overestimating small risks, but don't get that if we had the smallest tolerance for collective tail risks, we would not have made it for the past several million years.

Next let us consider layering, why systemic risks are in a different category from individual, idiosyncratic ones. Look at Figure 10: the worst-case scenario is not that an individual dies. It is worse if your family, friends and pets die. It is worse if you die and your arch enemy survives. They collectively have more life expectancy lost from a terminal tail event.

So there are layers. The biggest risk is that the entire ecosystem dies. The precautionary principle puts structure around the idea of risk for units expected to survive.

Ergodicity in this context means that your analysis for ensemble probability translates into time probability. If it doesn't, ignore ensemble probability altogether.

VII. What to do?

To summarize, we first need to make a distinction between Mediocristan and Extremistan, two separate domains that never overlap with one another. If we dont make that distinction, we dont have any valid analysis. Second, if we dont make the distinction between time probability (path dependent) and ensemble probability (path independent) we dont have a valid analysis.

The next phase of the *Incerto* project is to gain understanding of fragility, robustness, and, eventually, anti-fragility. Once we know something is fat-tailed, we can use heuristics to see how an exposure there reacts to random events: how much is a given unit harmed by them. It is vastly more effective to focus on being insulated from the harm of random events than try to figure them out in the required details (as we saw the inferential errors under fat tails are huge). So it is more solid, much wiser, more ethical, and more effective to focus on detection heuristics and policies rather than fabricate statistical properties.

The beautiful thing we discovered is that everything that is fragile has to present a concave exposure [13] similar –if not identical –to the payoff of a short option, that is, a negative exposure to volatility. It is nonlinear, necessarily. It has to have harm that accelerates with intensity, up to the point of breaking. If I jump 10m I am harmed more than 10 times than if I jump one metre. That is a necessary property of fragility. We just need to look at acceleration in the tails. We have built effective stress testing heuristics based on such an option-like property [14].

In the real world we want simple things that work [15]; we want to impress our accountant and not our peers. (My argument in the latest instalment of the *Incerto*, *Skin in the Game* is that systems judged by peers and not evolution rot from overcomplication). To survive we need to have clear techniques that map to our procedural intuitions.

The new focus is on how to detect and measure convexity and concavity. This is much, much simpler than probability.

Notes

[1]Let X be a random variable. The Cramer condition: for all $r > 0$,

$$\mathbb{E}\left(e^{rX}\right) < +\infty.$$

[2]More formally: let X be a random variable belonging to the class of distributions with a "power law" right tail:

$$\mathbb{P}(X > x) \sim L(x)\, x^{-\alpha} \tag{1}$$

where $L : [x_{\min}, +\infty) \to (0, +\infty)$ is a slowly varying function, defined as $\lim_{x\to+\infty} \frac{L(kx)}{L(x)} = 1$ for any $k > 0$. We can apply the same to the negative domain.

References

[1] N. N. Taleb, *Incerto: Antifragile, the Black Swan, Fooled by Randomness, the Bed of Procrustes, Skin in the Game*. Random House and Penguin, 2001-2018.

[2] F. Lundberg, *I. Approximerad framställning af sannolikhetsfunktionen. II. Återförsäkring af kollektivrisker. Akademisk afhandling... af Filip Lundberg,...* Almqvist och Wiksells boktryckeri, 1903.

[3] H. Cramér, *On the mathematical theory of risk*. Centraltryckeriet, 1930.

[4] P. Embrechts, *Modelling extremal events: for insurance and finance*. Springer, 1997, vol. 33.

[5] A. Fontanari, N. N. Taleb, and P. Cirillo, "Gini estimation under infinite variance," *Physica A: Statistical Mechanics and its Applications*, vol. 502, no. 256-269, 2018.

[6] N. N. Taleb and A. Pilpel, "I problemi epistemologici del risk management," *Daniele Pace (a cura di)" Economia del rischio. Antologia di scritti su rischio e decisione economica", Giuffrè, Milano*, 2004.

[7] P. Cirillo and N. N. Taleb, "On the statistical properties and tail risk of violent conflicts," *Physica A: Statistical Mechanics and its Applications*, vol. 452, pp. 29–45, 2016.

[8] L. Haan and A. Ferreira, "Extreme value theory: An introduction," *Springer Series in Operations Research and Financial Engineering (*, 2006.

[9] N. N. Taleb, *Dynamic Hedging: Managing Vanilla and Exotic Options*. John Wiley & Sons (Wiley Series in Financial Engineering), 1997.

[10] O. Peters and M. Gell-Mann, "Evaluating gambles using dynamics," *Chaos*, vol. 26, no. 2, 2016.

[11] J. L. Kelly, "A new interpretation of information rate," *Information Theory, IRE Transactions on*, vol. 2, no. 3, pp. 185–189, 1956.

[12] E. O. Thorp, "Optimal gambling systems for favorable games," *Revue de l'Institut International de Statistique*, pp. 273–293, 1969.

[13] N. N. Taleb and R. Douady, "Mathematical definition, mapping, and detection of (anti) fragility," *Quantitative Finance*, 2013.

[14] N. N. Taleb, E. Canetti, T. Kinda, E. Loukoianova, and C. Schmieder, "A new heuristic measure of fragility and tail risks: application to stress testing," *International Monetary Fund*, 2018.

[15] G. Gigerenzer and P. M. Todd, *Simple heuristics that make us smart*. New York: Oxford University Press, 1999.