On the Biases in the Estimation of Inequality Using Bracketed Quantile Contributions

Nassim Nicholas Taleb*, Raphael Douady†
*Retired option trader, NYU Engineering
†Risk Data, Centre d'Economie de la Sorbonne

***Abstract*—In fat-tailed domains, sample measures of top centile contributions to wealth are biased, unstable estimators extremely sensitive to sample size and concave in taking into account large deviations. They tend to vary over time merely from the increase of sample space, thus providing the illusion of structural changes in inequality. They are also inconsistent under aggregation. In addition, it can be shown that under fat tails, increases in wealth need to be accompanied with increased measurement of inequality. We compute the bias and standard error and show some small sample and asymptotic properties of the estimator and propose alternative extrapolative measurements of inequality and fat-tailedness.**

I. Introduction

Vilfredo Pareto noticed that 80% of the land in Italy belonged to 20% of the population, and vice-versa, thus both giving birth to the power law class of distributions and the popular saying 80/20. The self-similarity at the core of the property of power laws [1] and [2] allows us to recurse and reapply the 80/20 to the remaining 20%, and so forth until one obtains the result that the top percent of the population will own about 53% of the total wealth.

It looks like such measure of inequality can be seriously biased, depending on how it is measured, so it is very likely that the true ratio of inequality, that is, the share of the top percentile was closer to 70%, hences changes year-on-year would drift higher to converge to such level from larger sample. In fact, as we will show in this discussion, more complete samples resulting from technological progress, and also from population and economic growth will make such a measure converge by increasing over time, for no other reason than expansion in sample size.

The core of the problem is that, for the class one-tailed fat-tailed random variables, that is, bounded on the left and unbounded on the right, where the random variable $X \in [x_{min}, \infty)$, the in-sample quantile contribution (to income or wealth) is a biased estimator of the true value of the actual quantile contribution. Let us define the *quantile contribution*

$$\kappa_q = q \frac{\mathbb{E}[X|X > h(q)]}{\mathbb{E}[X]}$$

where $h(q) = \inf\{h \in [x_{min}, +\infty) \,, \mathbb{P}(X > h) \leq q\}$ is the exceedance threshold for the probability q.

For a given sample $(X_k)_{1 \leq k \leq n}$, its "natural" estimator $\widehat{\kappa}_q \equiv \frac{q^{th}\text{percentile}}{total}$, used in most academic studies, can be expressed, as

$$\widehat{\kappa}_q \equiv \frac{\sum_{i=1}^{n} \mathbb{1}_{X_i > \hat{h}(q)} X_i}{\sum_{i=1}^{n} X_i}$$

where $\hat{h}(q)$ is the estimated exceedance threshold for the probability q, i.e., $\hat{h}(q) = \inf\{h : \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{x > h} \leq q\}$. We shall see that the observed variable $\widehat{\kappa}_q$ is a downward biased estimator of the true ratio κ_q, the one that would hold out of sample, and such bias is in proportion to the fatness of tails and, for very fat tailed distributions, remains significant, even for very large samples.

II. Estimation

Let X be a random variable belonging to the class of distributions with a "power law" right tail, that is:

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

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 assume independent observations.

There is little difference for small exceedance quantiles (<50%) between the various possible distributions such as Dagum,[3],[4] Singh-Maddala distribution [5], or straight Pareto.

For exponents $1 \leq \alpha \leq 2$, as observed in [6], the law of large numbers operates, though *extremely* slowly. The problem is acute for α around (but above) 1 and severe as no convergence is possible at 1.

A. True value of estimator

Let us first consider $\phi_\alpha(x)$ the density of a α-Pareto distribution bounded from below by $x_{\min} > 0$, in other words: $\phi_\alpha(x) = \alpha x_{\min}^\alpha x^{-\alpha-1} \mathbb{1}_{x \geq x_{\min}}$, and $\mathbb{P}(X > x) = \left(\frac{x_{\min}}{x}\right)^\alpha$. Under these assumptions, the cutpoint of exceedance is $h(q) = x_{\min}\, q^{-1/\alpha}$ and we have:

$$\kappa_q = \frac{\int_{h(q)}^{\infty} x\, \phi(x) dx}{\int_{x_{min}}^{\infty} x\, \phi(x) dx} = \left(\frac{h(q)}{x_{\min}}\right)^{1-\alpha} = q^{\frac{\alpha-1}{\alpha}}$$

If the distribution of X is α-Pareto only beyond a cut-point x_{cut}, which we assume to be below $h(q)$, so that we have $\mathbb{P}(X > x) = \left(\frac{\lambda}{x}\right)^\alpha$ for some $\lambda > 0$, then we still have $h(q) = \lambda q^{-1/\alpha}$ and

$$\kappa_q = \frac{\alpha}{\alpha - 1} \frac{\lambda}{\mathbb{E}\left[X\right]} q^{\frac{\alpha-1}{\alpha}}$$

The estimation of κ_q hence requires that of the exponent α as well as that of the scaling parameter λ, or at least its ratio to the expectation of X.

In the general case, let us fix the threshold h and define:

$$\kappa_{q,h} = q\frac{\mathbb{E}[X|X>h]}{\mathbb{E}[X]}$$

so that we have $\kappa_q = \kappa_{q,h(q)}$. We also define the n-sample estimator:

$$\widehat{\kappa}_{q,h} \equiv \frac{\sum_{i=1}^{n} \mathbb{1}_{X_i>h} X_i}{\sum_{i=1}^{n} X_i}$$

where X_i are n independent copies of X.

B. Unusal Concavity-Driven Bias of the Estimator

Remark 1. *Let $X = (X)_{i=1}^n$ a random sample of size $n > \frac{1}{q}$, Y an extra single random observation, and define: $\kappa_{q,h}^{X\sqcup Y} = \frac{\sum_{i=1}^{n} \mathbb{1}_{X_i>h} X_i + \mathbb{1}_{Y>h} Y}{\sum_{i=1}^{n} X_i + Y}$. We observe that, whenever $Y > h$, one has:*

$$\frac{\partial^2 \kappa_{q,h}^{X\sqcup Y}}{\partial Y^2} \leq 0.$$

Note that, this inequality is still valid with κ_q as the value $h^{X\sqcup Y}(q)$ doesn't depend on the particular value of $Y > h^X(q)$.

Unlike the common small sample effect resulting from high impact from the rare observation in the tails less likely to show up in small samples, a bias which goes away by repetition of sample runs, we face a different situation. The concavity of the estimator constitutes a upper bound for the measurement in finite n, clipping large deviations, which leads to problems of aggregation as we will state below in Theorem 1. This remark is probably one of the reasons why, as observed in [7], $\widehat{\kappa}_q$ is a non-converging estimator of κ_q: only the number of exceedences of a well chosen sequence of thresholds is able to produce a converging estimator of the two key parameters to estimate κ_q, namely the exponent α and the scale λ. In practice, even in very large sample, the contribution of very large rare events to κ_q prevents the sample estimator to converge to the true value. One needs to use a different path, first estimating the distribution parameters (α, λ) and only then, estimating the tail contribution κ_q. Falk observes that, even with a proper estimator, the convergence is extremely slow, of the order of $n^{-\delta}/\ln n$, where the exponent δ depends on α and on the tolerance of the actual distribution vs. a theoretical Pareto, measured by the Hellinger distance. In particular, $\delta \to 0$ as $\alpha \to 1$, making the convergence really slow for low values of α. The next section shows how much the bias still remains for very large samples.

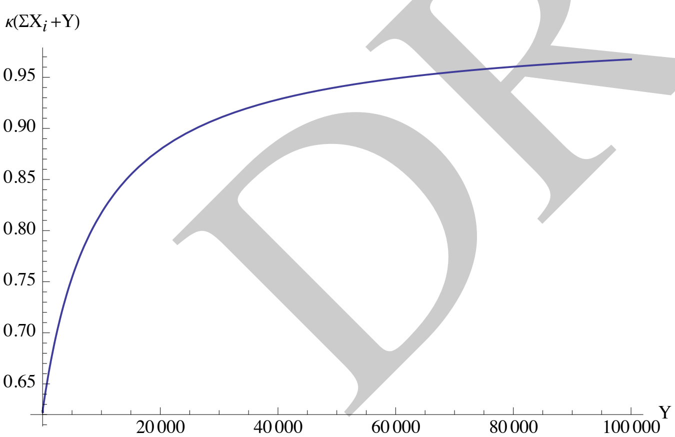

Figure 1. Effect of additional observations on κ

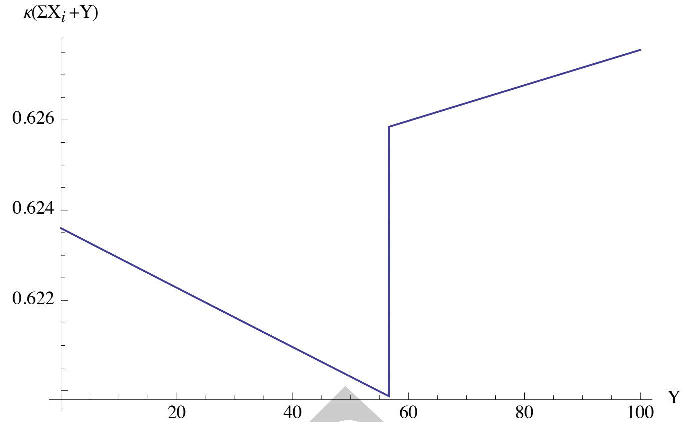

Figure 2. Effect of additional observations on κ, we can see convexity on both sides of h except for values of no effect to the left of h, an area of order $1/n$

C. Bias and Convergence

The table below shows the bias of $\widehat{\kappa}_q$ as an estimator of κ_q in the case of an α-Pareto distribution with $\alpha = 1.1$, a value chosen to be compatible with practical economic measures, such as the wealth distribution in the world or in a particular country, including developped ones.[1] In such a case, the estimator is extemely sensitive to *"small"* samples, *"small"* meaning in practice 10^8. We ran a trillion simulations across varieties of sample sizes. While $\kappa_{.01} \approx 0.657933$, even a sample size of 100 million remains severely biased as seen in Table I.

Naturally the bias is rapidly (and nonlinearly) reduced for α further away from 1, and becomes weak in the neighborhood of 2. It is also weaker outside the top 1% centile, hence this discussion focuses on the famed "one percent" and on the low α exponent.

III. An Inequality About Aggregating Inequality

For the estimation of the mean of a fat-tailed r.v. $(X)_i^j$, in m sub-samples of size n' each for a total of $n = mn'$, the allocation of the total number of observations n between i and j does not matter so long as the total n is unchanged. Here the allocation of n samples between m sub-samples does matter because of the concavity of κ.[2] Next we prove

[1]This value, which is lower than the estimated exponents one can find in the literature – around 2 [?] – is, still according to [?], a lower estimate which cannot be excluded from the observations.

[2]The same concavity – and general bias – applies when the distribution is lognormal.

Table I
BIASES OF ESTIMATOR OF $\kappa = 0.657933$ FROM 10^{12} MONTE CARLO REALIZATIONS

$\widehat{\kappa}(n)$	Mean	Median	STD across MC runs
$\widehat{\kappa}(10^3)$	0.405235	0.367698	0.160244
$\widehat{\kappa}(10^4)$	0.485916	0.458449	0.117917
$\widehat{\kappa}(10^5)$	0.539028	0.516415	0.0931362
$\widehat{\kappa}(10^6)$	0.581384	0.555997	0.0853593
$\widehat{\kappa}(10^7)$	0.591506	0.575262	0.0601528
$\widehat{\kappa}(10^8)$	0.626525	0.606915	TK

that global inequality as measured by $\widehat{\kappa}_q$ on a broad set of data will appear higher than local inequality, so aggregating European data, for instance, would give a $\widehat{\kappa}_q$ higher than the average measure of inequality across countries – an *"inequality about inequality"*. In other words, we claim that the estimation bias when using $\widehat{\kappa}_q(n)$ is even increased when dividing the sample into sub-samples and taking the weighted average of the measured values $\widehat{\kappa}_q(n_i)$.

Theorem 1. *Partition the n data into m sub-samples $N = N_1 \cup \ldots \cup N_m$ of respective sizes $n_1, \ldots, n_m$, with $\sum_{i=1}^m n_i = n$, and assume that the distribution of variables X_j is the same in all the sub-samples. Then we have:*

$$\mathbb{E}\left[\widehat{\kappa}_q(N)\right] \geq \sum_{i=1}^{m} \frac{n_i}{n} \mathbb{E}\left[\widehat{\kappa}_q(N_i)\right]$$

Proof. Denote by $S = \sum_{j=1}^{n} X_j$ and, for each sub-sample, $S_i = \sum_{j \in N_i} X_j$, so that we have $S = \sum_{i=1}^{m} S_i$. Also denote $h = h^N(q)$ for the full sample and, for each sub-sample, $h_i = h^{N_i}(q)$. Then we have:

$$\widehat{\kappa}_{q,h}(N) = \sum_{i=1}^{m} \frac{S_i}{S} \widehat{\kappa}_{q,h}(N_i)$$

We first observe that, in each sample:

$$\mathbb{E}\left[\widehat{\kappa}_{q,h}(N_i)\right] \geq \mathbb{E}\left[\widehat{\kappa}_{q,h_i}(N_i)\right]$$

This is simply due to the fact that the larger the ratio $\widehat{\kappa}_{q,h}(N_i)$, the higher h_i is likely to be, hence reducing the contribution of the "tail" of N_i beyond h_i to the total sum S_i.

The second observation is that, for a given threshold h, the random variables R_i and $\widehat{\kappa}_{q,h}(N_i)$ are positively correlated, indeed, conditionally to a given upward bias of R_i with respect to its expectation, due to fat tails, the bias is more likely due to a large contribution of the tail beyond h than to an accumulation of values above average but below h. Hence we have:

$$\mathbb{E}\left[\frac{S_i}{S}\widehat{\kappa}_{q,h}(N_i)\right] \geq \mathbb{E}\left[\frac{S_i}{S}\right]\mathbb{E}\left[\widehat{\kappa}_{q,h}(N_i)\right]$$

Finally, it is easy to show that, with fat-tailed distributions, we have $\mathbb{E}\left[S_i/S\right] \geq n_i/n$, again due to the fact that, conditionnaly to a given total sum S, large sample values in a sub-sample N_i have a higher impact on the ratio R_i than small values, inducing a convexity effect leading to this inequality.

Putting these three inequalities together yields the result. ☐

IV. MORE WEALTH IMPLIES INCREASE IN $\hat{\kappa}$

INCOMPLETE SECTION (Where we show that since the distribution is fat tailed, the maximum is of the same order as the sum, hence additional wealth means more measured inequality. It is quite absurd to assume that wealth is coming from the bottom or even the middle.)

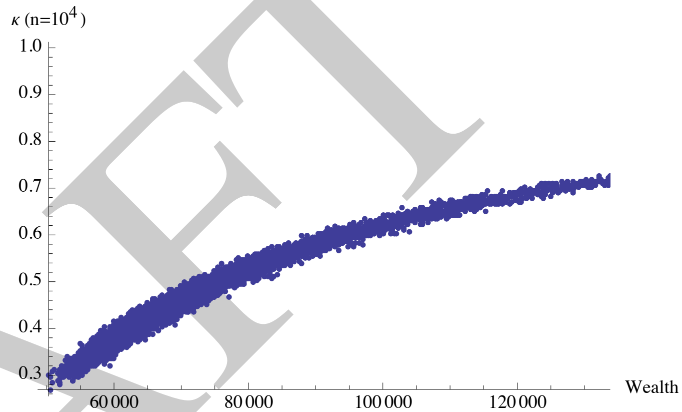

Figure 3. Effect of additional wealth on $\hat{\kappa}$

V. CONCLUSION AND PROPER ESTIMATION OF INEQUALITY

Inequality can be severe at the level of the generator, but in small n we observe a lower κ. So we get a historical illusion of rise in wealth inequality when it has been there all along.[3]

Even the estimation of α can be biased in some domains where one does not see the entire picture: in the presence of uncertainty about the "true" α, it can be shown that, unlike other parameters, the one to use is not the probability-weighted exponents (the standard average) but the minimum across a section of exponents [6].

It did not escape our attention that some theories are built based on claims of such "increase" in inequality, as in [9], without taking into account the true nature of κ, and promulgating theories about the "variation" of inequality without reference to the stochasticity of the estimation $-$and the lack of consistency of κ. We also faced the argument that "the sudy is based on a complete set of data" and estimating in sample $\hat{\kappa}$ from "robust" methods works. Robust methods, alas, tend to fail with fat-tailed data.

Take an insurance (or, better, reinsurance) company. The "accounting" profits in a year in which there were few claims do not reflect on the "economic" status of the company. The

[3]Accumulated wealth is typically thicker tailed than income,[8].

"accounting" profits are not used to predict variations year-on-year, rather the exposure to tail (and other) events, analyses that take into account the stochastic nature of the performance. This difference between "accounting" (deterministic) and "economic" (stochastic) values matters for policy making, particularly under fat tails.

How Should We Measure Inequality?

Risk managers now tend to compute CVaR and other metrics, methods that are extrapolative and nonconcave, such as the information from the α exponent (taking the one from the lower bound of the range of exponents) and rederiving the corresponding κ, are less biased and do not get mixed up with problems of aggregation.[4] By extrapolative, we mean the built-in extension of the tail in the measurement by taking into account realizations outside the sample path that are in excess of the extrema observed.[5]

Acknowledgment

The late Benoît Mandelbrot, Dominique Guéguan, Felix Salmon, Bruno Dupire, the staff at Luciano Restaurant in Brooklyn and Naya in Manhattan.

References

[1] B. Mandelbrot, "The pareto-levy law and the distribution of income," *International Economic Review*, vol. 1, no. 2, pp. 79–106, 1960.
[2] ——, "The stable paretian income distribution when the apparent exponent is near two," *International Economic Review*, vol. 4, no. 1, pp. 111–115, 1963.
[3] C. Dagum, "Inequality measures between income distributions with applications," *Econometrica*, vol. 48, no. 7, pp. 1791–1803, 1980.
[4] ——, *Income distribution models*. Wiley Online Library, 1983.
[5] S. Singh and G. Maddala, "A function for size distribution of incomes: reply," *Econometrica*, vol. 46, no. 2, 1978.
[6] N. N. Taleb, "Silent risk: Lectures on fat tails,(anti) fragility, and asymmetric exposures," *Available at SSRN 2392310*, 2014.
[7] M. Falk *et al.*, "On testing the extreme value index via the pot-method," *The Annals of Statistics*, vol. 23, no. 6, pp. 2013–2035, 1995.
[8] X. Gabaix, "Power laws in economics and finance," National Bureau of Economic Research, Tech. Rep., 2008.
[9] T. Piketty, "Capital in the 21st century," 2014.
[10] T. Piketty and E. Saez, "The evolution of top incomes: a historical and international perspective," National Bureau of Economic Research, Tech. Rep., 2006.

[4] Some authors such as [10] use Pareto interpolation for unsufficient information about the tails (based on tail parameter), filling-in the bracket with conditional average bracket contribution, which is not the same thing as using full power-law extension, hence retains the bias.

[5] Even using a lognormal distribution, by fitting the scale parameter, works to some extent as a rise of the standard deviation extrapolates probability mass into the right tail.

THIS IS FROM N N TALEB'S "SILENT RISK" AS APPENDIX SHOWING PROBLEMS IN ESTIMATION OF TAIL ALPHAS UNDER PARAMETER UNCERTAINTY

4 | Effects of Higher Orders of Uncertainty

***Chapter Summary* 4: The Spectrum Between Uncertainty and Risk. There has been a bit of discussions about the distinction between "uncertainty" and "risk". We believe in gradation of uncertainty at the level of the probability distribution itself (a "meta" or higher order of uncertainty.) One end of the spectrum, "Knightian risk", is not available for us mortals in the real world. We show how the effect on fat tails and on the calibration of tail exponents and reveal inconsistencies in models such as Markowitz or those used for intertemporal discounting (as many violations of "rationality" aren't violations .**

4.1 Meta-Probability Distribution

When one assumes knowledge of a probability distribution, but has uncertainty attending the parameters, or when one has no knowledge of which probability distribution to consider, the situation is called "uncertainty in the Knightian sense" by decision theorisrs(Knight, 1923). "Risk" is when the probabilities are computable without an error rate. Such an animal does not exist in the real world. The entire distinction is a lunacy, since no parameter should be rationally computed witout an error rate. We find it preferable to talk about degrees of uncertainty about risk/uncertainty, using metadistribution, or metaprobability.

The Effect of Estimation Error, General Case

The idea of model error from missed uncertainty attending the parameters (another layer of randomness) is as follows.

Most estimations in social science, economics (and elsewhere) take, as input, an average or expected parameter,

$$\bar{\alpha} = \int \alpha \; \phi(\alpha) \; d\alpha, \tag{4.1}$$

where α is ϕ distributed (deemed to be so a priori or from past samples), and regardless of the dispersion of α, build a probability distribution for x that relies on the mean estimated parameter, $p(X = x) = p\left(x \middle| \bar{\alpha}\right)$, rather than the more appropriate metaprobability adjusted probability for the density:

$$p(x) = \int \phi(\alpha) \, \mathrm{d}\alpha \tag{4.2}$$

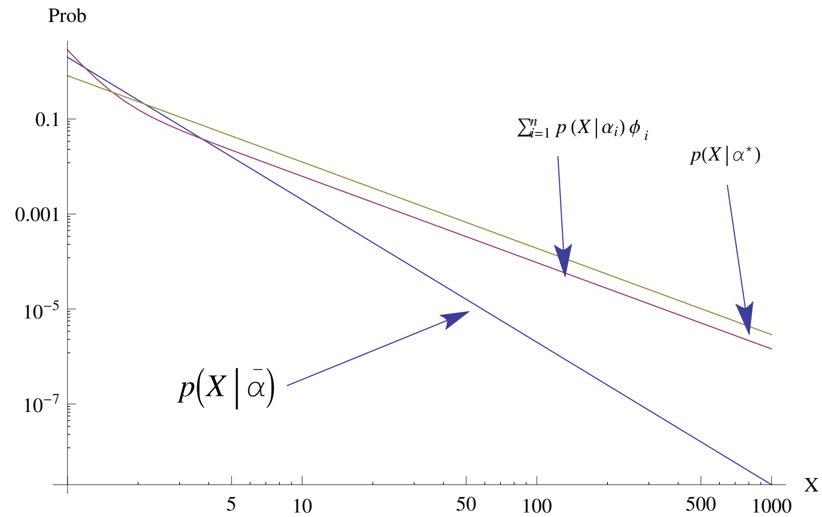

Figure 4.1: Log-log plot illustration of the asymptotic tail exponent with two states.

In other words, if one is not certain about a parameter α, there is an inescapable layer of stochasticity; such stochasticity raises the expected (metaprobability-adjusted) probability if it is $< \frac{1}{2}$ and lowers it otherwise. The uncertainty is fundamentally epistemic, includes incertitude, in the sense of lack of certainty about the parameter.

The model bias becomes an equivalent of the Jensen gap (the difference between the two sides of Jensen's inequality), typically positive since probability is convex away from the center of the distribution. We get the bias ω_A from the differences in the steps in integration

$$\omega_A = \int \phi(\alpha)\, p(x|\alpha)\, \mathrm{d}\alpha - p\left(x|\int \alpha\phi(\alpha)\, \mathrm{d}\alpha\right)$$

With $f(x)$ a function , $f(x) = x$ for the mean, etc., we get the higher order bias $\omega_{A'}$

$$\omega_{A'} = \int \left(\int \phi(\alpha)\ f(x)\ p(x|\alpha)\ \ d\alpha\right) \mathrm{d}x - \int f(x)\ p\left(x|\int \alpha\ \phi(\alpha)\, \mathrm{d}\alpha\right) \mathrm{d}x \quad (4.3)$$

Now assume the distribution of α as discrete n states, with $\alpha = (\alpha_i)_{i=1}^n$ each with associated probability $\phi = \phi_i$ _i=1^n, $\sum_{i=1}^n \phi_i = 1$. Then 4.2 becomes

$$p(x) = \phi_i \left(\sum_{i=1}^{n} p\left(x\,|\alpha_i\,\right)\right) \quad (4.4)$$

So far this holds for α any parameter of any distribution.

4.2 Metadistribution and the Calibration of Power Laws

Remark 1. *In the presence of a layer of metadistributions (from uncertainty about the parameters), the asymptotic tail exponent for a powerlaw corresponds to the lowest possible tail exponent regardless of its probability.*

This explains "Black Swan" effects, i.e., why measurements tend to chronically underestimate tail contributions, rather than merely deliver imprecise but unbiased estimates.

When the perturbation affects the standard deviation of a Gaussian or similar non-powerlaw tailed distribution, the end product is the weighted average of the probabilities. However, a powerlaw distribution with errors about the possible tail exponent will bear the asymptotic properties of the *lowest* exponent, not the average exponent.

Now assume p(X=x) a standard Pareto Distribution with α the tail exponent being estimated, $p(x|\alpha) = \alpha x^{-\alpha-1} x_{\min}^{\alpha}$, where $x_{\min}$is the lower bound for x,

$$p(x) = \sum_{i=1}^{n} \alpha_i x^{-\alpha_i - 1} x_{\min}^{\alpha_i} \phi_i$$

Taking it to the limit

$$\underset{x\to\infty}{\text{limit}}\; x^{\alpha^*+1} \sum_{i=1}^{n} \alpha_i x^{-\alpha_i - 1} x_{\min}^{\alpha_i} \phi_i = K$$

where K is a strictly positive constant and $\alpha^* = \underset{1\leq i\leq n}{min}\, \alpha_i$. In other words $\sum_{i=1}^{n} \alpha_i x^{-\alpha_i - 1} x_{\min}^{\alpha_i} \phi_i$ is asymptotically equivalent to a constant times x^{α^*+1}. The lowest parameter in the space of all possibilities becomes the dominant parameter for the tail exponent.

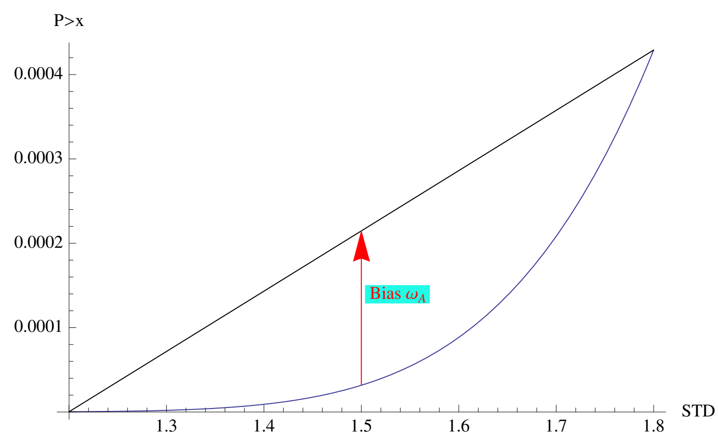

Figure 4.2: Illustration of the convexity bias for a Gaussian from raising small probabilities: The plot shows the STD effect on P>x, and compares P>6 with a STD of 1.5 compared to P> 6 assuming a linear combination of 1.2 and 1.8 (here a(1)=1/5).

Figure 4.1 shows the different situations: a) $p(x|\bar{\alpha})$, b) $\sum_{i=1}^{n} p\left(x\,|\alpha_i\,\right) \phi_i$ and c) $p\left(x\,|\alpha^*\,\right)$. We can see how the last two converge. The asymptotic Jensen Gap ω_A becomes $p(x|\alpha^*) - p(x|\bar{\alpha})$.

Implications

Whenever we estimate the tail exponent from samples, we are likely to underestimate the thickness of the tails, an observation made about Monte Carlo generated α-stable variates and the estimated results (the “Weron effect”)[74].

The higher the estimation variance, the lower the true exponent.

The asymptotic exponent is the lowest possible one. It does not even require estimation.

Metaprobabilistically, if one isn't sure about the probability distribution, and there is a probability that the variable is unbounded and "could be" powerlaw distributed, then it is powerlaw distributed, and of the lowest exponent.

The obvious conclusion is to in the presence of powerlaw tails, focus on changing payoffs to clip tail exposures to limit $\omega_{A'}$ and "robustify" tail exposures, making the computation problem go away.

4.3 The Effect of Metaprobability on Fat Tails

Recall that the tail fattening methods in 2.4 and 2.6.These are based on randomizing the variance. Small probabilities rise precisely because they are convex to perturbations of the parameters (the scale) of the probability distribution.

4.4 Fukushima, Or How Errors Compound

"Risk management failed on several levels at Fukushima Daiichi. Both TEPCO and its captured regulator bear responsibility. First, highly tailored geophysical models predicted an infinitesimal chance of the region suffering an earthquake as powerful as the Tohoku quake. This model uses historical seismic data to estimate the local frequency of earthquakes of various magnitudes; none of the quakes in the data was bigger than magnitude 8.0. Second, the plant's risk analysis did not consider the type of cascading, systemic failures that precipitated the meltdown. TEPCO never conceived of a situation in which the reactors shut down in response to an earthquake, and a tsunami topped the seawall, and the cooling pools inside the reactor buildings were overstuffed with spent fuel rods, and the main control room became too radioactive for workers to survive, and damage to local infrastructure delayed reinforcement, and hydrogen explosions breached the reactors' outer containment structures. Instead, TEPCO and its regulators addressed each of these risks independently and judged the plant safe to operate as is."Nick Werle, n+1, published by the n+1 Foundation, Brooklyn NY

4.5 The Markowitz inconsistency

Assume that someone tells you that the probability of an event is exactly zero. You ask him where he got this from. "Baal told me" is the answer. In such case, the person is coherent, but would be deemed unrealistic by non-Baalists. But if on the other hand, the person tells you "I estimated it to be zero," we have a problem. The person is both unrealistic and inconsistent. Something estimated needs to have an estimation error. So probability cannot be zero if it is estimated, its lower bound is linked to the estimation error; the higher the estimation error, the higher the probability, up to a point. As with Laplace's argument of total ignorance, an infinite estimation error pushes the probability toward $\frac{1}{2}$. We will return to the implication of the mistake; take for now that anything estimating a parameter and then putting it into an equation is different from estimating the equation across parameters. And Markowitz was inconsistent by starting his "seminal" paper with "Assume you know E and V" (that is, the expectation and the variance). At the end of the paper he accepts that they need to be estimated, and what is worse, with a combination of statistical techniques and the "judgment of practical men." Well, if these parameters need to be estimated, with an error, then the derivations