The Law of Large Numbers Under Fat Tails

Nassim Nicholas Taleb
Tandon School of Engineering, New York University and Real World Risk Institute, LLC.

***Abstract*—To get the same error in measuring the mean for what is commonly called a "Pareto 80/20" as one gets for a Gaussian distributed variable of the same scale, one needs in excess of 10^{13} more data. Convergence according the law of large numbers is considerably slower in these cases.**

The implication is that the mean for fat tailed variables, that is with exponent $1 < \alpha \leq 2$ with asymptotic Lévy-stable distribution cannot be possibly obtained from sample averages in cases when α is close to 1. However maximum likelihood methods work well since α is asymptotically normal and, for a large class of power law distributed variables, the "shadow mean" (or true mean) can be derived with a considerably small error rate.

This note establishes the sample equivalence between the two classes, examines the effect of model uncertainty concerning α and gives a hint of the errors in the statistical literature.

We also get exlicit and semi-explicit expressions for partial means above (below) a certain level K for a Lévy-stable variable as well as the explicit distribution of the mean obtained indirectly via Hill estimators.

I. Introduction

You observe data and get some confidence that the average is represented by the sample thanks to a standard metrified "n". Now what if the data were fat tailed? How much more do you need? What if the model were uncertain –we had uncertainty about the parameters or the probability distribution itself? Let us call "sample equivalence" the sample size that is needed to correspond to a Gaussian sample size of n.

It appears that 1) the statistical literature has been silent on the subject of sample equivalence –since the sample mean is not a good estimator under fat tailed distributions, 2) errors in the estimation of the mean can be several order of magnitudes higher than under corresponding thin tails, 3) many operators writing "scientific" papers aren't aware of it (which includes many statisticians), 4) model error compounds the issue.

We show that fitting tail exponents via ML methods have a small error in delivering the mean.

Main Technical Results In addition to the qualitative discussions about commonly made errors in violating the sample equivalence, the technical contribution is as follows:

- explicit extractions of partial expectations for alpha stable distributions
- the expression of how uncertainty about parameters (quantified in terms of parameter volatility) translates into a larger (or smaller) required n. In other words, the effect of model uncertainty, how the degree of model uncertainty worsens inference, in a quantifiable way.

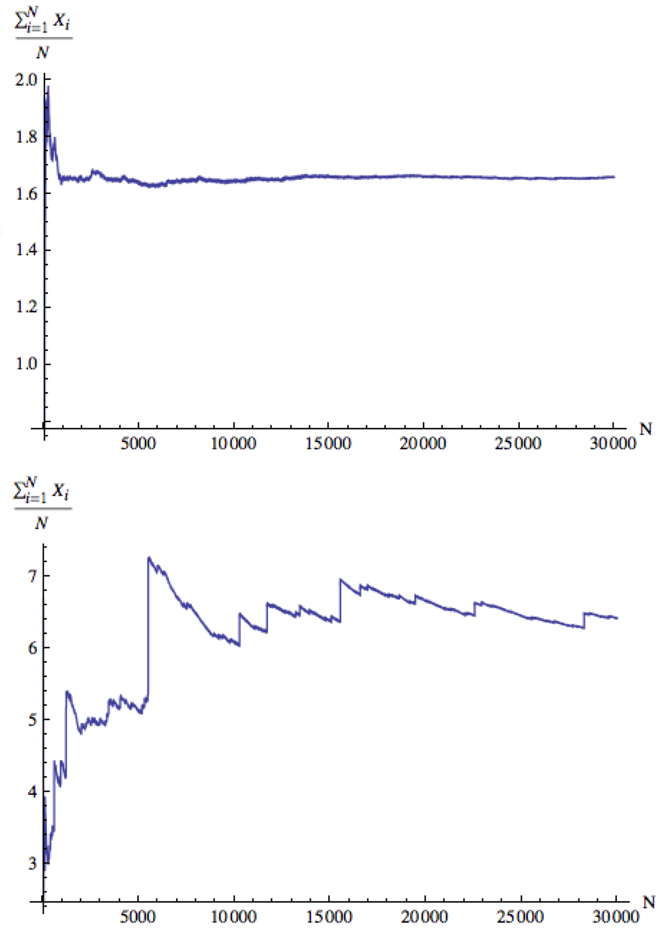

Fig. 1: How thin tails (Gaussian) and fat tails (1< α ≤2) converge to the mean.

II. Introduction

The first discussion examines the issue of "sample equivalence" without any model uncertainty.

A. Background of problem

Let us summarize the standard convergence theorem. By the weak law of large numbers, a sum of random variables $X_1, \ldots, X_n$ with finite mean m, that is $\mathbb{E}(X) < \infty$, then $\frac{1}{n}\sum_{1 \leq i \leq n} X_i$ converges to m in probability, as n $\rightarrow \infty$. Or, for any $\epsilon > 0$ $\lim_{n\to\infty} \mathbb{P}\big(|\overline{X}_n - m| > \epsilon\big) = 0$. In other words: the sample mean will end up converging to the true mean, should the latter exist.

But the result holds at infinity, while we live with finite n. There are several regimes of convergene.

- **Case 1a** when the variance and all other moments exist, and the data is i.i.d., there are two convergence effects at play, one, convergence to the Gaussian (by central limit), the second, the l.l.n., which accelerates the convergence. Some subcategories with higher kurtosis than the Gaussian, such as regime switching situations, or distributions entailing Poisson jumps or similar large deviations with small probability converge more slowly but these are special cases that we can ignore in this discussion since Case 2 is vastly more consequential in effect (it requires an extremely high kurtosis to slow down the central limit).
- **Case 1b** when the variance exists, but higher moments don't, the central limit theorem doesn't really work in practice (it is too slow for "real time") and the law of large numbers works more slowly than Case 1a, but works nevertheless. We consider this as "intermediate" case, more particularly with finite-variance power laws, those with the tail exponent ≥ 2 (or, more accurately, if the distribution is two-tailed, the lower of the left or right tail exponent equal to or exceeding 2).
- **Case 2** when the mean exists, but the variance doesn't, the law of large numbers converges very, very slowly when the exponent is closer to 1.

It is Case 2 that is the main object of this paper. More particularly cases where the lowest tail exponent $1 < \alpha \leq 2$. Of particular relevance is "80/20" where the $\alpha \approx 1.16$.

B. Discussion of the result about sample equivalence for fat tails

We assume that Case 1a converge quickly to a Gaussian, hence approach the "Gaussian basin" which is the special case of stable distributions.

Table I shows the equivalence of number of summands between processes.

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

The "equivalence" is not straightforward.

Exposition of the problem

Let $X_{\alpha,1}, X_{\alpha,2}, \ldots, X_{\alpha,n_\alpha}$ be a sequence of i.i.d. powerlaw distributed variables with tail exponent $1 < \alpha \leq 2$ in at least one of the tails, that is, belonging to the class of distributions with at least one "power law" tail, that is:

$$\mathbb{P}(|X_\alpha| > |x|) \sim L(x)\,|x|^{-\alpha} \quad (1)$$

where $L : [x_0, \pm\infty) \to (0, \pm\infty)$ is a slowly varying function, defined as $\lim_{x\to\pm\infty} \frac{L(kx)}{L(x)} = 1$ for any $k > 0$.

Let $X_{g,1}, X_{g,2}, \ldots, X_{g,n}$ be a sequence of Gaussian variables with mean μ and scale σ. We are looking for values of n' corresponding to a given n_g:

$$n_{\min} = \inf\left\{ n_\alpha : \mathbb{E}\left(\left|\sum_{i=1}^{n_\alpha} \frac{X_{\alpha,i} - m_p}{n_\alpha}\right|\right) \leq \mathbb{E}\left(\left|\sum_{i=1}^{n_g} \frac{X_{g,i} - m_g}{n_g}\right|\right), n_\alpha > 0 \right\} \quad (2)$$

Instability of Mean Deviation and use of L^1 norm

And since we know that convergence for the Gaussian happens at speed $n_g^{\frac{1}{2}}$ (something we will redo using stable distributions), we can compare to convergence of other classes.

> The idea is to limit convergence to L^1 norm; we know clearly that there is no point using the L^2 norm, and even when (as in finite variance power laws, there is some convergence in L^2 (central limit), we ignore such situation for its difficulties in real time. As to the distribution of the maximum, that is, L^∞, fughedoubadit.

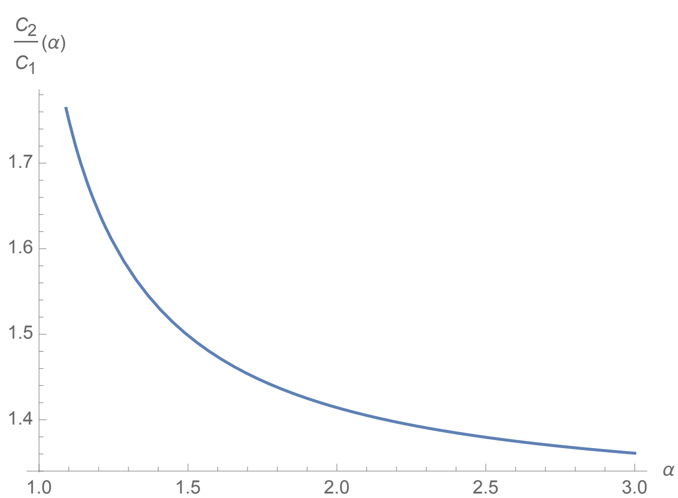

Fig. 2: The ratio of cumulants $\frac{C_2}{C_1}$ for a symmetric powerlaw, as a function of the tail exponent α.

We are expressing in Equation 2 the expected error (that is, a risk function) in $\mathbf{L}^1$ as mean absolute deviation from the observed average, to accommodate absence of variance –but assuming of course existence of first moment without which there is no point discussing averages.

Typically, in statistical inference, one uses standard deviations of the observations to establish the sufficiency of n. But in fat tailed data standard deviations do not exist, or, worse, when they exist, as in powerlaw with tail exponent > 3, they are extremely unstable, particularly in cases where kurtosis is infinite.

Using mean deviations of the samples (when these exist) doesn't accommodate the fact that fat tailed data hide properties. The "volatility of volatility", or the dispersion around the mean deviation increases nonlinearly as the tails get fatter. For instance, a stable distribution with tail exponent at $\frac{3}{2}$ matched to exactly the same mean deviation as the Gaussian will deliver measurements of mean deviation 1.4 times as unstable as the Gaussian.

Using mean absolute deviation for "volatility", and its mean deviation "volatility of volatility" expressed in the $\mathbf{L}^1$ norm, or C_1 and C_2 cumulant:

$$C_1 = \|.\|_1 = \mathbb{E}(|X-m|)$$

$$C_2 = \|(\|.\|_1)\|_1 = \mathbb{E}\left(|X - \mathbb{E}(|X-m|)|\right)$$

We can compare that matching mean deviations does not go very far matching cumulants.(see Appendix 1)

Further, a sum of Gaussian variables will have its extreme values distributed as a Gumbel while a sum of fat tailed will follow a Fréchet distribution *regardless of the the number of summands*. The difference is not trivial, as shown in figures , as in 10^6 realizations for an average with 100 summands, we can be expected observe maxima $> 4000\times$ the average while for a Gaussian we can hardly encounter more than $> 5\times$.

III. Generalizing Mean Deviation as Partial Expectation

It is unfortunate that even if one matches mean deviations, the dispersion of the distributions of the mean deviations (and their skewness) would be such that a "tail" would remain markedly different in spite of a number of summands that allows the matching of the first order cumulant $\|.\|_1$. So we can match the special part of the distribution, the expectation $> K$ or $< K$, where K can be any arbitrary level.

Let $\Psi(t)$ be the characteristic function of the random variable. Let θ be the Heaviside theta function. Since $\text{sgn}(x) = 2\theta(x) - 1$

$$\Psi^{\theta,K}(t) = \int_{-\infty}^{\infty} e^{itx}\left(2\theta(x-K)-1\right)\,\mathrm{d}x = \frac{2ie^{iKt}}{t}$$

And define the expectation of $|X-K|$ as

$$\mathbb{E}_{|K|} := \int_K^{\infty}(x-K)\,\mathrm{d}F(x) - \int_{-\infty}^{K}(x-K)\,\mathrm{d}F(x).$$

The special expectation becomes, by convoluting the Fourier transforms; where F is the distribution function for x:

$$\mathbb{E}_{|K|} = -i\frac{\partial}{\partial t}\int_{-\infty}^{\infty}\Psi(t-u)\Psi^{\theta,K}(u)\mathrm{d}u|_{t=0} \quad (3)$$

We note that when the distribution is symmetric and $\mathbb{E}_{|K|}$ becomes the mean average deviation. Otherwise, we can pull more information as follows. Writing the mean $\mu = \mu_{K_+} + \mu_{K_-}$ for the portions above and below K, and $\mathbb{E}_{|K|} = \mu_{K_+} - \mu_{K_-}$ we can find the downside expected shortfall and upside partial expectation (unconditional) , μ_{K_+} and μ_{K_-} respectively, i.e., $\mu_{K_+} = \frac{1}{2}(\mu + \mathbb{E}_{|K|})$.

Our method also allows the computation of a conditional tail or "CVar" in the language of finance and insurance, with $\mu_{K_-} = \mathbb{E}(X|_{X>K})\mathbb{P}(X>K)$ or just $\mathbb{E}(X|_{X>K})$.

Note a similar approach using the Hilbert Transform for the absolute value of a Lévy stable r.v., see Hlusel, [1], Pinelis [2].

Mean deviation (under a symmetric distribution with mean μ, i.e. $\mathbb{P}(X>\mu) = \frac{1}{2}$) becomes a special case of equation 3, $\mathbb{E}(|X-\mu|) = \left(\int_{\mu}^{\infty}(x-\mu)\,\mathrm{d}F(x) - \int_{-\infty}^{\mu}(x-\mu)\,\mathrm{d}F(x)\right) = \mathbb{E}_{\mu}^{+}$.

IV. Class of Stable Distributions

Assume alpha-stable the class $\mathfrak{S}$ of probability distribution that is closed under convolution: $\mathbf{S}(\alpha,\beta,\mu,\sigma)$ represents the stable distribution with tail index $\alpha \in (0,2]$, symmetry parameter $\beta \in [0,1]$, location parameter $\mu \in \mathbb{R}$, and scale parameter $\sigma \in \mathbb{R}^+$. The Generalized Central Limit Theorem gives sequences a_n and b_n such that the distribution of the shifted and rescaled sum $Z_n = \left(\sum_i^n X_i - a_n\right)/b_n$ of n i.i.d. random variates X_i the distribution function of which $F_X(x)$ has asymptotes $1 - cx^{-\alpha}$ as $x \to +\infty$ and $d(-x)^{-\alpha}$ as $x \to -\infty$ weakly converges to the stable distribution

$$S(\wedge_{\alpha,2}, \mathbb{1}_{0<\alpha<2}\frac{c-d}{c+d}, 0, 1).$$

We note that the characteristic functions are real for all symmetric distributions. [We also note that the convergence is not clear across papers [3] but this doesn't apply to symmetric distributions.]

Note that the tail exponent α used in non stable cases is somewhat, but not fully, different for $\alpha = 2$, the Gaussian case where it ceases to be a powerlaw –the main difference is in the asymptotic interpretation. But for convention we retain the same symbol as it corresponds to tail exponent but use it differently in more general non-stable power law contexts.

The characteristic function $\Psi(t)$ of a variable X^α with scale σ will be, using the expression for $\alpha > 1$, See Zolotarev [4], Samorodnitsky and Taqqu [5]:

$$\Psi_\alpha = \exp\left(i\mu t - |t\sigma|^\alpha\left(1 - i\beta\tan\left(\frac{\pi\alpha}{2}\right)\text{sgn}(t)\right)\right)$$

which, for an n-summed variable (the equivalent of mixing with equal weights), becomes:

$$\Psi_\alpha(t) = \exp\left(i\mu nt - \left|n^{\frac{1}{\alpha}}t\sigma\right|^\alpha\left(1 - i\beta\tan\left(\frac{\pi\alpha}{2}\right)\text{sgn}(t)\right)\right)$$

A. Results

Let $X^\alpha \in \mathfrak{S}$, be the centered variable with a mean of zero, $X^\alpha = (Y^\alpha - \mu)$. We write $\mathbb{E}_{|K|}(\alpha,\beta,\mu,\sigma,K) :=$

$\mathbb{E}(X^\alpha|_{X^\alpha>K}\mathbb{P}(X^\alpha > K))$ under the stable distribution above. From Equation 3:

$$\mathbb{E}_{|K|}(\alpha,\beta,\mu,\sigma,K) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\alpha\sigma^\alpha |u|^{\alpha-2}\left(1+i\beta\tan\left(\frac{\pi\alpha}{2}\right)\operatorname{sgn}(u)\right) \exp\left(|u\sigma|^\alpha\left(-1-i\beta\tan\left(\frac{\pi\alpha}{2}\right)\operatorname{sgn}(u)\right)+iKu\right)du \tag{4}$$

with explicit solution for $K=\mu=0$:

$$\mathbb{E}_{|K|}(\alpha,\beta,0,\sigma,0) = -\sigma\frac{1}{\pi\alpha}\Gamma\left(-\frac{1}{\alpha}\right)\left(\left(1+i\beta\tan\left(\frac{\pi\alpha}{2}\right)\right)^{1/\alpha}+\left(1-i\beta\tan\left(\frac{\pi\alpha}{2}\right)\right)^{1/\alpha}\right). \tag{5}$$

and semi-explicit generalized form for $K\neq\mu$:

$$\mathbb{E}_{|K|}(\alpha,\beta,\mu,\sigma,K) = \sigma\frac{\Gamma\left(\frac{\alpha-1}{\alpha}\right)\left(\left(1+i\beta\tan\left(\frac{\pi\alpha}{2}\right)\right)^{1/\alpha}+\left(1-i\beta\tan\left(\frac{\pi\alpha}{2}\right)\right)^{1/\alpha}\right)}{2\pi} + \sum_{k=1}^{\infty}\frac{i^k(K-\mu)^k\Gamma\left(\frac{k+\alpha-1}{\alpha}\right)\left(\beta^2\tan^2\left(\frac{\pi\alpha}{2}\right)+1\right)^{\frac{1-k}{\alpha}}}{2\pi\sigma^{k-1}k!} \left((-1)^k\left(1+i\beta\tan\left(\frac{\pi\alpha}{2}\right)\right)^{\frac{k-1}{\alpha}}+\left(1-i\beta\tan\left(\frac{\pi\alpha}{2}\right)\right)^{\frac{k-1}{\alpha}}\right) \tag{6}$$

Our formulation in Equation 6 generalizes and simplifies the commonly used one from Wolfe [6] from which Hardin [7] got the explicit form, promoted in Samorodnitsky and Taqqu [5] and Zolotarev [4]:

$$\mathbb{E}(|X|) = \frac{1}{\pi}\sigma\left(2\Gamma\left(1-\frac{1}{\alpha}\right)\left(\beta^2\tan^2\left(\frac{\pi\alpha}{2}\right)+1\right)^{\frac{1}{2\alpha}} \cos\left(\frac{\tan^{-1}\left(\beta\tan\left(\frac{\pi\alpha}{2}\right)\right)}{\alpha}\right)\right) \tag{7}$$

Which allows us to prove the following statements:

1) Relative convergence: The general case with $\beta\neq 0$: for so and so, assuming so and so, (precisions) etc.,

$$n_\alpha^\beta = 2^{\frac{\alpha}{1-\alpha}}\pi^{\frac{\alpha}{2-2\alpha}}\left(\Gamma\left(\frac{\alpha-1}{\alpha}\right)\sqrt{n_g}\left(\left(1-i\beta\tan\left(\frac{\pi\alpha}{2}\right)\right)^{\frac{1}{\alpha}}+\left(1+i\beta\tan\left(\frac{\pi\alpha}{2}\right)\right)^{\frac{1}{\alpha}}\right)\right)^{\frac{\alpha}{\alpha-1}} \tag{8}$$

with alternative expression:

$$n_\alpha^\beta = \pi^{\frac{\alpha}{2-2\alpha}}\left(\frac{\sec^2\left(\frac{\pi\alpha}{2}\right)^{-\frac{1}{2}/\alpha}\sec\left(\frac{\tan^{-1}\left(\tan\left(\frac{\pi\alpha}{2}\right)\right)}{\alpha}\right)}{\sqrt{n_g}\,\Gamma\left(\frac{\alpha-1}{\alpha}\right)}\right)^{\frac{\alpha}{1-\alpha}} \tag{9}$$

Which in the symmetric case $\beta=0$ reduces to:

$$n_\alpha = \pi^{\frac{\alpha}{2(1-\alpha)}}\left(\frac{1}{\sqrt{n_g}\,\Gamma\left(\frac{\alpha-1}{\alpha}\right)}\right)^{\frac{\alpha}{1-\alpha}} \tag{10}$$

2) Speed of convergence: $\forall k\in\mathbb{N}^+$ and $\alpha\in(1,2]$

$$\mathbb{E}\left(\left|\sum_i^{kn_\alpha}\frac{X_i^\alpha-m_\alpha}{kn_\alpha}\right|\right)/\mathbb{E}\left(\left|\sum_i^{n_\alpha}\frac{X_i^\alpha-m_\alpha}{n_\alpha}\right|\right) = k^{\frac{1}{\alpha}-1} \tag{11}$$

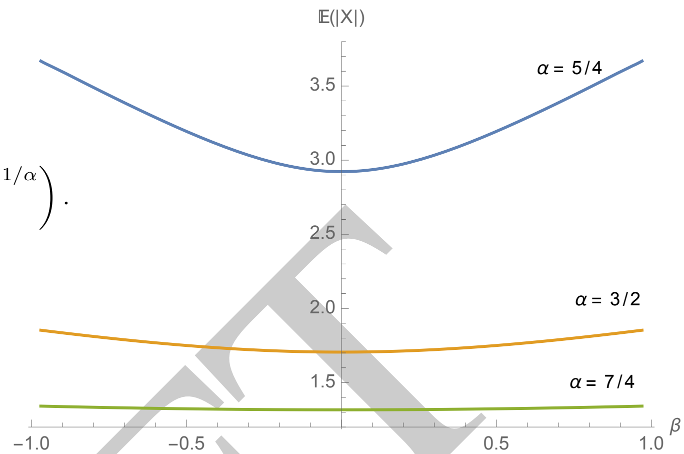

Fig. 3: Asymmetries and Mean Deviation.

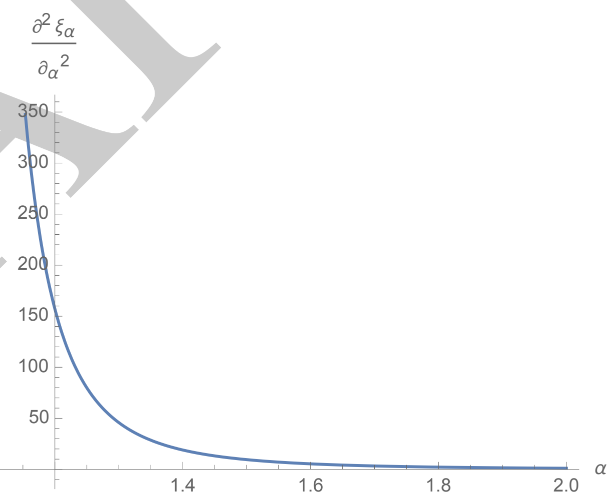

Fig. 4: Mixing distributions: the effect is pronounced at lower values of α, as tail uncertainty creates more fat-tailedness.

Remark 1. *The ratio mean deviation of distributions in $\mathfrak{S}$ is homogeneous of degree $k^{\frac{1}{\cdot}\alpha-1}$. This is not the case for other classes "nonstable".*

Proof. (Sketch) From the characteristic function of the stable distribution. Other distributions need to converge to the basin $\mathfrak{S}$. □

B. Stochastic Alpha or Mixed Samples

Define mixed population X_α and $\xi(X_\alpha)$ as the mean deviation of ...

Proposition 1. *For so and so*

$$\xi(X_{\bar{\alpha}}) \geq \sum_{i=1}^{m}\omega_i\xi(X_{\alpha_i})$$

where $\bar{\alpha} = \sum_{i=1}^{m} \omega_i \alpha_i$ *and* $\sum_{i=1}^{m} \omega_i = 1$.

Proof. A sketch for now: $\forall \alpha \in (1,2)$, where γ is the Euler-Mascheroni constant ≈ 0.5772, $\psi^{(1)}$ the first derivative of the Poly Gamma function $\psi(x) = \Gamma'[x]/\Gamma[x]$, and H_n the n^{th} harmonic number:

$$\frac{\partial^2 \xi}{\partial \alpha^2} = \frac{2\sigma\Gamma}{\pi\alpha^4}\left(\frac{\alpha-1}{\alpha}\right) n^{\frac{1}{\alpha}-1}\left(\psi^{(1)}\left(\frac{\alpha-1}{\alpha}\right) + \left(-H_{-\frac{1}{\alpha}} + \log(n) + \gamma\right)\left(2\alpha - H_{-\frac{1}{\alpha}} + \log(n) + \gamma\right)\right)$$

which is positive for values in the specified range, keeping $\alpha < 2$ as it would no longer converge to the Stable basin. □

Which is also negative with respect to α as can be seen in Figure 4. The implication is that one's sample underestimates the required "n". (Commentary).

V. Alternative Methods for Mean

The sample average of a Pareto distributed variable is not an adequate estimator. We saw that there are two ways to get the mean:

- The observed mean from data,
- For a one-tailed distribution, the observed scale, location and shape α from data, with corresponding distribution of the mean.

We will compare both –in fact there is a very large difference between the properties of both estimators. This exercise selects the simplest Pareto distribution, one in which the scale is the minimum value, here simplified to 1.

Let X be distributed according to Pareto distribution with PDF:

$$\mathbb{P}(x) \triangleq \alpha x^{-\alpha-1}, \, x \in [1, \infty) \tag{12}$$

with mean $z = \frac{\alpha}{\alpha-1}$, $\alpha > 1$.

Let the random variable z "derived mean" be $z = \frac{n-1}{n}\frac{\hat{\alpha}}{\hat{\alpha}-1}$, where $\hat{\alpha}$ is obtained via the maximum likelihood estimator $\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \log(x_i)}$, selecting only values of $\hat{\alpha} \geq 1+\epsilon$

Theorem 1. *The moments of order* m *of the sample "indirect" mean of a random variable distributed according to Eq. 12, obtained via maximum likelihood measurement of* α *for selecting solely observations mapping to values of* $\alpha > 1+\epsilon$ *are:*

$$\mathbb{E}(Z^m) = e^{\alpha(1-n)}(\alpha(n-1))^{n-1}\sum_{i=0}^{n-1}(-1)^i\binom{n-1}{i}\left(\frac{1}{\alpha-\alpha n}\right)^{i-m} \frac{(\Gamma(i-m+1,\alpha-n\alpha)-\Gamma(i-m+1,-(n-1)\alpha\epsilon))}{\Gamma(n)-\Gamma\left(n,\frac{(n-1)\alpha}{\epsilon+1}\right)} \tag{13}$$

Proof. First let us derive the standard maximum likelihood estimator for α. The likelihood function is $\mathcal{L} = \prod_{i=1}^{n} \alpha x_i^{-\alpha-1}$. Maximizing the Log of the likelihood function (assuming we set the minimum value) $\log(\mathcal{L}) = n(\log(\alpha) + \alpha\log(1)) - (\alpha+1)\sum_{i=1}^{n}\log(x_i)$ yields: $\hat{\alpha} = \frac{n}{\sum_{i=1}^{n}\log(x_i)}$. Now consider $l = -\frac{\sum_{i=1}^{n}\log X_i}{n}$. Using the characteristic function to get the distribution of the average logarithm yield:

$$\psi(t)^n = \left(\int_1^\infty f(x)\exp\left(\frac{it\log(x)}{n}\right)dx\right)^n = \left(\frac{\alpha n}{\alpha n - it}\right)^n$$

which is the characteristic function of the gamma distribution $(n, \frac{1}{\alpha n})$. A standard result is that $\hat{\alpha}' \triangleq \frac{1}{l}$ will follow the inverse gamma distribution with density:

$$\phi_{\hat{\alpha}}(a) = \frac{e^{-\frac{\alpha n}{\hat{\alpha}}}\left(\frac{\alpha n}{\hat{\alpha}}\right)^n}{\hat{\alpha}\Gamma(n)}, \, a > 0$$

.

1) Debiasing: Since $\mathbb{E}(\hat{\alpha}) = \frac{n}{n-1}\alpha$ we elect another –unbiased– random variable $\hat{\alpha'} = \frac{n-1}{n}\hat{\alpha}$ which, after scaling, will have for distribution $\phi_{\hat{\alpha'}}(a) = \frac{e^{\frac{\alpha-\alpha n}{a}}\left(\frac{\alpha(n-1)}{a}\right)^{n+1}}{\alpha\Gamma(n+1)}$.

2) Truncating for $\alpha > 1$*:* Given that values of $\alpha \leq 1$ lead to infinite mean (hence no mean) we restrict the distribution to values greater than $1+\epsilon$, $\epsilon > 0$. Our sampling now applies to lower-truncated values of the estimator, those strictly greater than 1, with a cut point $\epsilon > 0$, that is, $\sum \frac{n-1}{\log(x_i)} > 1+\epsilon$, or $\mathbb{E}(\hat{\alpha}|_{\hat{\alpha}>1+\epsilon})$: $\phi_{\hat{\alpha''}}(a) = \frac{\phi_{\hat{\alpha'}}(a)}{\int_{1+\epsilon}^{\infty}\phi_{\hat{\alpha'}}(a)\,\mathrm{d}a}$, hence the distribution of the values of the exponent conditional of it being greater than 1 becomes:

$$\phi_{\hat{\alpha''}}(a) = \frac{e^{\frac{\alpha n^2}{a-an}}\left(\frac{\alpha n^2}{a(n-1)}\right)^n}{a\left(\Gamma(n)-\Gamma\left(n,\frac{n^2\alpha}{(n-1)(\epsilon+1)}\right)\right)}, \, a \geq 1+\epsilon \tag{14}$$

3) Distribution of $\frac{n-1}{n}\frac{\hat{\alpha}}{\hat{\alpha}-1}$*:* With $z \in (1, 1+\frac{1}{\epsilon})$,

$$\Phi(z) = \frac{(\alpha(n-1))^n(z-1)^{n-1}z^{-n-1}e^{\frac{(z-1)(\alpha-\alpha n)}{z}}}{\Gamma(n)-\Gamma\left(n,\frac{(n-1)\alpha}{\epsilon+1}\right)} \tag{15}$$

4) Moments: Trying to solve the following integral, with $n, m \in \mathbb{Z}^+$, $\alpha > 1$, $0 < \epsilon < 1$, and $\Gamma(.)$ and $\Gamma(.,.)$ the gamma and incomplete gamma functions, respectively:

$$\mathbb{E}(Z) = \frac{(\alpha(n-1))^n}{\Gamma(n)-\Gamma\left(n,\frac{(n-1)\alpha}{\epsilon+1}\right)}\int_1^{1/\epsilon} e^{\frac{(z-1)(\alpha-\alpha n)}{z}}(z-1)^{n-1}z^{m-n-1}dz \tag{16}$$

for small ϵ, replacing $1+1/\epsilon$ by $1/\epsilon$. We can express the integral with gamma functions. Since

$$(y-1)^{n-1} = \sum_{i=0}^{n-1}(-1)^i\binom{n-1}{i}y^{n-1-i},$$

and since

$$\int_1^{1/\epsilon} y^{m-2-i}e^{\frac{(y-1)}{y}(\alpha-\alpha n)}dy = e^{\alpha(1-n)}\left(\frac{1}{\alpha-\alpha n}\right)^{i-m}\frac{(\Gamma(i-m+1,\alpha-n\alpha)-\Gamma(i-m+1,-(n-1)\alpha\epsilon))}{\alpha(n-1)}, \tag{17}$$

we get as a possible solution a summation showing ratios of differences of various gamma functions. Which ends our proof. □

A. Cubic Student T (Gaussian Basin)

The Student T with 3 degrees of freedom is of special interest in the literature. It is often mistakenly approximated to be Gaussian owing to the finiteness of its variance. Asymptotically, we end up with a Gaussian, but this doesn't tell us anything about the speed. The exercise below allows us to gage the speed of the law of large numbers.

The pdf Cubic Student T distribution (tail exponent equals 3).

$$p(x) = \frac{6\sqrt{3}}{\pi \left(x^2+3\right)^2},\ x \in (-\infty, \infty) \tag{18}$$

Proposition 2. *Let $MD(n)$ be the mean absolute deviation for n summands for a cubic alpha with density as in Eq. 18.*

$$MD(n) = \begin{cases} \frac{2\sqrt{3}}{\pi} & \textit{for}\ \ n=1 \\ \frac{2\sqrt{3}}{\pi}\left(\frac{1}{n}\right)^{\frac{\log\left(e^n n^{-n}\Gamma(n+1,n)-1\right)}{\log(n)}} & \textit{for}\, n>1 \end{cases} \tag{19}$$

We note the "speed" of convergence $\gamma = \left\{\gamma : \frac{MD(n)}{MD(1)} = \left(\frac{1}{n}\right)^{1-\frac{1}{\gamma}}\right\}$:
$\gamma = \frac{\log(n)}{\log(e^n n E_{-n}(n)-1)}$ where $E_{(.)}(.)$ is the exponential integral $E_n z = \int_1^\infty \frac{e^{t(-z)}}{t^n} dt$.
.

Proof. We have the Fourier:

$$\varphi(\omega) = \mathbb{E}[e^{i\omega X}] = (1+\sqrt{3}\,|\omega|)\, e^{-\sqrt{3}\,|\omega|}$$

hence the n-summed characteristic function is:

$$\varphi(\omega) = (1+\sqrt{3}|\omega|)^n\, e^{-n\sqrt{3}\,|\omega|}$$

The pdf of Y is given by:

$$p(x) = \frac{1}{\pi}\int_0^\infty (1+\sqrt{3}\,\omega)^n\, e^{-n\sqrt{3}\,\omega} \cos(\omega x)\, d\omega$$

Hence the mean absolute deviation

$$MD(n) = \frac{1}{\pi}\int_{-\infty}^\infty \int_0^\infty |x|\,(1+\sqrt{3}\,\omega)^n\, e^{-n\sqrt{3}\,t} \cos(\omega x)\, d\omega\, dx \tag{20}$$

□

VI. Additional Applications: Convergence of the Variance

The finance literature measures the standard deviation or "volatility" of power law distributed variables with finite variance. It is erroneous as the variance itself will be power law distributed with half the exponent, hence can have infinite variance itself.

So one can obtain the same results as above showing that the sum of the squares of a random variable, say Student T or Double-Pareto distributed obeys the same convergence restrictions. [TO BE COMPLETED]

VII. Acknowledgement

Colman Humphrey, Michael Lawler...

References

[1] M. Hlusek, "On distribution of absolute values," 2011.
[2] I. Pinelis, "Characteristic function of the positive part of a random variable and related results, with applications," *Statistics & Probability Letters*, vol. 106, pp. 281–286, 2015.
[3] V. V. Uchaikin and V. M. Zolotarev, *Chance and stability: stable distributions and their applications*. Walter de Gruyter, 1999.
[4] V. M. Zolotarev, *One-dimensional stable distributions*. American Mathematical Soc., 1986, vol. 65.
[5] G. Samorodnitsky and M. S. Taqqu, *Stable non-Gaussian random processes: stochastic models with infinite variance*. CRC Press, 1994, vol. 1.
[6] S. J. Wolfe, "On the local behavior of characteristic functions," *The Annals of Probability*, pp. 862–866, 1973.
[7] C. D. Hardin Jr, "Skewed stable variables and processes." DTIC Document, Tech. Rep., 1984.