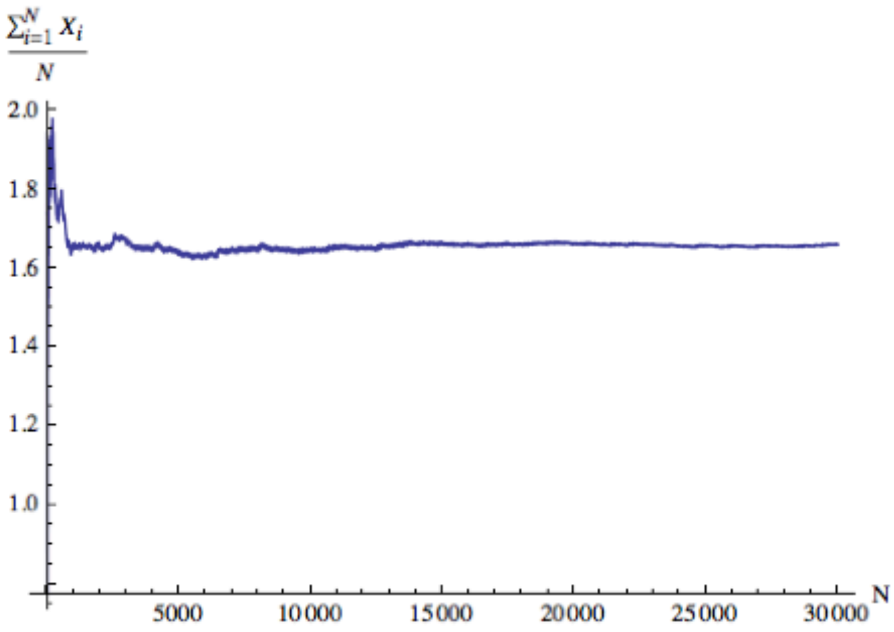
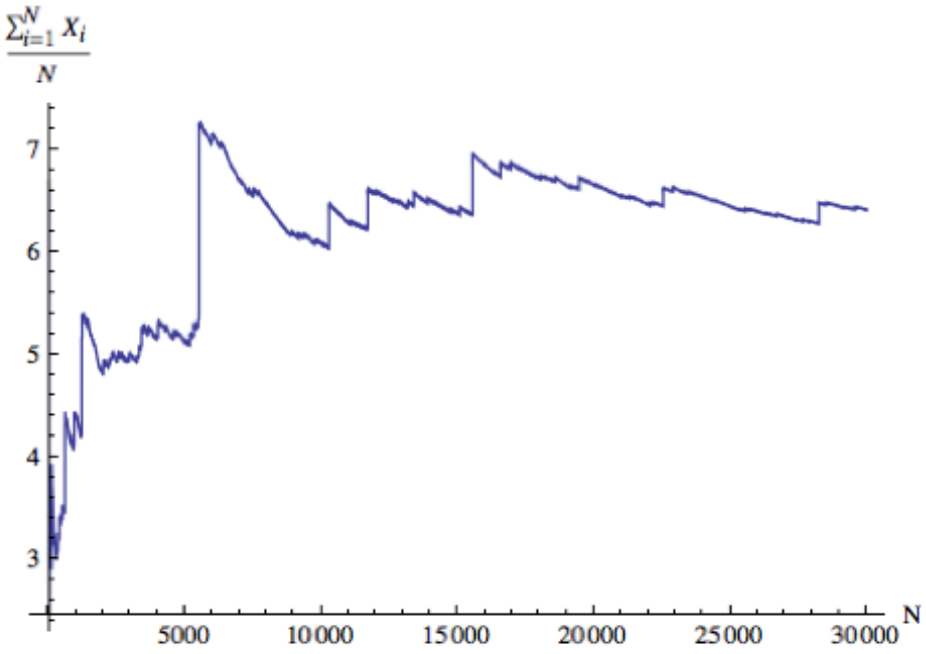
The Law of Large Numbers in the Real World

Nassim Nicholas Taleb

Extreme Risk Initiative, NYU School of Engineering

Law of Large Numbers
Convergence in norm L^1



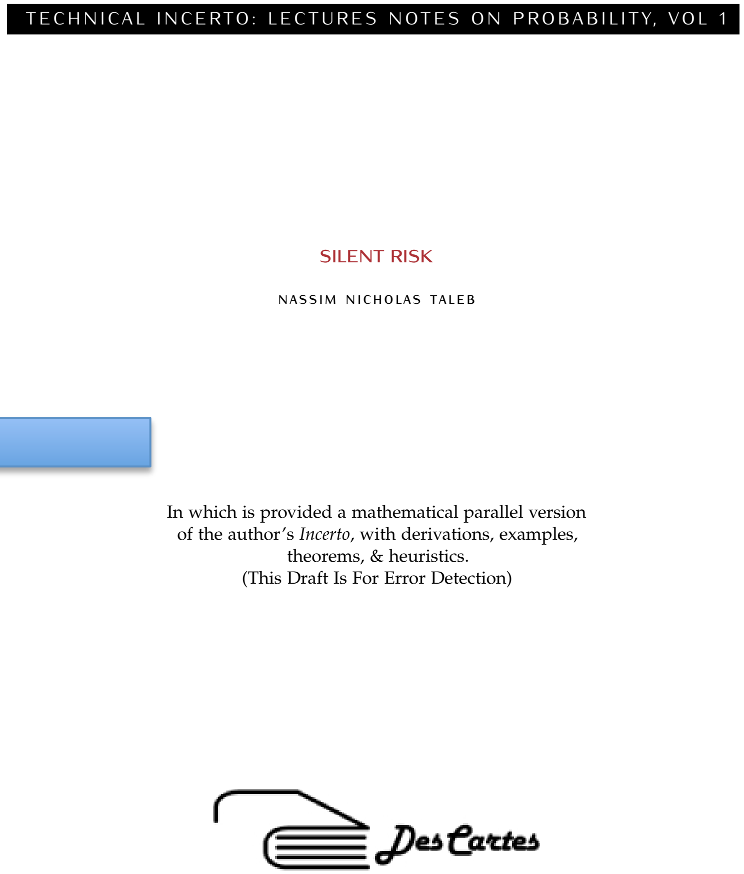


Real vs. Incremental Research

For Journals (Incremental)

- w/ Cirillo, 2015, (Underestimation of Violence)
- w/Douady, 2015,(Centiles) *Physica A*
- 2015, in preparation (Gini)
- -------, in preparation (CLT)

More Comprehensive Work



TECHNICAL INCERTO: LECTURES NOTES ON PROBABILITY, VOL 1

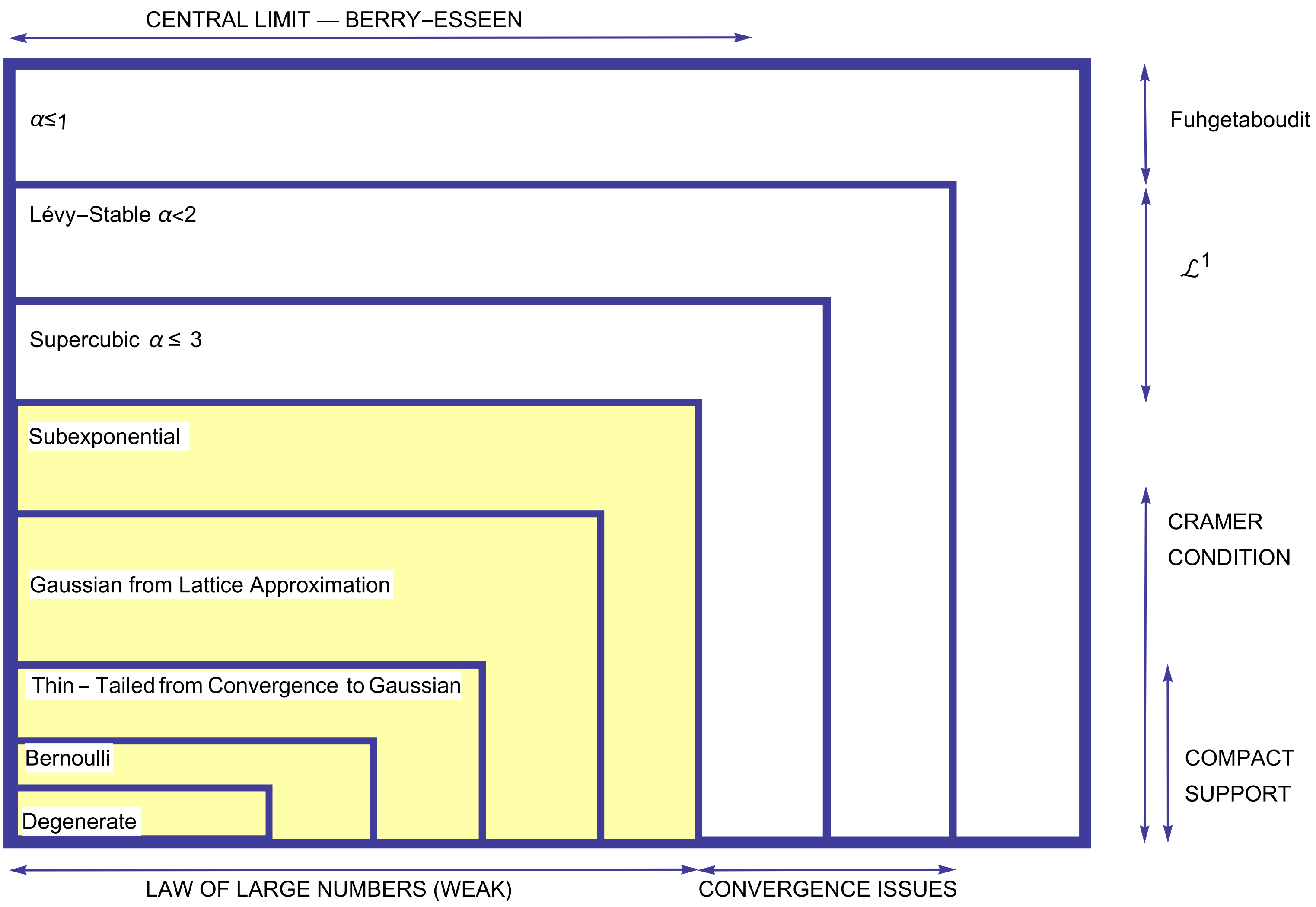
SILENT RISK

NASSIM NICHOLAS TALEB

In which is provided a mathematical parallel version
of the author's *Incerto*, with derivations, examples,
theorems, & heuristics.
(This Draft Is For Error Detection)

DesCartes

Classes of distributions



CENTRAL LIMIT — BERRY–ESSEEN

$\alpha \leq 1$

Lévy–Stable $\alpha < 2$

Supercubic $\alpha \leq 3$

Subexponential

Gaussian from Lattice Approximation

Thin – Tailed from Convergence to Gaussian

Bernoulli

Degenerate

Fuhgetaboudit

$\mathcal{L}^1$

CRAMER
CONDITION

COMPACT
SUPPORT

LAW OF LARGE NUMBERS (WEAK)

CONVERGENCE ISSUES

Definition Paretian Tail

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 $k > 0$.

Simplest Forms

One-Tailed

- 1-2 Parameter: Minimum value, tail (Pareto I)
- 3 Parameters: location m, scale s, tail α (Pareto 2-Lomax)
- 4 Parameters: Generalized Beta 2nd kind (includes many known distributions such as Singh-Maddala, etc.)

Two Tailed

- Student T distribution (finance papers galore)
- Lévy-Stable (to which all those with $\alpha<2$ converge)
- Other (double Pareto, etc.)

Sample Equivalence

Dispersion of outcomes: we cannot use standard deviation and other tools since no second moment. Only MAD , mean absolute deviation of the mean from "true" mean (or 0).

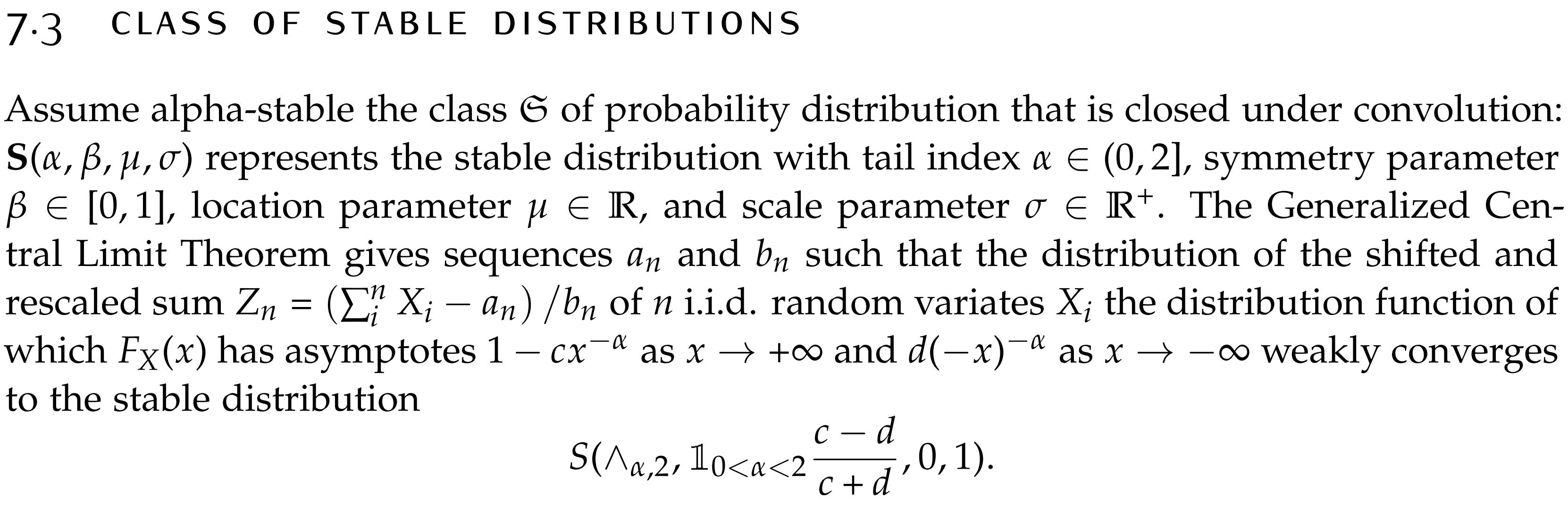
By the weak law of large numbers, consider a sum of random variables X_1, X_2,..., X_n independent and identically distributed with finite mean m, that is E$[X_i] < \infty$, then $\frac{1}{n}\sum_{1\leq i\leq n} X_i$ converges to m **in probability**, as n $\rightarrow \infty$. And the idea is that we live with finite n.

Indexing by *p* for powerlaw and *g* Gaussian:

$$\inf\left\{n_p : E\left(\left|\sum^{n_p}\frac{X_i^p - m_p}{n_p}\right|\right) \leq E\left(\left|\sum^{n_g}\frac{X_i^g - m_g}{n_g}\right|\right)\ , n_p > 0\right\}$$

Sum of all powerlaw r.v. w/ exponent α<2

Basin of convergence (Generalized CLT)



7.3 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

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

Derivations

- Working with Fourier Transforms, we are able to derive partial expectation above *K*, E^K|*X*|(option-like) in semi-closed form. For MAD= 2 E^K|*X*|, *K*=0 or *K*=μ:

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

$$n_\alpha^\beta = 2^{\frac{\alpha}{1-\alpha}} \pi^{\frac{\alpha}{2-2\alpha}} \left(\Gamma\left(\frac{\alpha-1}{\alpha}\right) \sqrt{n_g} \left(\left(1 - i\beta \tan\left(\frac{\pi\alpha}{2}\right)\right)^{\frac{1}{\alpha}} + \left(1 + i\beta \tan\left(\frac{\pi\alpha}{2}\right)\right)^{\frac{1}{\alpha}} \right) \right)^{\frac{\alpha}{\alpha-1}} \qquad (7.6)$$

Sample size *n* equivalent

Pareto's "80/20" gives α≈1.16

α	n_α	$n_\alpha^{\beta=\pm\frac{1}{2}}$	$n_\alpha^{\beta=\pm 1}$
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

Speed of convergence

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

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

Back door working

- Sample mean ("realized a average" in language of finance) is never Gaussian when α<2 (even when >2, **another story on CLT (next chapter in *Silent Risk*)**)
- Tail α can be estimated with MLE
- Tail α from MLE is asymptotically Gaussian (preasymptotically Lognormal with low variance reaches v. quickly)
- We can fit GPD or EVD with same tail exponent α, further reducing variance.

Remark 1. *If* $\hat{\theta}$ *is the maximum likelihood estimator (MLE) of* θ*, then for an absolutely continuous function* ϕ*,* $\phi(\hat{\theta})$ *is the MLE estimator of* $\phi(\theta)$.

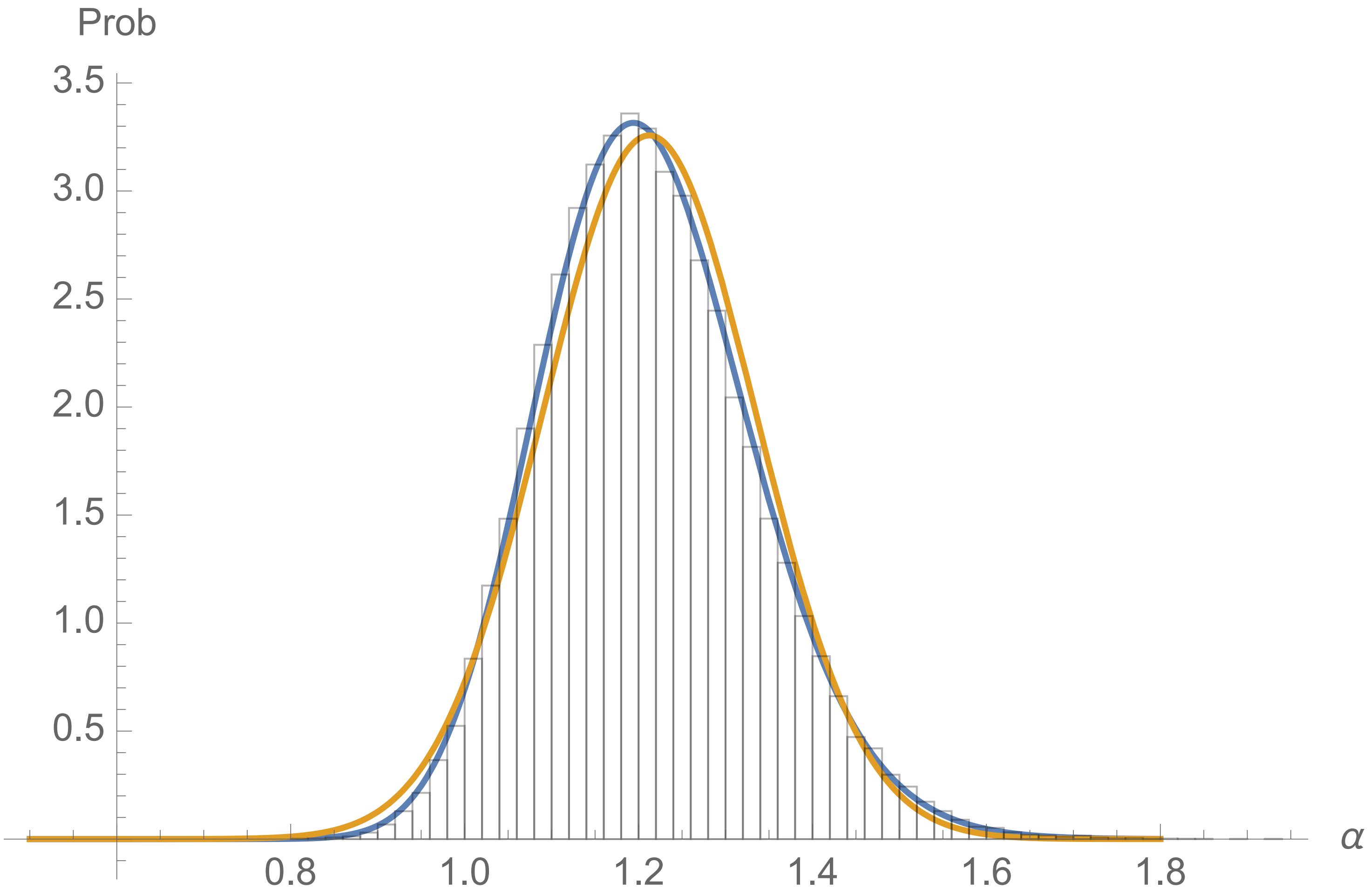
MLE for α

Lognormal

vs

Normal

Difference fades quickly!



Case 1

Pareto distribution with a single parameter to fit (actually even 2).

$$f(x) = \begin{cases} \alpha L^{\alpha} x^{-\alpha-1} & x \geq L \\ 0 & \text{Otherwise} \end{cases}$$

$$\mathbb{E}(x) = \begin{cases} L \frac{\alpha}{\alpha-1} & \alpha > 1 \\ \text{Indeterminate} & \text{Otherwise} \end{cases}$$

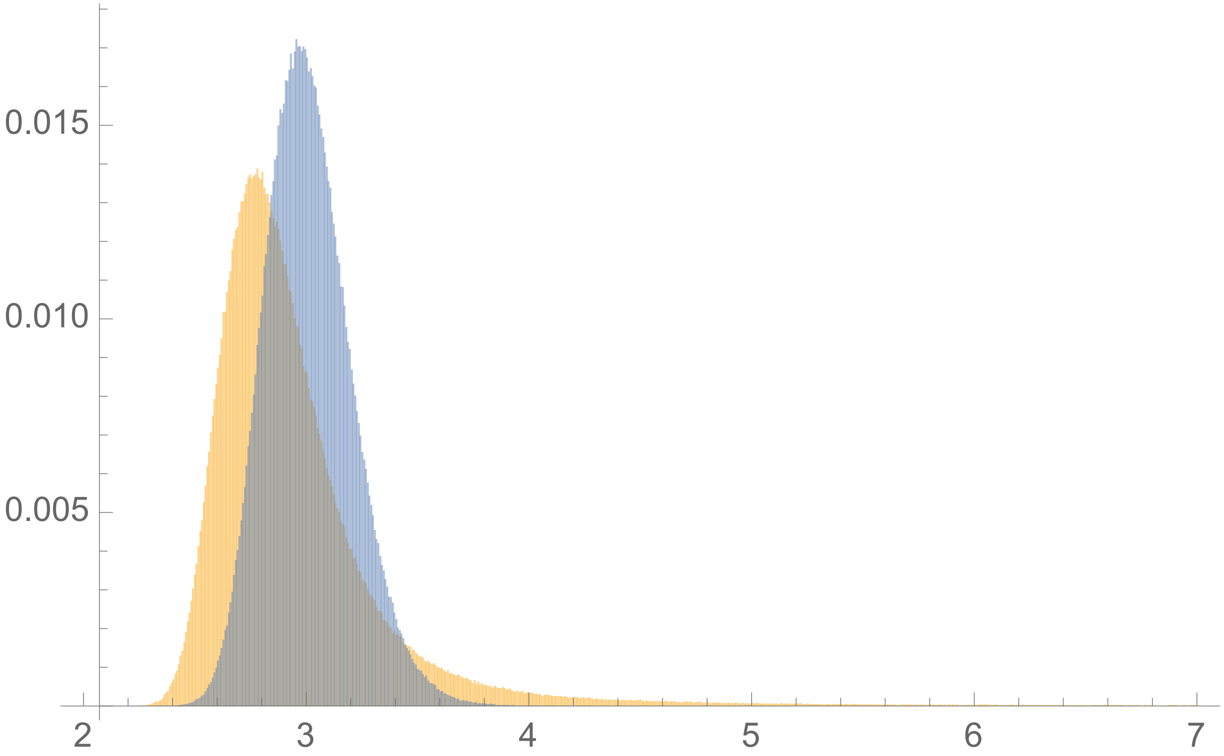
“Realized” vs MLE smaller samples

n= 10^3

Ratio of Std: infinite (in sample v. unstable/high)

Ratio of Mean deviations: 2

Skewness: huge for realized



“Realized” vs MLE larger samples

***n*= 10⁵**

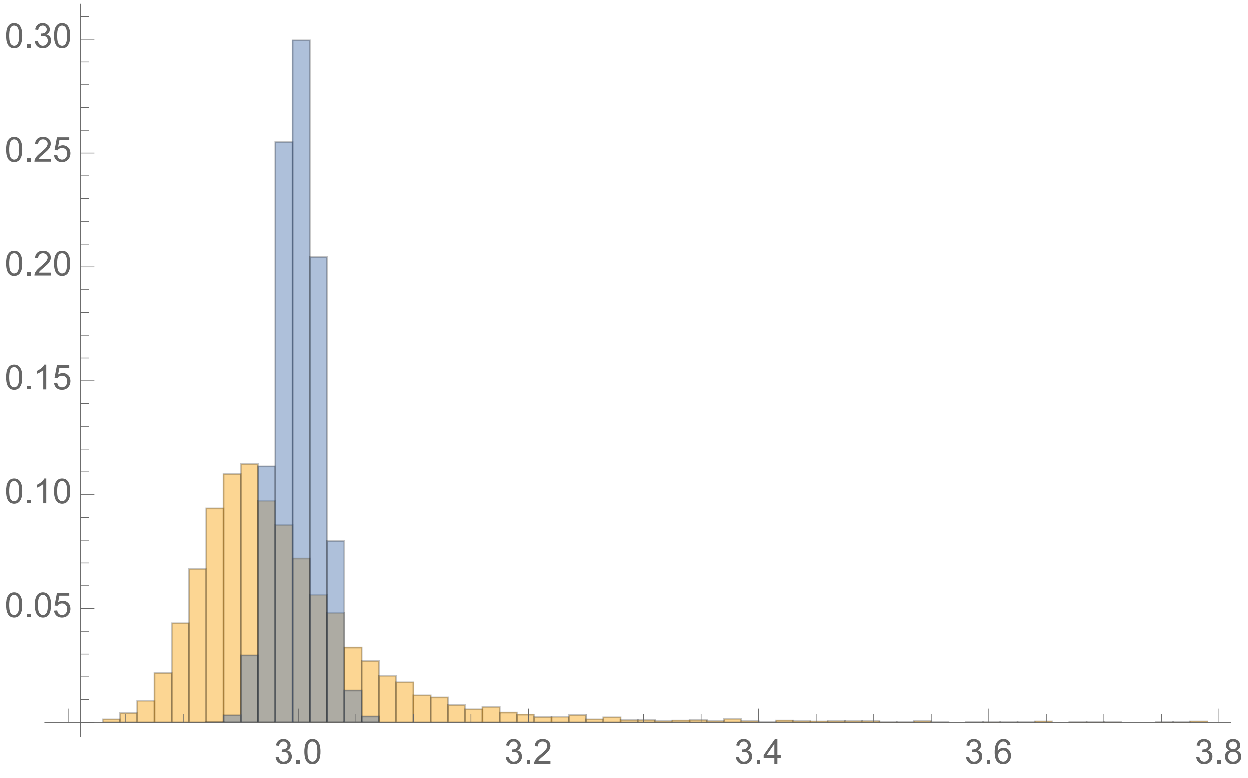
Ratio of Std: infinite.

Ratio of Mean deviations: 4.7

Skewness: remains huge

for realized

Realized mean usually *lower*…



Simple conclusion

We cannot use the mean to estimate the mean.

Tail α dominates *especially* in larger samples.

We can show it analytically from $L\frac{\alpha}{\alpha-1}$ when α is Lognormal, and Normal

(see *Silent Risk* for distribution)

A few points

When $1 < \alpha < 2$ we can safely say that the sample mean is insignificant and underestimates the mean (reason: *infinite skewness*). For *all* sample size.

In some cases, with $\alpha < 1$, we can extract the "true" mean, albeit stochastic. Case study on violence (Cirillo and Taleb, 2015).

Compact Support

Next, with today's maximum population H and L the naively rescaled minimum for our definition of conflict, we introduce a smooth rescaling function $\varphi : [L, H] \to [L, \infty)$ satisfying:

i φ is "smooth": $\varphi \in C^\infty$,
ii $\varphi^{-1}(\infty) = H$,
iii $\varphi^{-1}(L) = \varphi(L) = L$.

In particular, we choose:

$$\varphi(x) = L - H \log\left(\frac{H - x}{H - L}\right). \tag{1}$$

Perform Analytics on Power Law

First, work on statistical significance of power laws on transformed variables and derive parameters

The distribution of x can be rederived as follows from the distribution of X':

$$\int_L^{\infty} f(x')\,\mathrm{d}x' = \int_L^{\varphi^{-1}(\infty)} g(x)\,\mathrm{d}x,$$

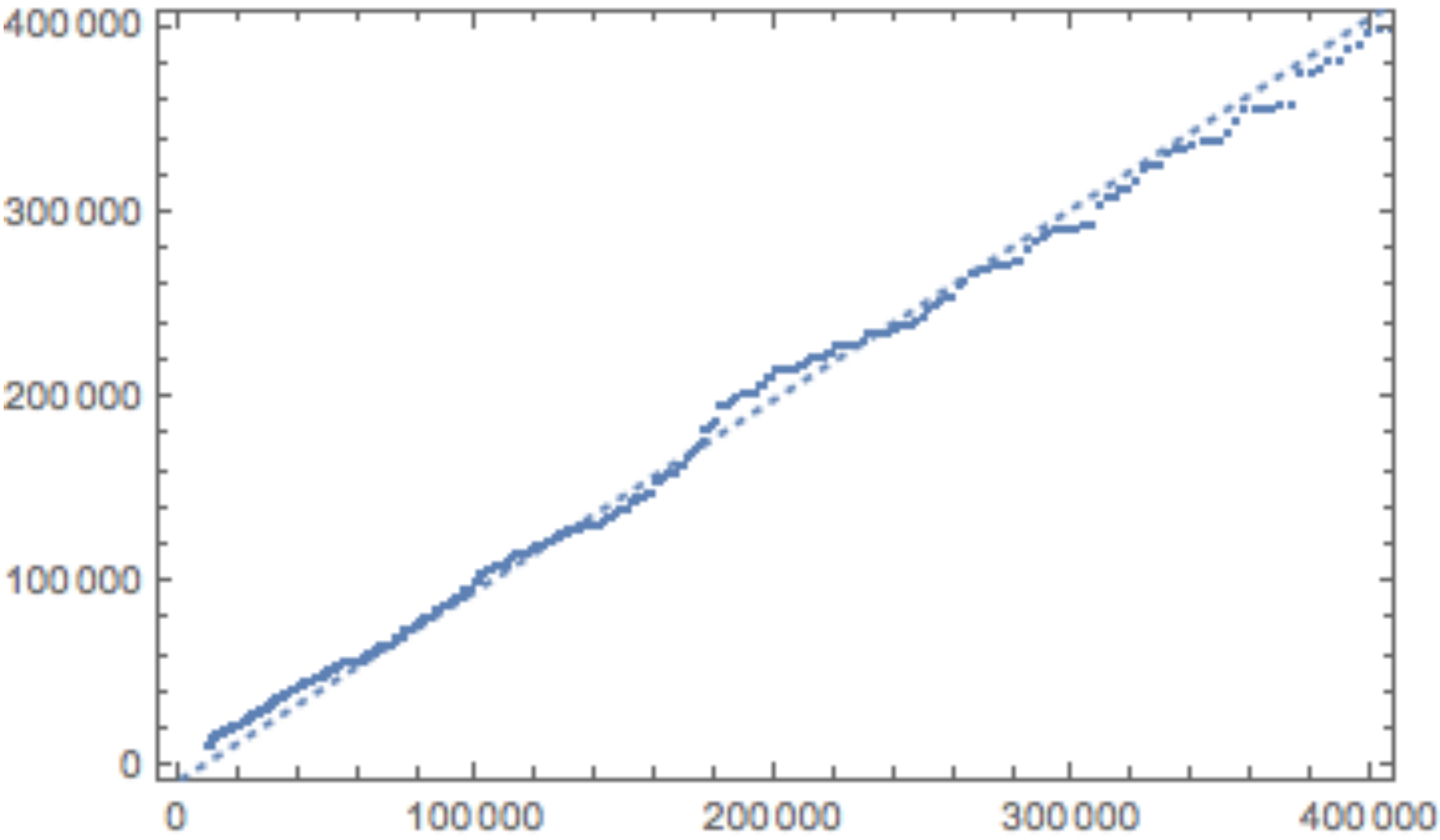
where $\varphi^{-1}(u) = (L - H)e^{\frac{L-u}{H}} + H$.

Parameters equivalence

$$f(x_r) = \frac{\alpha \left(\frac{-L+\sigma+x_r}{\sigma}\right)^{-\alpha-1}}{\sigma}, x_r \in [L, \infty)$$

$$g(x) = \frac{\alpha H \left(\frac{\sigma - H \log\left(\frac{H-x}{H-L}\right)}{\sigma}\right)^{-\alpha-1}}{\sigma(H-x)}, x \in [L, H],$$

Near tail Q-Q Plot



GPD for far tail

GPS's parameter 1/ζ matches near tail α:

➔ Use Pareto Lomax

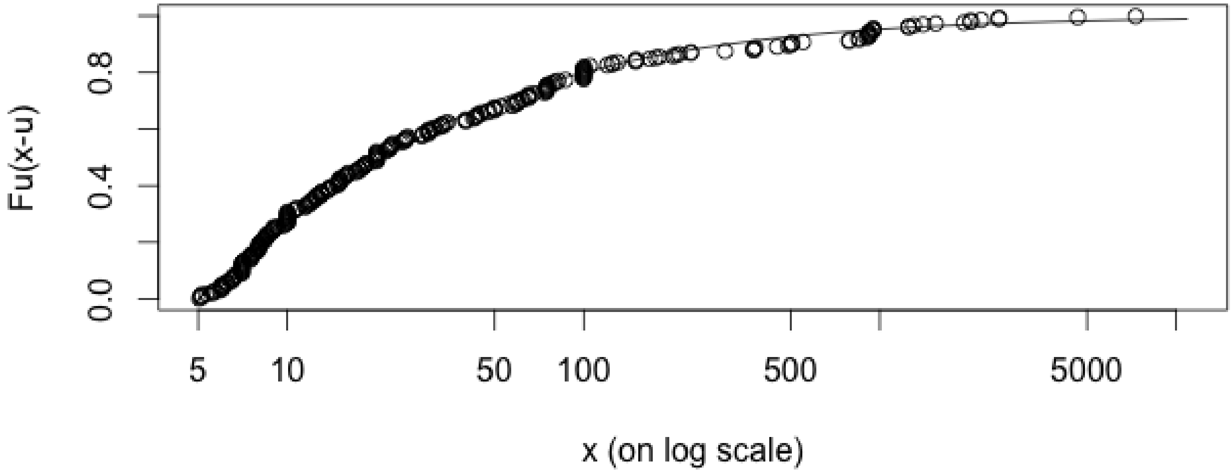


Fig. 10. GPD cumulative distribution fitting to actual casualties' data (in 10k). Parameters as per Table IV, first line.

“True mean” vs Realized Mean

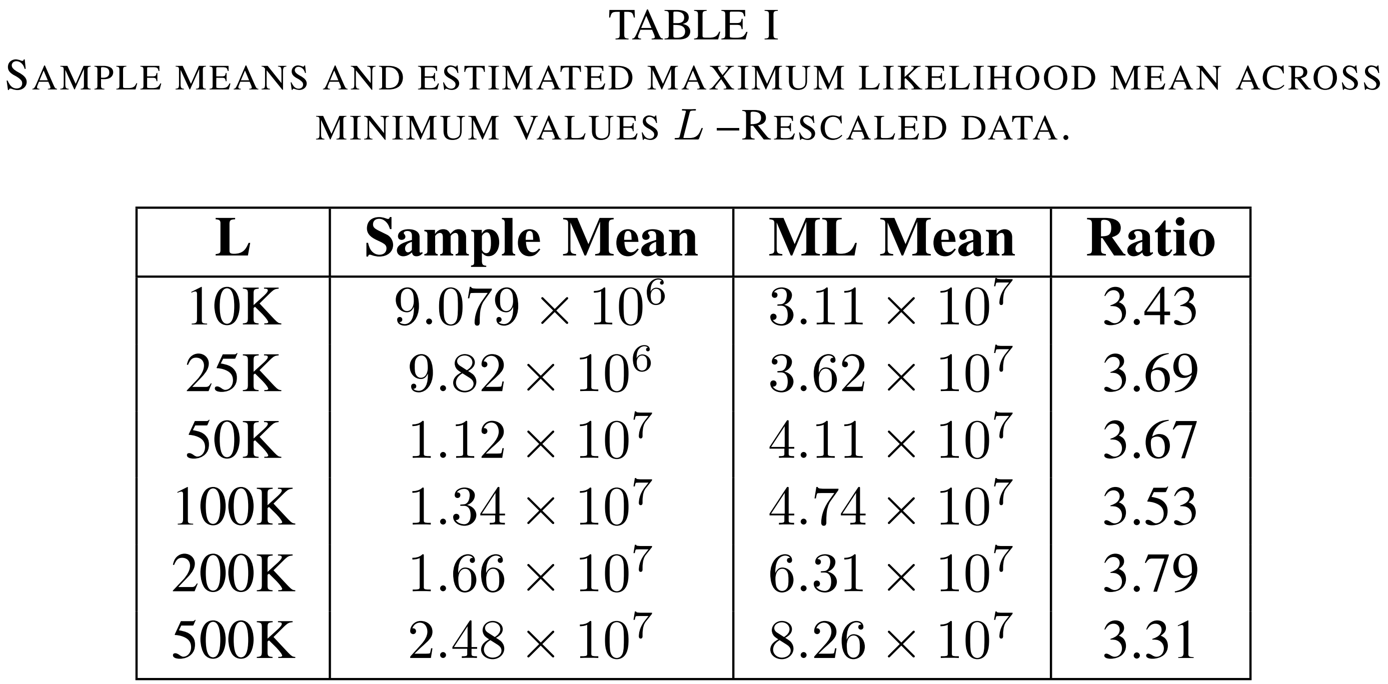


TABLE I

Sample means and estimated maximum likelihood mean across minimum values L –Rescaled data.

L	Sample Mean	ML Mean	Ratio
10K	9.079×10^6	3.11×10^7	3.43
25K	9.82×10^6	3.62×10^7	3.69
50K	1.12×10^7	4.11×10^7	3.67
100K	1.34×10^7	4.74×10^7	3.53
200K	1.66×10^7	6.31×10^7	3.79
500K	2.48×10^7	8.26×10^7	3.31

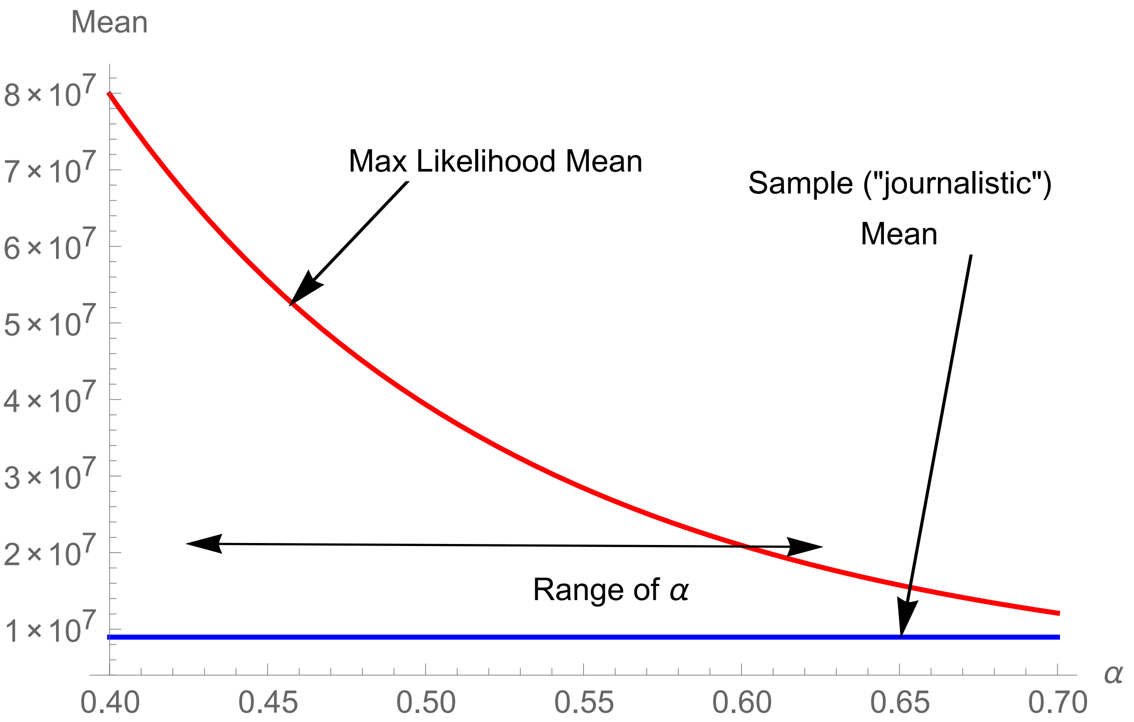


Fig. 5. Observed "journalistic" mean compared to MLE mean (derived from rescaling back the data to compact support) for different values of α (hence for permutations of the pair (σ_α, α)). The "range of α is the one we get from possible variations of the data from bootstrap and reliability simulations.

Application: the Piketty craze.

Centile contribution: is Pareto 80/20 true?

80/20 by recursing ➔ Top 1% has 53%

Huuuuuuuuge bias in mean measurement as bracketed

➔ y-o-y changes suspicious

Quantile Contribution

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

$$\hat{h}(q) = \inf\{h : \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{x>h} \leq q\}$$

Bias

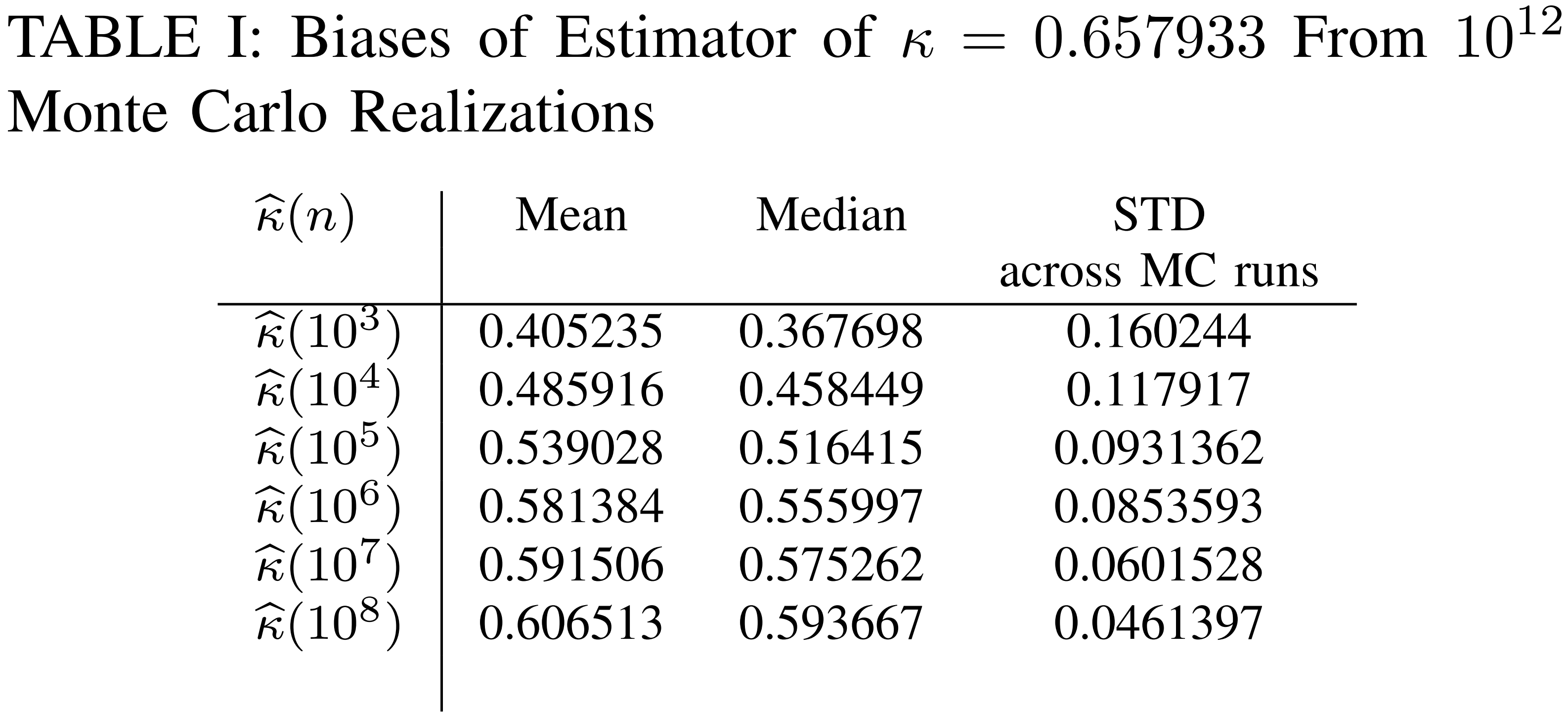


TABLE I: Biases of Estimator of $\kappa = 0.657933$ From 10^{12} Monte Carlo Realizations

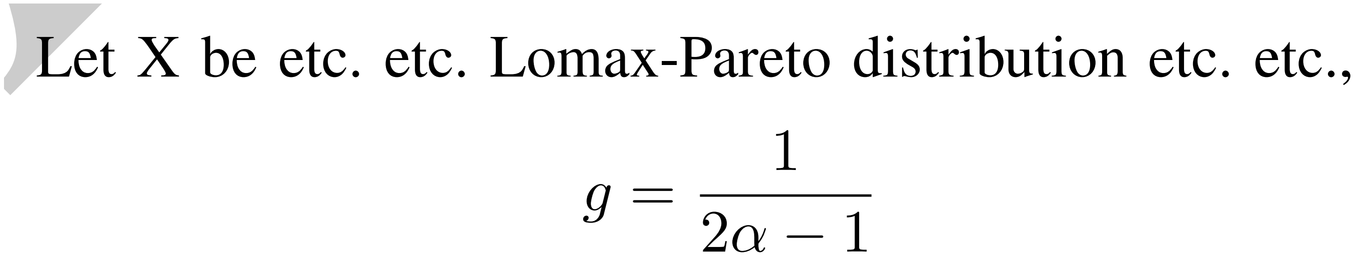
$\widehat{\kappa}(n)$	Mean	Median	STD across MC runs
$\widehat{\kappa}(10^3)$	0.405235	0.367698	0.160244
$\widehat{\kappa}(10^4)$	0.485916	0.458449	0.117917
$\widehat{\kappa}(10^5)$	0.539028	0.516415	0.0931362
$\widehat{\kappa}(10^6)$	0.581384	0.555997	0.0853593
$\widehat{\kappa}(10^7)$	0.591506	0.575262	0.0601528
$\widehat{\kappa}(10^8)$	0.606513	0.593667	0.0461397

GINI Coefficient

$$g = \frac{1}{2} \frac{\mathbb{E}\left(|X - X'|\right)}{\mu}.$$

$$\widehat{G}(Y) = \frac{\sum_{j=1}^{n} \sum_{i=1}^{n} |Y_i - Y_j|}{2(n-1) \sum_{i=1}^{n} Y_i}$$

Bias in Gini



Let X be etc. etc. Lomax-Pareto distribution etc. etc.,

$$g = \frac{1}{2\alpha - 1}$$

True value *g**=.8333 with α=1.1

Population, not sample:

Population= 10^4, g^= .75, std=.043

Population= 10^5, g^= .77, std=.042

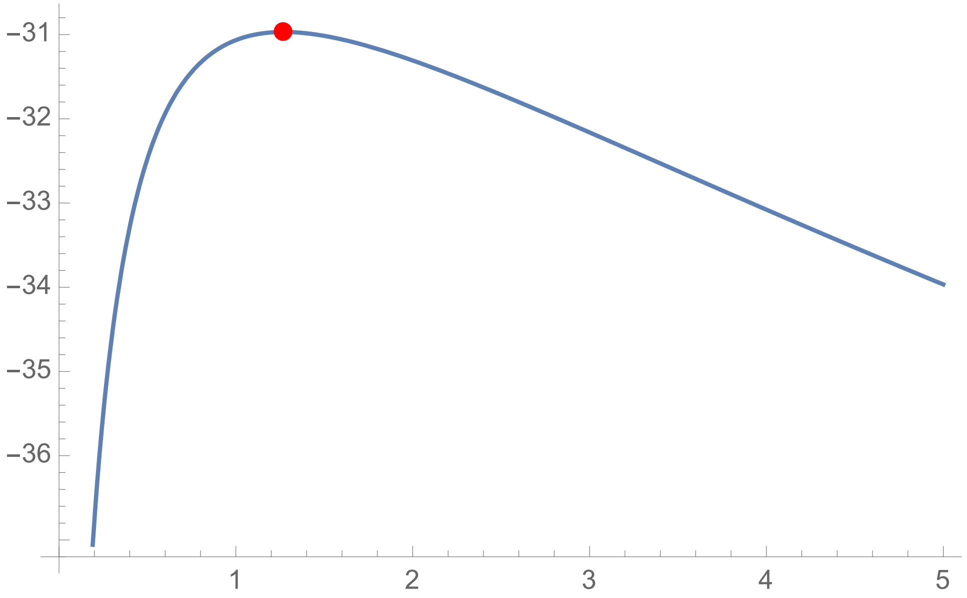
You get direct GINI from α *<u>without bias!</u>*

Now the ~~bad~~ excellent news

“Stochastic vol” with Student T (a lot of papers…) with 0 mean. 2 Parameters to estimate, *α, s* vs (central or noncental) 2nd moment.

$$f(x) = \frac{\left(\frac{\alpha}{\alpha+\frac{x^2}{s^2}}\right)^{\frac{\alpha+1}{2}}}{\sqrt{\alpha}\; s\, \mathrm{B}\left(\frac{\alpha}{2},\ \frac{1}{2}\right)} \qquad \sigma = \sqrt{\frac{\alpha}{\alpha-2}}\; s$$

MLE for *s*



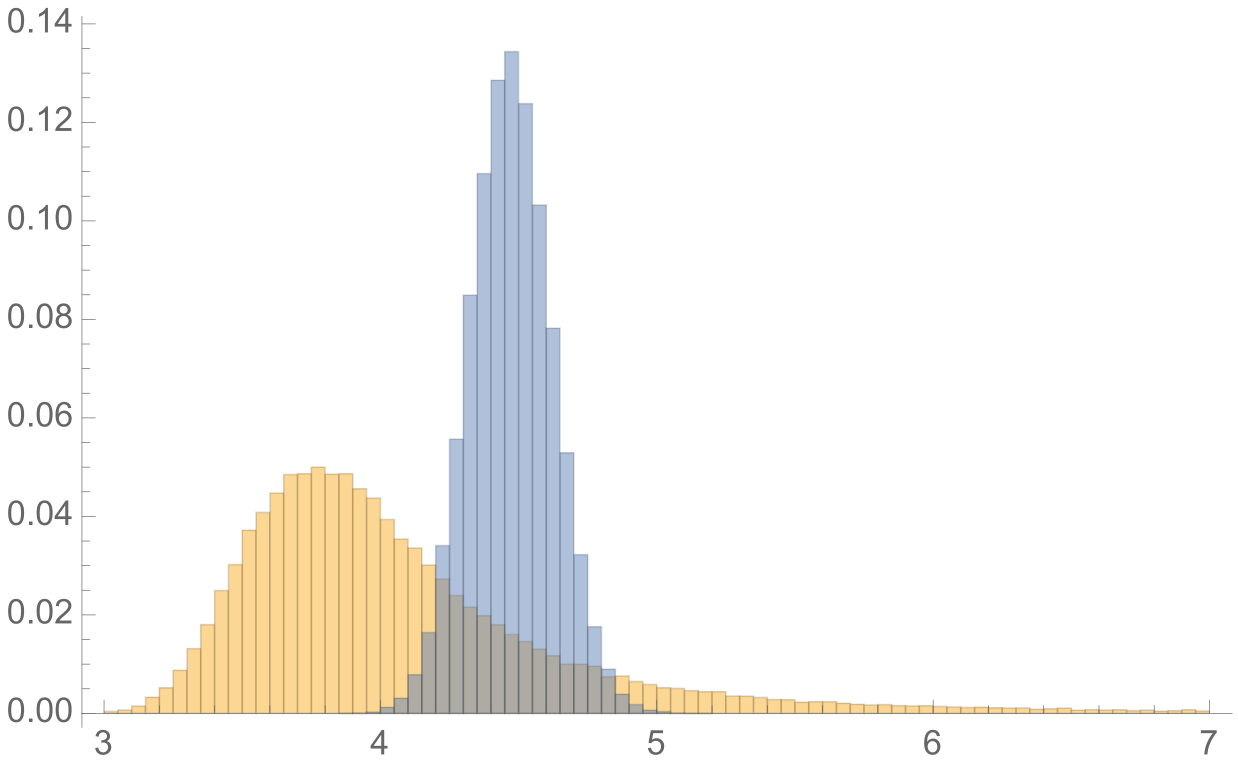
Both α and *s* vs realized vol

Tail α=5/2

No variance of variance (x^2 follows FRatio(1,α): tail exponent ½ α

Realized volatility is not an estimator of realized volatility

MD ratio: **5** ➔ **delivered volatility 5 x more volatile than true volatility.**



Conclusion

- Much of finance, social science, relies on *bogus* estimators. For instance Pinker's problem is quite insidious with mechanistic users of statistics.
- We "recalibrate" models because **they are not estimators**, chasing past fitness.
  - As the late Benoit Mandelbrot said: *when a lightning hits we do not change the laws of nature*.
- Excellent news: rigorous methods, including using extreme value theory and development of new estimators clears up a lot of problems