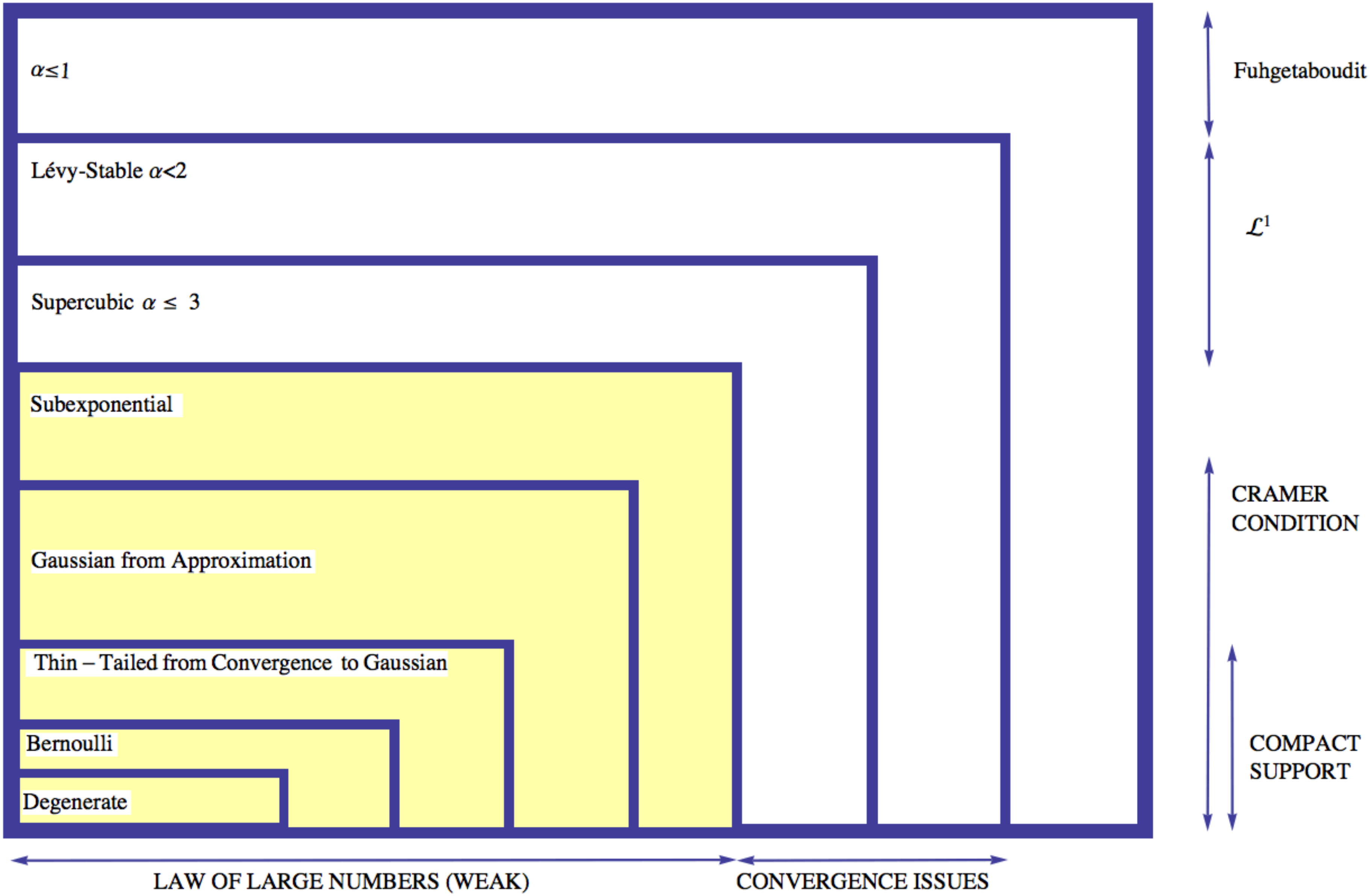
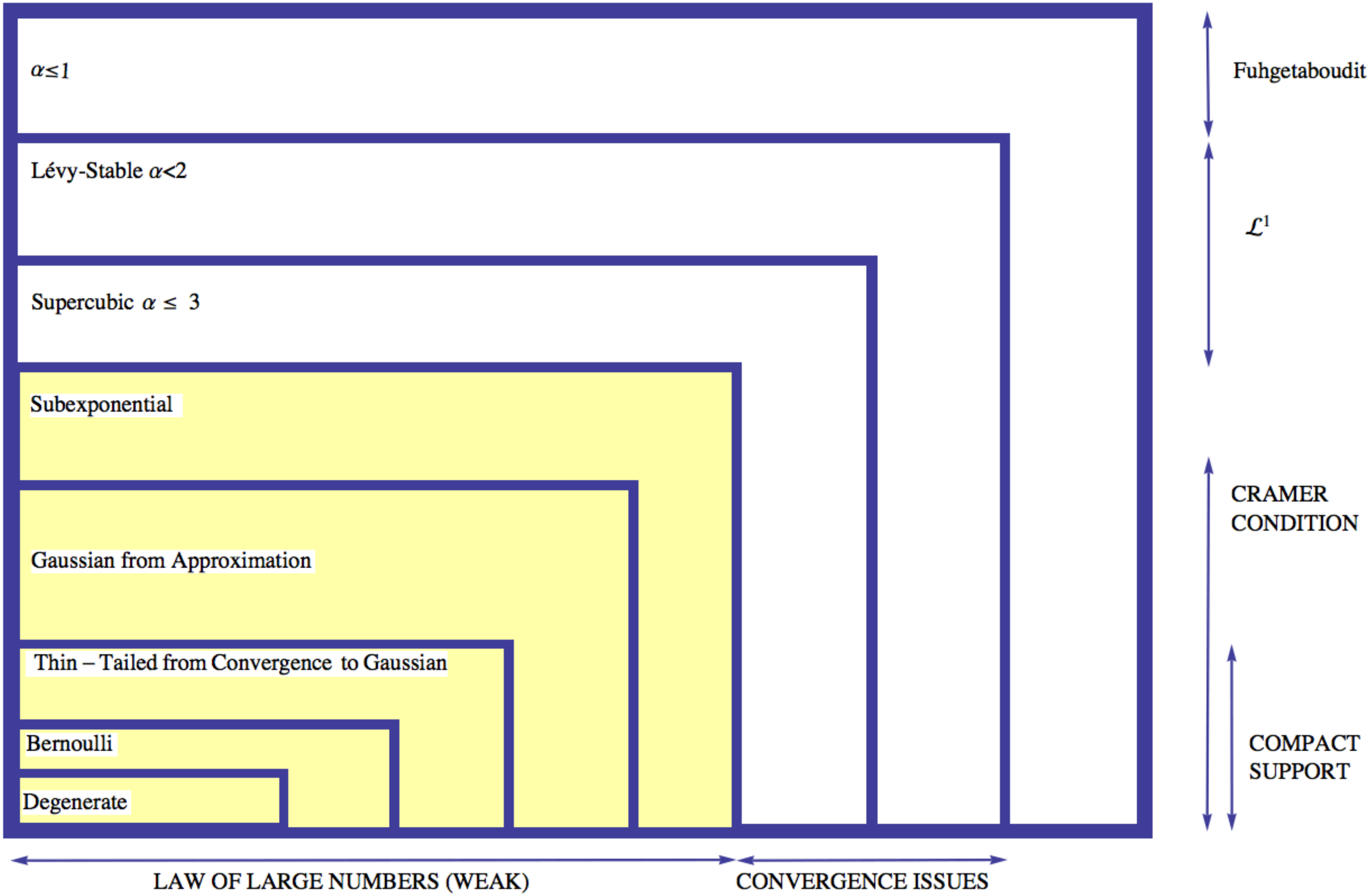
Statistics and Fat Tails


α≤1
Lévy-Stable α<2
Supercubic α ≤ 3
Subexponential
Gaussian from Approximation
Thin – Tailed from Convergence to Gaussian
Bernoulli
Degenerate
Fuhgetaboudit
$\mathcal{L}^1$
CRAMER CONDITION
COMPACT SUPPORT
LAW OF LARGE NUMBERS (WEAK)
CONVERGENCE ISSUES

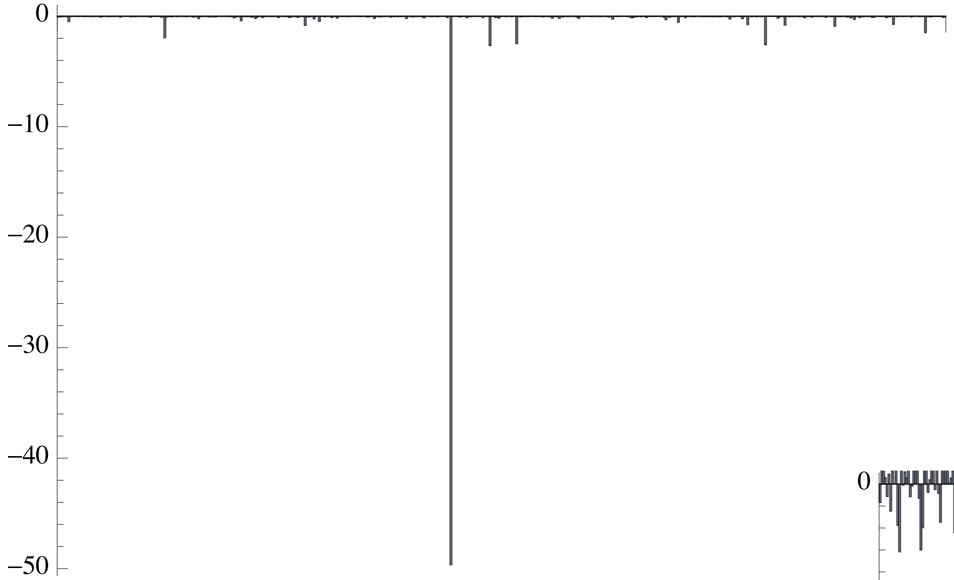
Mediocristan

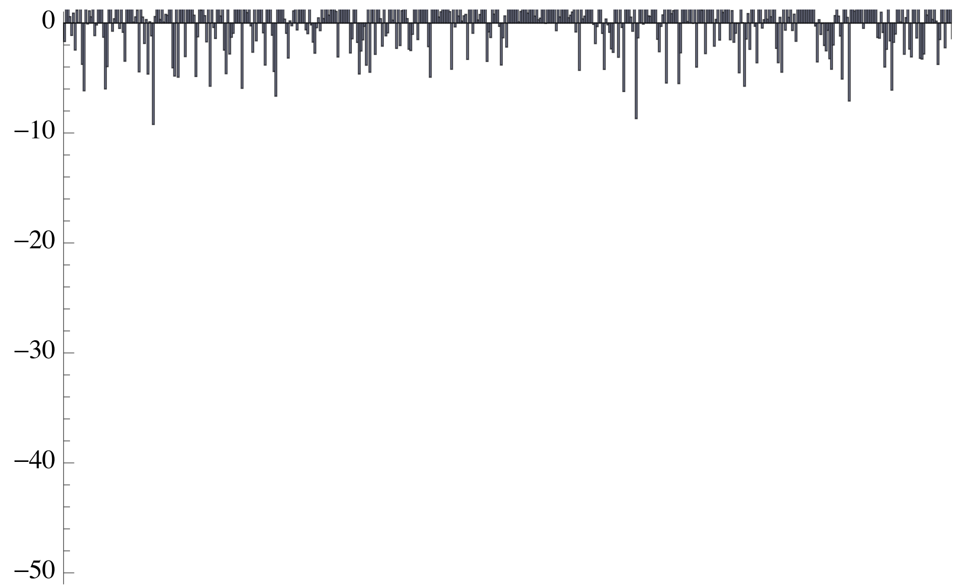
Extremistan

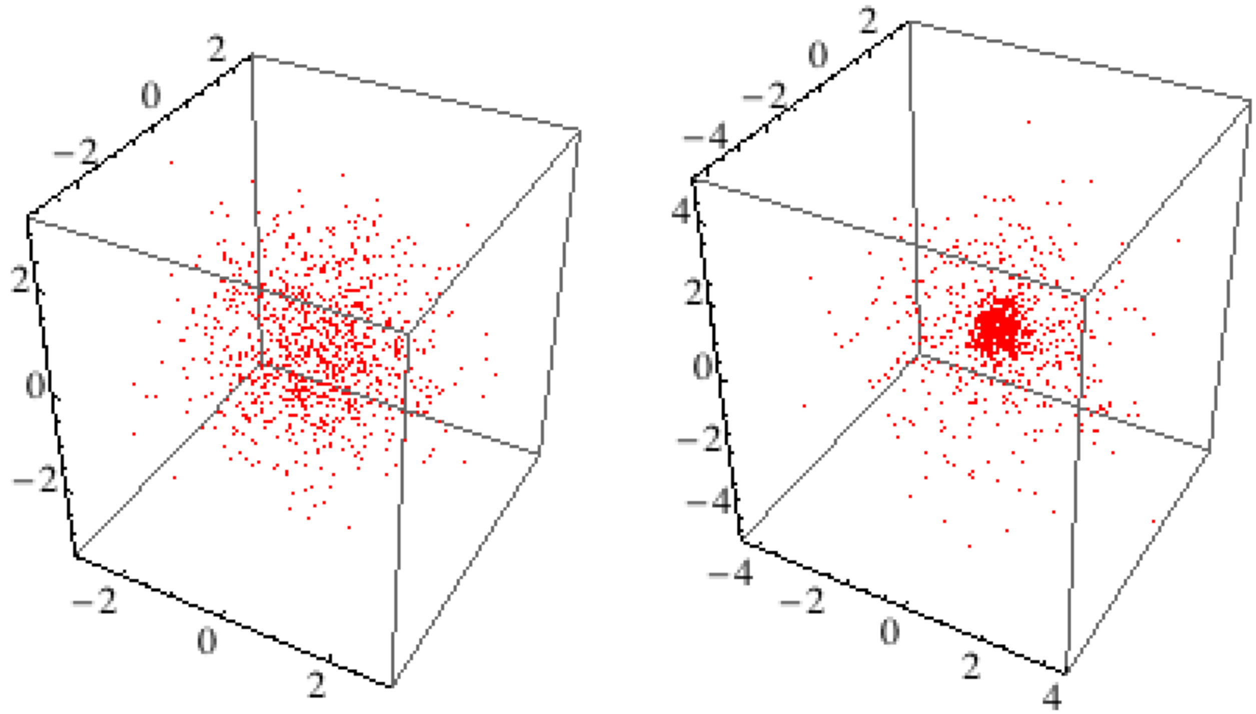


$\alpha \leq 1$
Lévy-Stable $\alpha < 2$
Supercubic $\alpha \leq 3$
Subexponential
Gaussian from Approximation
Thin – Tailed from Convergence to Gaussian
Bernoulli
Degenerate
Fuhgetaboudit
$\mathcal{L}^1$
CRAMER CONDITION
COMPACT SUPPORT
LAW OF LARGE NUMBERS (WEAK)
CONVERGENCE ISSUES

Mediocristan/Extremistan

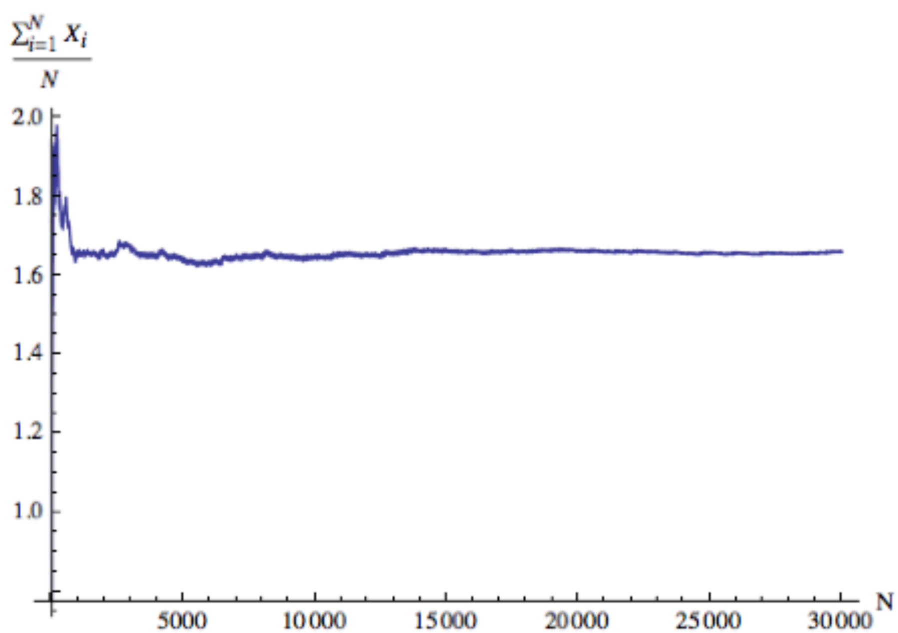


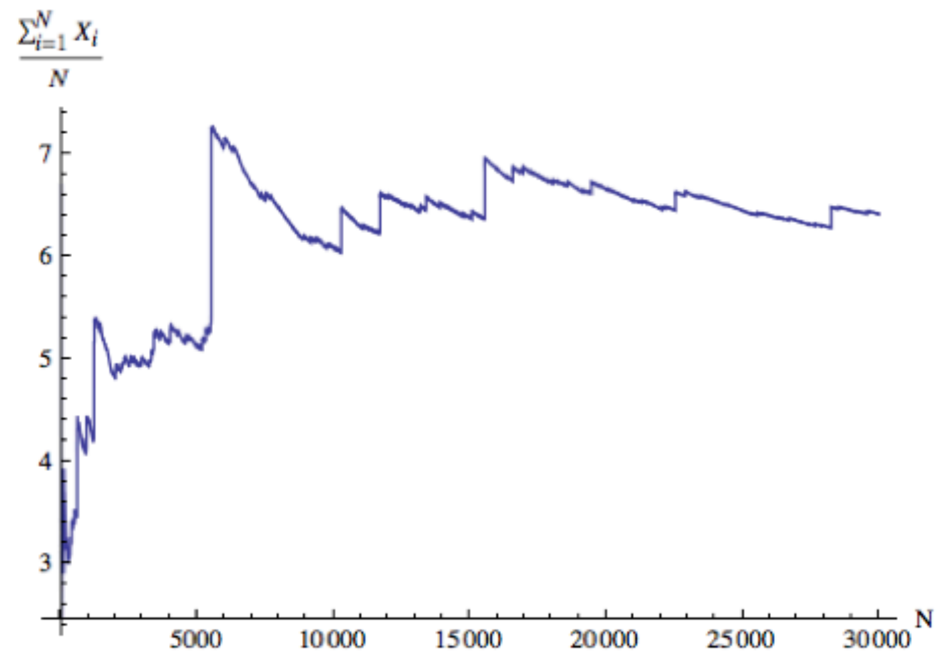




Multidimensional Fat Tails: For a 3 dimentional vector, thin tails (left) and fat tails (right) of the same variance.

Law of Large Numbers





Problem central with **Geostatistics**

- Synthesis of paper Taleb-Douady currently under (small) revision *Physica A: Statistical Mechanics and Applications*
- This is another problem where statistical methods fail with fat tails.
- Ignoring fat tails self-contradictory when people alarmed at (wealth or other) "concentration" since

concentration ⇔fattailedness

- Centile measures **super-additive**.
- Both Piketty and his detractors (*Financial Times*) made the same mistake.

Centile Contribution

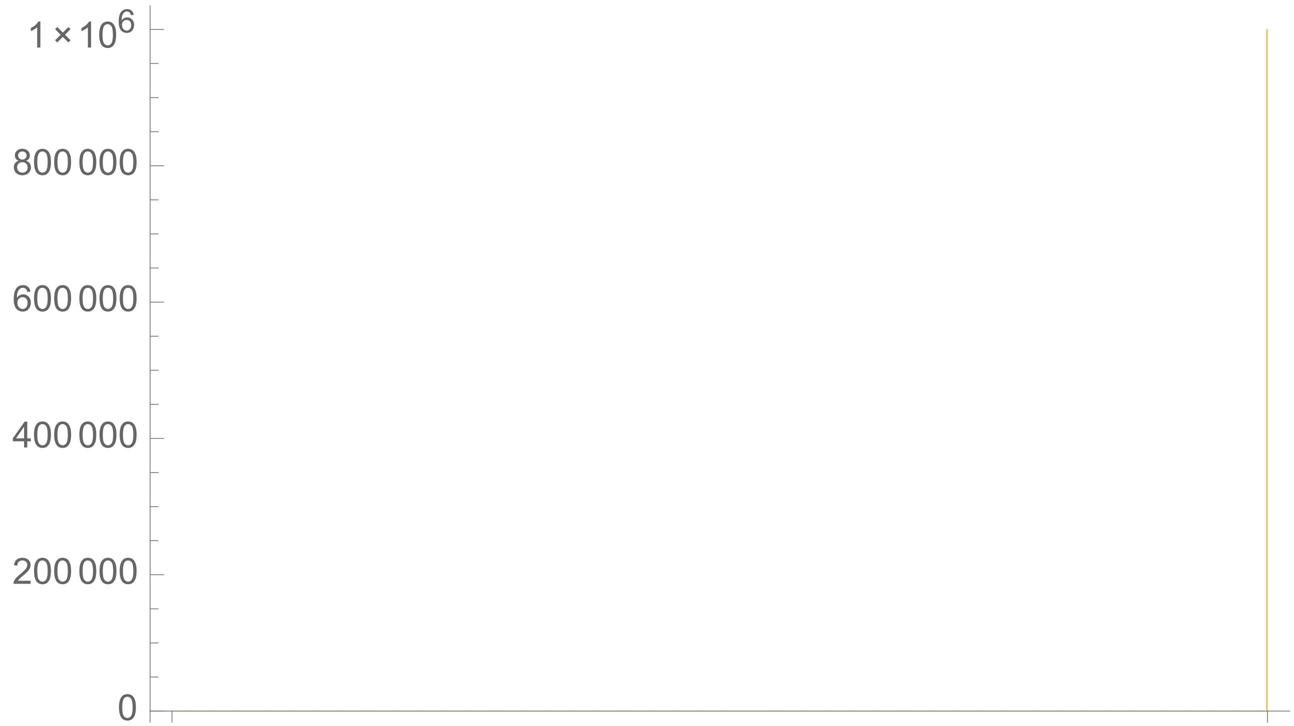
- Share of the top q% with *n* observations, where X is wealth (or something else), h is Centile:

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

$$h(q) = \inf\{h \in [x_{min}, +\infty), \mathbb{P}(X > h) \leq q\}$$

Country A (Extreme Concentration)

$\kappa_{q,A} = 0.999$



q= .01 in all exercise Both A and B countries have same *n* $=10^3$

Country B (Low Concentration)

$\kappa_{q,B}$ = 0.098

The pool of A U B $\kappa_{q,A \cup B}$ = 0.99812

The average of A and B $\frac{(\kappa_{q,A} + \kappa_{q,B})}{2}$ = 0.5486

Adjusting by the Mean Income

$$\frac{(\kappa_{q,A} + \kappa_{q,B})}{2} = 0.5486$$

Case 1: Making A and B have *exactly* same average, but keeping their previous concentration.

$$\kappa_{q,\omega_A A \cup \omega_B B} = 0.5676$$

Case 2: Making A and B have variations in average makes concentration.

$$\kappa_{q,\omega_A A \cup \omega_B' B} = 0.7114$$

BIG DATA

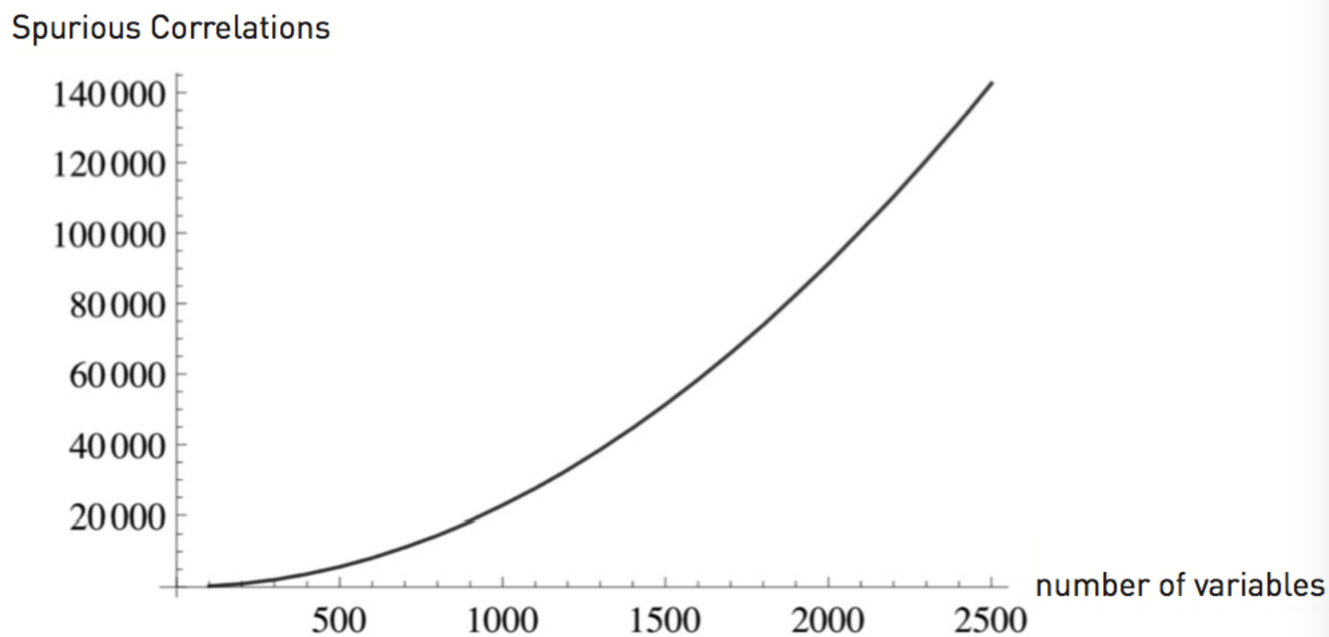


FIGURE 18. The Tragedy of Big Data. The more variables, the more correlations that can show significance in the hands of a "skilled" researcher. Falsity grows faster than information; it is nonlinear (convex) with respect to data.