

塔勒布量化开篇之作《肥尾分布的统计效应》(上)

作者：戴国晨

导读：随机性是我们所生活世界的本质属性之一，身体健康情况，股票市场收益，经济增长……一系列的未来状态都可以用随机变量抽象表示。在以《黑天鹅》为代表的 INCERTO 系列中，塔勒布从哲学的角度对随机性进行了探讨，这部《肥尾分布的统计效应》则更进一层，从数学角度出发，对不同分布的统计性质进行了系统性研究。

面对随机性事件，人们可能面对的风险有如下几种 (1) 不知道分布的类型和性质 (2) 知道分布的类型和性质，错误的估计了参数 (3) 黑天鹅。

对于 (1)，如果错误理解了分布类型，结果很可能是灾难性的。比如估计分布为正态分布（薄尾），极端情况下的波动可能就是 3 个标准差的事件，但是幂律分布（肥尾）同样的概率可以出现 10 个标准差的事件，如果尾部进一步加厚，对数帕累托分布下同样的概率可以出现 100 个“标准差”的事件。不同分布对大数定律和中心极限定理的收敛性质有很大区别，因此体系的第一层是了解不同的分布，明确各自的统计性质。

对于 (2)，如果已经可以判断分布的类型，比如幂律分布（肥尾），但是由于样本量有限，很可能因为真正的尾部没有出现而导致分布参数估计错误，其后果同样是灾难性的。如果想要在小样本下决策，现有的统计性质就需要对“看不见的尾”进行修正，此为体系的第二层。

本书的价值正是建立这样一套体系，通过技术手段解决 (1) 和 (2)，使我们能用数学框架来认识随机性事件，并合理度量其尾部效应。这里塔勒布创造性的提出了中数视角，在小样本下大数定律尚未起作用之时，定量化描述了不同分布呈现出的统计特征。

当然 (3) 是永远无法量化的，不是受困于技术，而是受制于逻辑。火鸡也许可以预测每天早上食物的多寡，但不可能预测感恩节的到来，所有的数学分布和统计估计依然是经验框架下的产物。

认知真相无比之难，经验之外不可知的那一部分总是让人寝食难安，此时应该如何应对呢？塔勒布的哲学拨云见日：反脆弱，不用做预测大师，只需要改变赔付关系即可。在黑天鹅降临之时，具备反脆弱特性的事物不但不会受损，反而还能有巨大收获。

那既然有哲学妙法从根本上化解风险，为什么我们还要挖空心思在处理 (1) (2) 的量化技术中精进？因为反脆弱的道和数学框架的术并不是割裂的，想要实践反脆弱哲学改变赔付关系，首先要明确改变赔付的成本，追根溯源则需要对肥尾分布有定量化的理解。因此在本书的上半部分让我们先聚焦数学，深入理解肥尾效应。

序言

了解世界越是浅薄，决策越是轻易

我们生活的世界极度复杂，充斥着不透明和不确定性，但是人们却对如何理解和应对极端不确定性少有研究。在承受巨大风险的同时，却又笼统的把罕见的事件归为一类，套用“黑天鹅”和“肥尾”来进行解释。这里的谬误在于(1) 黑天鹅代表完全超出经验框架的事件，罕见事件仍有经验可依不属于此范畴。(2) 知晓肥尾概念和真正了解其统计效应之间仍然存在鸿沟，在不同类型的肥尾下人们所面临的冲击有天壤之别，只有真正去了解尾部性质，差异化的采取保护措施才能有效管理风险。

塔勒布的哲学思想包含一整套对尾部风险的认知框架，命名为 INCERTO（意大利语中为不确定性）。从极端风险的道到期权交易的术，当真正踏入市场成为利益攸关的风险接受者时，就会一层层看到其中的玄妙。

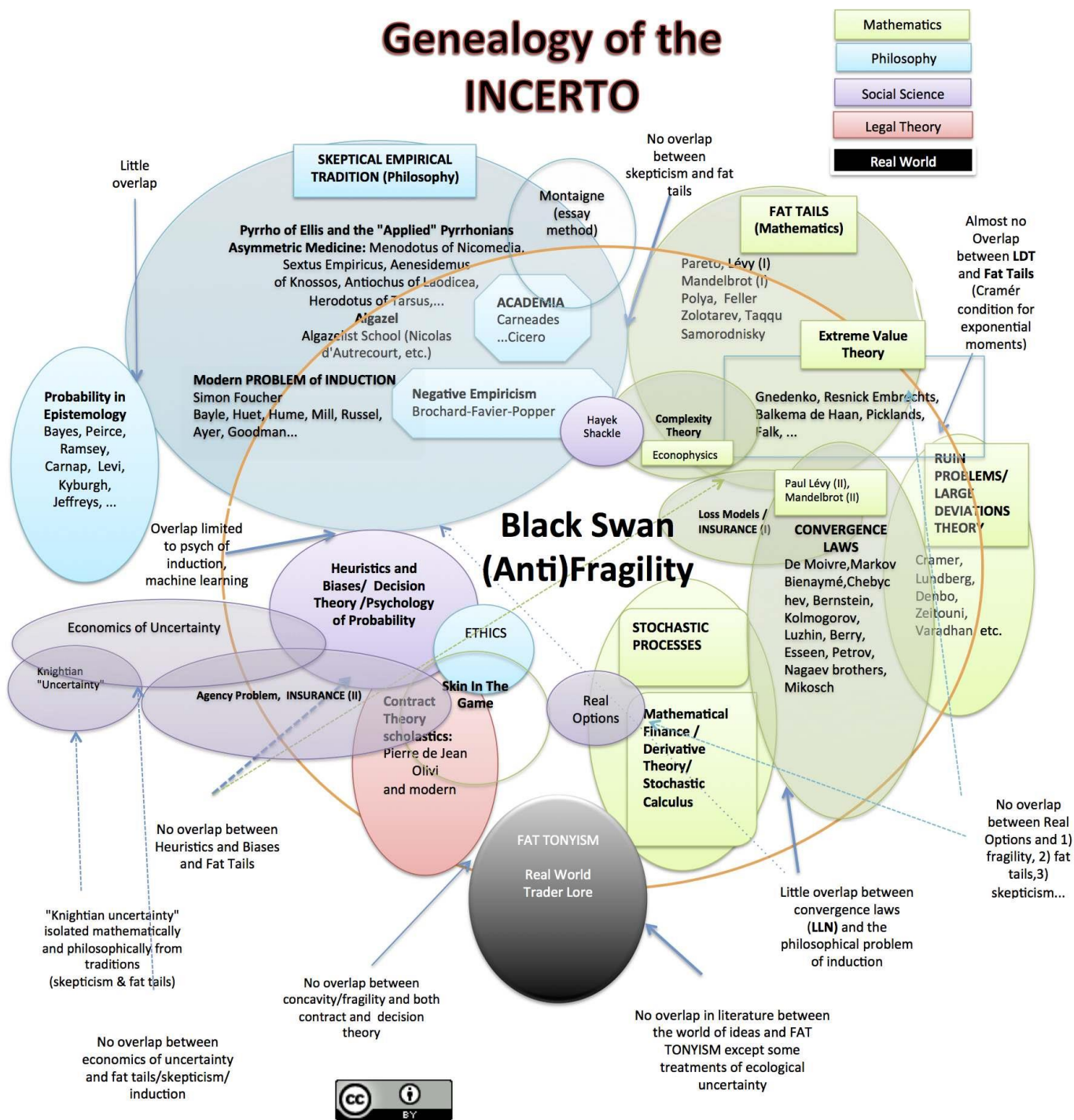
INCERTO 横跨五个与极端尾部风险相关的领域：数学，哲学，社会科学，契约论和决策论，目的在于探究尾部风险与现实世界的联系。其中契约论主要结合期权，决策论讨论如何保护自身甚至反脆弱。该书《肥尾分布的统计效应》为 INCERTO 数量化研究中的一部分，主要集合了一些研究论文和对学界的犀利批判。

INCERTO 系列丛书 (2001-2018)

- 随机漫步的傻瓜 FOOLED BY RANDOMNESS

- 黑天鹅 THE BLACK SWAN
- 普洛克路斯忒斯之床 THE BED OF PROCRUSTES
- 反脆弱 ANTIFRAGILE
- 利益攸关 SKIN IN THE GAME

INCERTO 谱系



词汇表

概率 \mathbb{P} ，一般以 $\mathbb{P}(X > x)$ 表示， X 为随机变量
期望 \mathbb{E} ， $\mathbb{E}(X) = \int_{\Omega} X dP$

方差 V , $V(X) = E[(X - E(X))^2] = E(X^2) - E(X)^2$

平均绝对偏差 M , $M(X) = E[|X - E(X)|]$

PDF 概率密度函数 $\varphi(\cdot)$ 或 $f(\cdot)$

CDF 累积概率密度函数 $F(\cdot)$, 对于定义在 R 上的随机变量, $F(a) = P(X \leq a) = \int_{-\infty}^a f(x)dx$

特征函数 $\chi(t) = E(e^{itX_s})$

\xrightarrow{D} 表示收敛于某分布, $X_n \xrightarrow{D} X$ 代表随机变量 X_n 满足 $\lim_{n \rightarrow \infty} F_n(x) = F(x)$

\xrightarrow{P} 表示收敛于某概率, 对于任意 $\varepsilon > 0$ 满足 $\lim_{n \rightarrow \infty} \Pr(|X_n - X| > \varepsilon) = 0$

$\xrightarrow{a.s.}$ 表示必然收敛, 更强的收敛条件, 满足 $P(\lim_{n \rightarrow \infty} X_n = X) = 1$

α , α 表示尾部衰减指数, 也即当 K 很大时满足 $P_{X>K} = Cx^{-\alpha}$ 的尾部 α , $\alpha \in (0, \infty)$

高斯正态分布 $\mathcal{N}(\mu, \sigma)$

对数正态分布 $\mathcal{L}(\cdot, \cdot)$

稳定分布 $\mathcal{S}(\alpha_s, \beta, \mu, \sigma)$, 尾部参数 $\alpha_s \in (0, 2]$, 对称指数 $\beta \in (-1, 1)$, 中心参数 μ 和离散参数 σ

幂律分布 \mathfrak{P}

亚指数分布 \mathfrak{G}

狄拉克函数 $\delta(\cdot)$

阶跃函数 $\theta(\cdot)$

误差函数 $\text{erf}(\cdot)$, $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$

向量范数 $\|X\|_p \triangleq \left(\frac{1}{n} \sum_{i=1}^n |x_i|^p\right)^{1/p}$

肥尾效应简介

为了衡量我们所面对的尾部, 可以定义平均斯坦和极端斯坦如下:

- **平均斯坦 (薄尾)**
 - 当样本量变大时, 单一观测值无法改变整体统计属性
 - 随机变量连续两次大于 X 的概率高于一次大于 $2X$ 的概率
 - 如果市场在两天内下跌了 30%, 很可能是每天下跌 10%-20%
 - 服从平均斯坦的分布: 正态分布, 身高分布
- **极端斯坦 (肥尾)**
 - 尾部的小概率事件会极大影响整体统计性质
 - 随机变量连续两次大于 X 的概率小于一次大于 $2X$ 的概率
 - 如果市场在两天内下跌了 30%, 很可能是一天-29%, 一天-1%
 - 服从极端斯坦的分布: 财富分布, 灾难分布

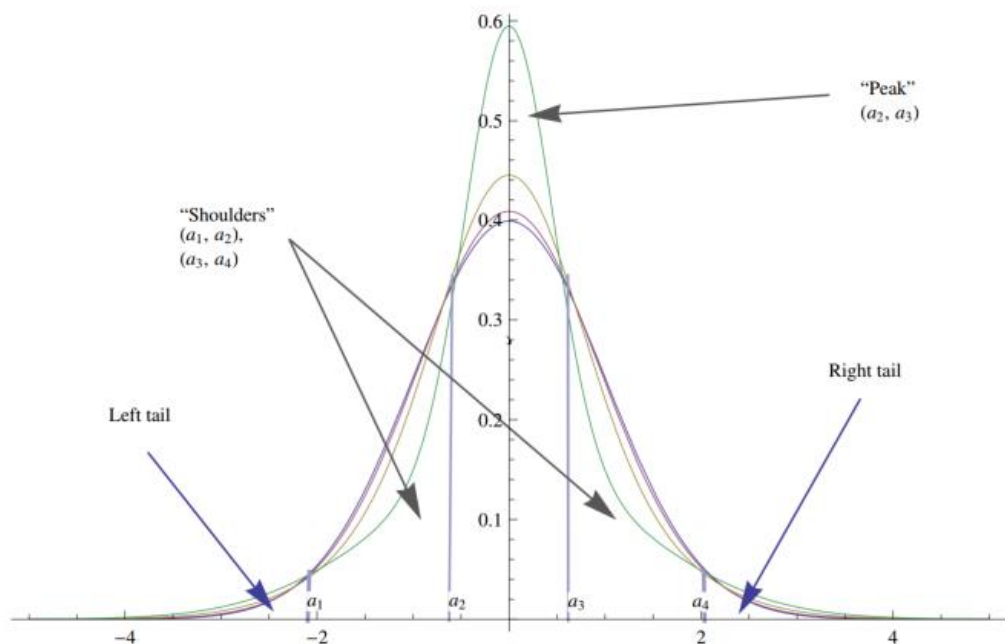
灾难原则

平均斯坦中保险合约才能被合理定价, 永远不要在极端斯坦下卖保险, 因为基于历史赔付定出的保费很可能不够。

平均斯坦中的灾难是“好事”(可控且能帮助人们完善对分布的认知), 极端斯坦中灾难是坏事(可能直接摧毁一切), 举例对比不同环境中的两种灾难事件 (1) 某架飞机失事会导致机组人员丧生和 (2) 某架飞机失事会导致所有坐过飞机的人丧生, 前者可以帮助飞机制造商和保险公司评估飞机失事概率, 并作出改进, 但后者的破坏力量太大, 可能会导致人类灭亡。因此对于前一种事件, 我们的目标是如何降低发生概率。而对后一种事件我们不算概率, 目标

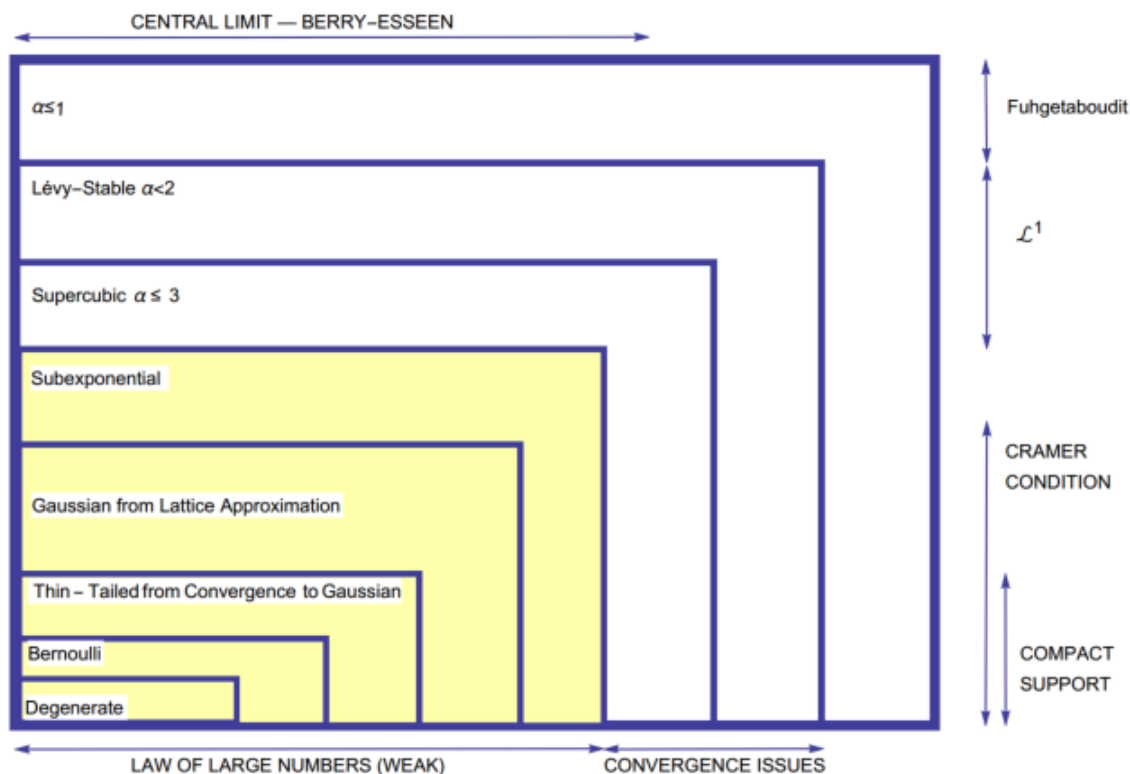
在于如何降低其影响，或是完全杜绝它发生。在金融市场中类似的两种状态是（1）出现危机会导致企业利润降低（2）出现危机会导致企业破产，甚至加剧危机的蔓延，遗憾的是后者非常常见。

肥尾分布在肥尾的同时也带来了尖峰，增大了样本出现在中心区域的概率，在小样本下给人分布更稳定的幻觉。



肥尾程度（不同分布由轻到重）

- 高斯分布
- 普通厚尾分布：峰度大于 3
- 亚指数分布：灾难原则开始生效，各阶矩依然存在
- 幂律分布：帕累托尾部， n 阶矩可能会不存在



上图为分布肥尾程度表，黄色部分的分布满足大数定律，白色部分的分布尾部服从幂律，随着由内向外延伸中心极限定理逐渐失效，最后矩不再收敛。幂律尾部指数用 α 表示：

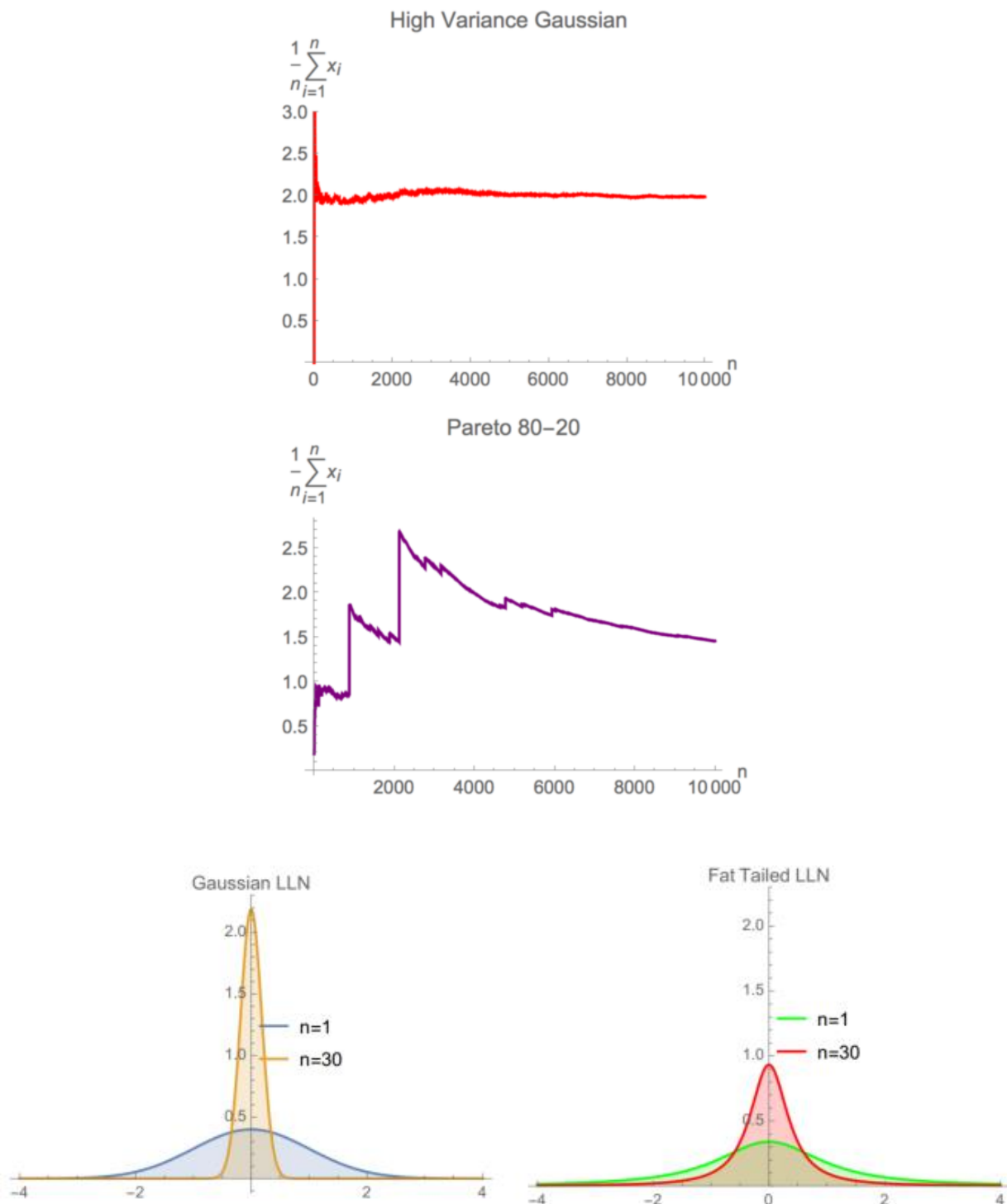
- $\alpha \leq 3$ 超立方分布, 分布只有均值和方差, 高阶矩不存在
- $\alpha \leq 2$ 列维稳定分布, 分布方差不再存在, 只有一阶矩
- $\alpha < 1$ Fuhgetaboudit (别整了), 各阶中心矩不存在, 所有过去的经验无法得出任何有意义的预示

因此在这里可以对肥尾 (Fat Tail) 和厚尾 (Thick Tail) 做一定区分。当人们讨论厚尾分布时, 往往指的只是峰度大于正态的分布, 但是肥尾分布对应的是“极度厚尾分布”, 分布的峰度甚至都不存在, 和分布也不再收敛于正态分布。

在很多学术研究中, 人们明知变量满足肥尾分布的条件下, 依然采用常规方法做统计检验是非常危险的, 极可能因为样本量不足或是统计特征不存在而得出一些“伪结论”。

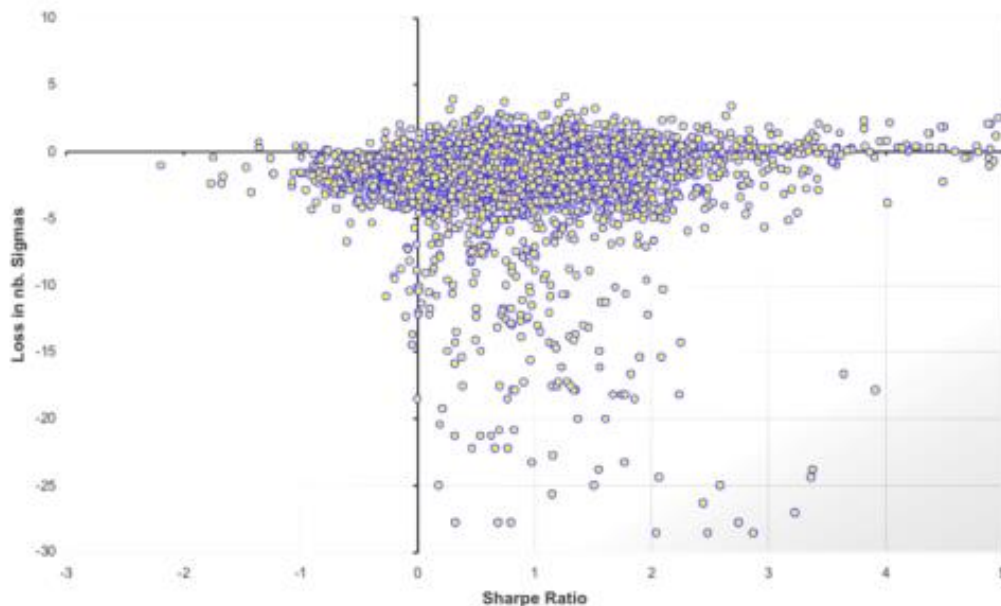
大数定律的收敛情况

极端斯坦 (下图帕累托 80-20 分布, $\alpha=1.13$) 下求样本均值收敛的速度远远慢于平均斯坦 (高斯分布)。从另一个角度讲, 不同分布下做统计归纳的样本量需求可能差异极大, 下图给出了样本量为 30 时, 高斯分布和肥尾分布的均值分布。



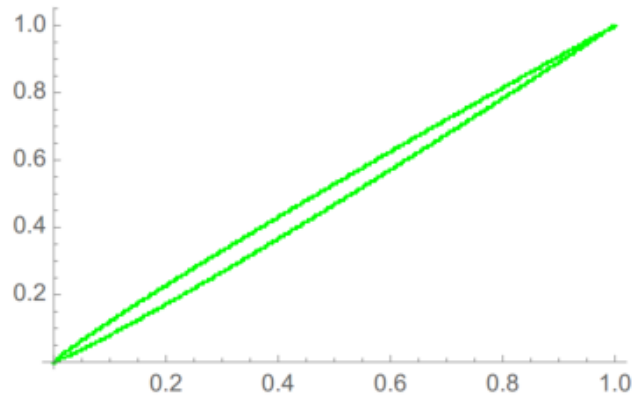
肥尾效应的一些后果

- 现实世界中大数定律可能有效，但是会收敛的太慢。
- 由于样本规模的问题，样本均值一般总是会偏离分布均值，尤其是在不对称分布中
- 在金融领域所有的分布都是肥尾分布，一些统计量如方差和标准差并不可信，由此衍生出的 **beta**，**夏普比率**，**协方差矩阵等金融量也不可信**。下图中的横轴为 08 年金融危机前对冲基金的历史夏普比率，纵轴为金融危机中的基金损失，可以看到历史波动对规避未来损失并无预测作用，在危机中很多基金出现了 10 个“标准差”以外的损失。

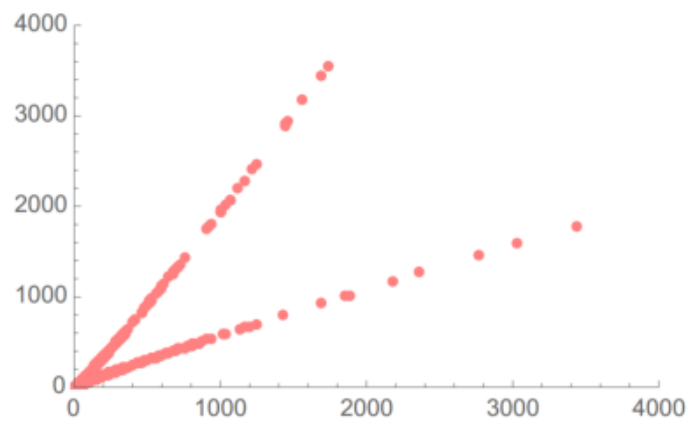


- 稳健统计并不稳健（修正极值会导致信息丧失），**经验分布的经验有限**（最糟情况会超出经验）
- **最小二乘法线性回归会失效**（方差不存在，小样本效应）
- 即使是肥尾分布，对于一些统计量的极大似然估计依然有效，此时通过估计尾部参数 α 反推变量分布，再求解分布均值要比对样本直接求均值准确的多。
- **证实和证伪人们经验的难度比想象的要大。**
- 主成分分析和因子分析会失效，肥尾条件下为得到稳定结果所需的样本量极大。
- 对样本求统计量会失效，分布的高阶矩很可能不存在，导致每个样本矩都不同且不收敛。
- 对极端情况下发生的损失量级很难进行估计。在高斯分布下 CVAR 等极端统计量还有一定意义，而肥尾分布中出现的损失可能会大到超出想象。
- 类似于基尼系数这样的参数不具备可加性。欧洲各个国家基尼系数的加权平均会小于整个欧洲大陆的基尼系数。其他如 1%的人拥有 x%的财富的指标也不具备可加性。
- **动态对冲无法解决期权的真正风险。**
- **重视概率忽视赔付在肥尾条件下会导致更大的问题。**现实世界中的回报并不是基于概率，而是基于赔付，但是人们在直观上更在意概率预测。一个例子是交易员预测市场上涨，但是做空市场——因为上涨空间有限而下跌空间巨大。同样的道理，一个大概率赔钱的策略不一定是糟糕的策略，只要没有破产风险且小概率能获得巨大收益即可，如尾部对冲策略（Universa）；一个胜率 99.99%的策略也不一定是好策略，如果不能完全规避破产风险前期盈利都会归零，如杠杆统计套利（LTCM）。
- **肥尾条件下对实际分布估计的微小偏离都可能带来巨大的赔付偏差。**下图中的 x, y 轴为实际频率和预测频率，向上或向下偏离 $y=x$ 分别为低估和高估。可以看到在预测概率相差不大的情况下（绿线），对 $\alpha=1.15$ 的帕累托分布进行下注，其赔付结果相距甚远（红点）。由于存在非线性关系，**市场参与者的概率预测误差和最终赔付误差完全是两类分布**，概率预测误差是统计量，在 0 到 1 之间，因此误差分布是薄尾的，而赔付的误差分布是肥尾的。

Corresponding MisCalibration in Probability



MisCalibration in Payoff



大数定律

在大数定律下，我们不断加入新的样本后均值会逐渐收敛到理论均值，但是肥尾条件下收敛速度可能会很慢，不同尾部 α 下收敛到同样精度所需的样本数量如下表所示。

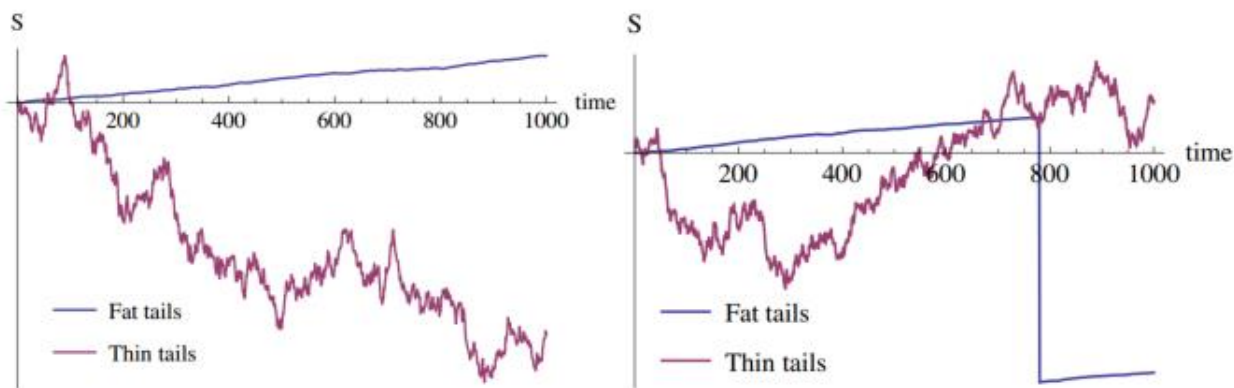
α	n_α Symmetric	$n_\alpha^{\beta=\pm\frac{1}{2}}$ Skewed	$n_\alpha^{\beta=\pm 1}$ One-tailed
1	<i>Fughedaboudit</i>	-	-
$\frac{9}{8}$	6.09×10^{12}	2.8×10^{13}	1.86×10^{14}
$\frac{5}{4}$	574,634	895,952	1.88×10^6
$\frac{11}{8}$	5,027	6,002	8,632
$\frac{3}{2}$	567	613	737
$\frac{13}{8}$	165	171	186
$\frac{7}{4}$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

对于帕累托 80/20 分布，我们需要 10^{11} 量级的数据，均值才能达到正态分布 30 个样本的精度，也就是说在肥尾分布中，对有限的样本求均值来进行统计估计几乎毫无意义。

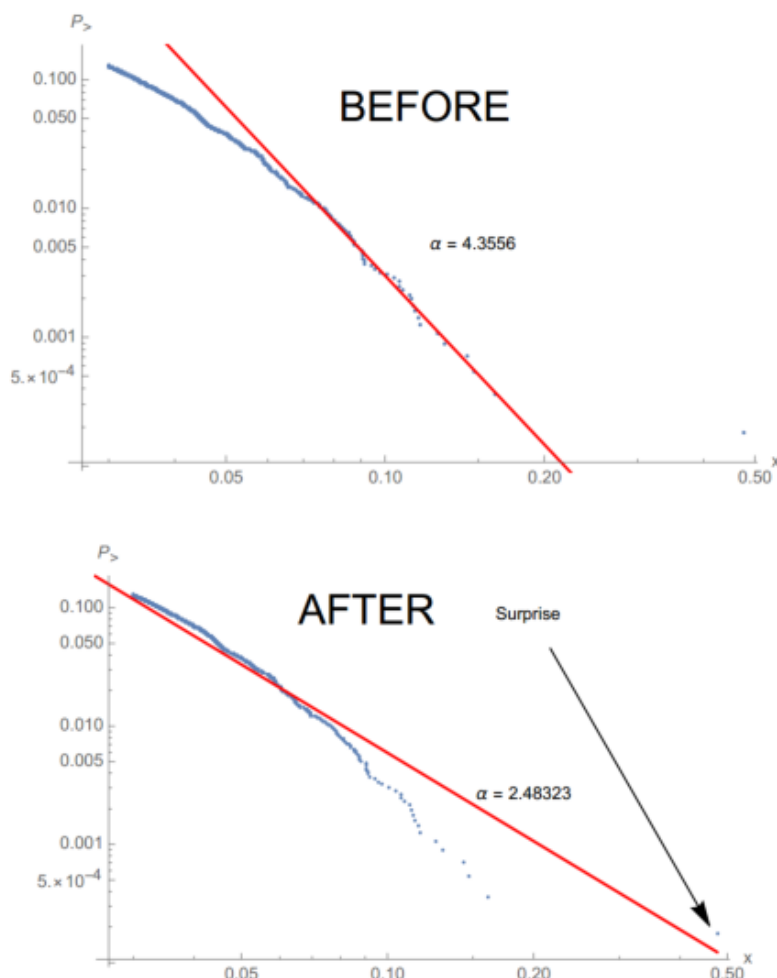
认识论和不对称推理

- 我们无法直接观察到统计分布，只能通过结果进行推测
- 统计分布无法保证结果与其一致
- 需要通过元概率来研究尾部事件

推测统计分布的难度是不对称的。罪犯装普通人比普通人装罪犯要容易得多，**肥尾分布也很容易在小样本下“伪装”成薄尾分布，但薄尾分布几乎不可能呈现出肥尾特征**。如果我们观测到一个 20 倍标准差的事件，可以确定该分布是肥尾分布，但如果没有观测到极端事件却不能确定分布为薄尾。**观测到多少只白天鹅能够估计未来黑天鹅出现的概率？**再大的样本都不能，而一旦出现一只黑天鹅会让认知瞬间出现巨大变动。下图中可以看到肥尾分布在灾难发生前比薄尾分布稳定的多。



下图为阿根廷股市的表现以及 2019.08.12 之前和之后的回报分布，之前多年的平缓上涨隐藏了分布的真实尾部，一个单日暴跌就大幅改变了尾部 α 值（从 4.35 到 2.48）：



帕累托分布（幂律分布）

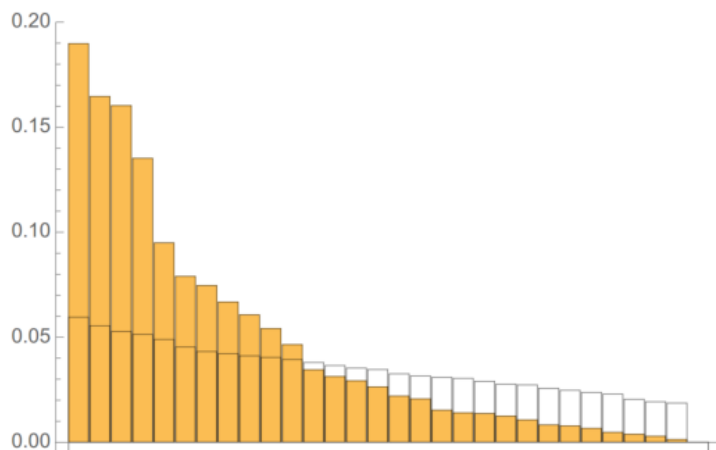
在帕累托分布下对于随机变量 X 和尾部值 x ，我们有：

$$\frac{P(X > kx)}{P(X > x)} = \frac{P(X > knx)}{P(X > nx)}$$

如在财富分布中，拥有 2 亿财富和 1 亿财富的人数比等于拥有 2000 万和 1000 万财富的人数比。帕累托分布没有高阶矩，因此样本统计量很不稳定。下表中展示了不同金融资产过去数十年时间序列的**单日最大峰度贡献**，正态分布下的单日回报峰度贡献理论值在 $0.8\% \pm 0.28\%$ 之间，但是标普 500 的结果是 80%，也即整个分布峰度的 80% 是由一天的收益率贡献的。因此，我们从历史数据出发估计金融市场真实分布的峰度误差将会极大，四阶矩是如此，二阶矩也是如此，所有基于历史波动率和方差的模型在这里都要打一个问号。

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.

因此肥尾分布会极大地增加求解统计量的样本量要求。在主成分分析中该影响也类似，对肥尾分布的随机变量进行主成分分析很可能会找出“伪主成分”。下图为独立正态分布随机变量在 100 和 1000 样本量下的主成分分析结果（1000 个样本下主成分间的差异已经很小，较容易识别变量独立性），

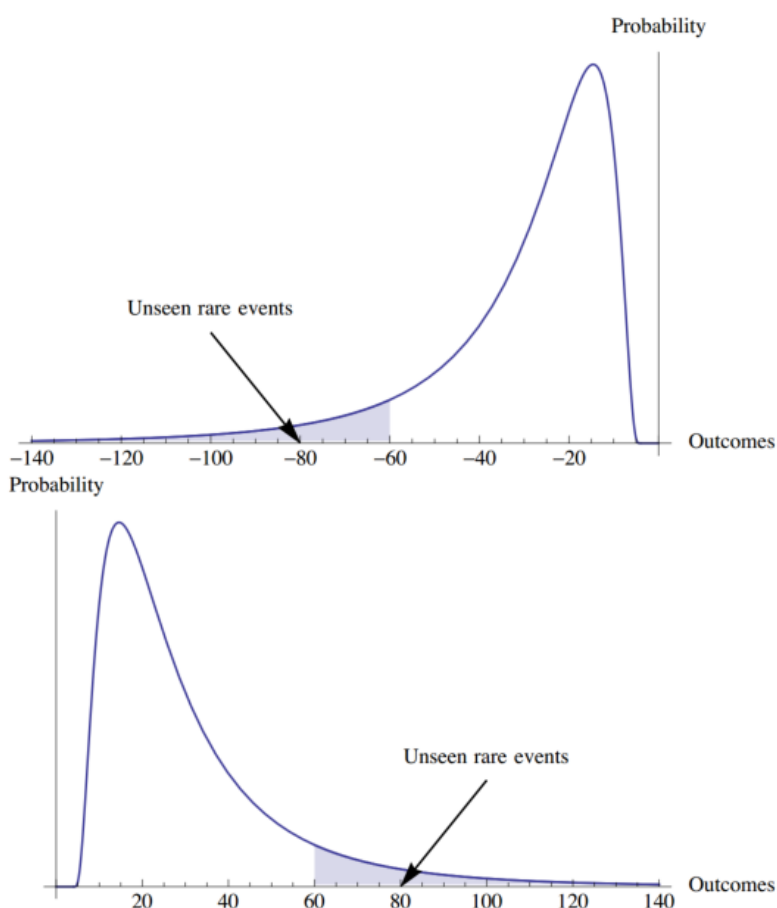


而下图是在 100 和 1000 样本量下，独立肥尾变量（尾部 $\alpha=1.5$ ）的主成分分析结果，可以看到 1000 样本量下依然存在着显著的“伪主成分”。



统计量估计

对于肥尾分布的样本来说，由于（1）肥尾事件很少发生（2）分布不一定对称，简单求样本平均估计均值的方法不再有效。一个比较好的办法是先推测分布，再从分布反求均值（极大似然估计）。比如对于帕累托分布，98%的样本会小于均值。如下图所示，阴影部分的尾部值在小样本中很可能不会出现，但是会极大影响整体均值。



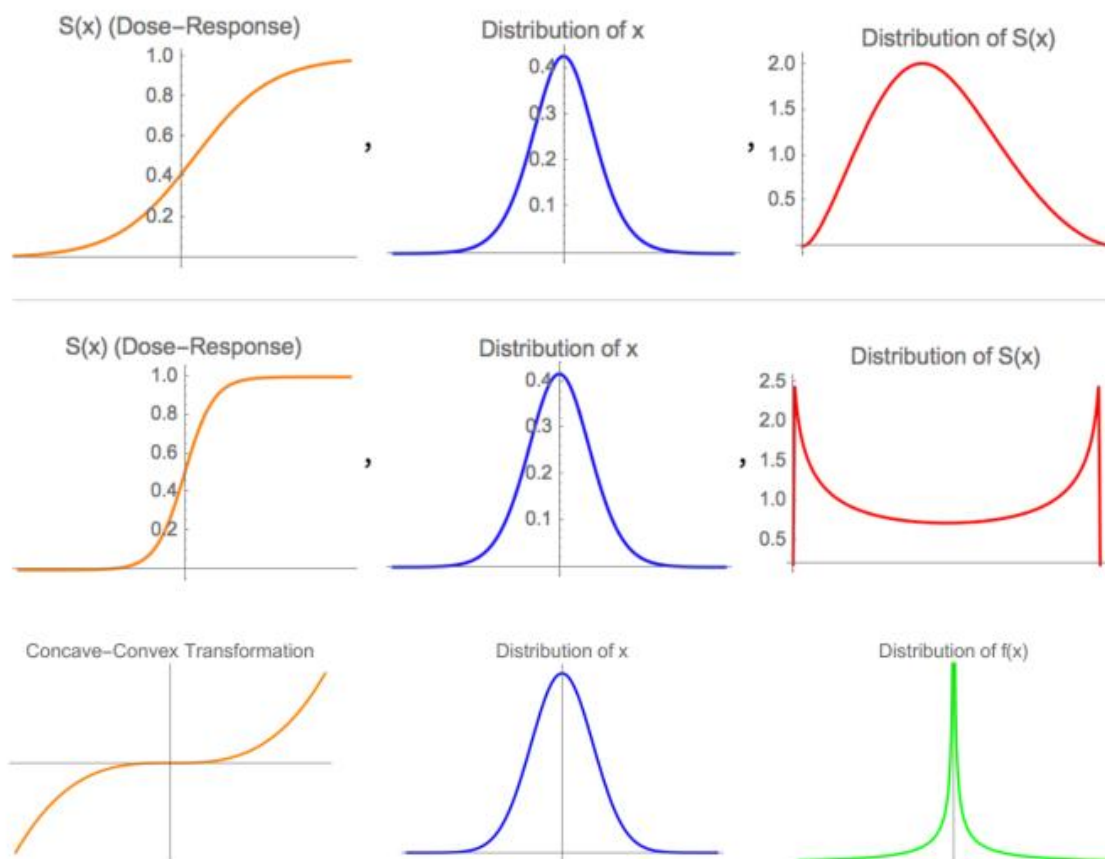
非线性影响，脆弱与反脆弱

如果把金融产品的回报看做是随机变量 X ，该回报施加到个人身上的影响就是 $F(X)$ ， X 和 $F(X)$ 大概率不是线性的，让潜在的非线性结果对我们有利，或者至少不对我们有害是反脆弱的必要条件。比如下列几种情况：

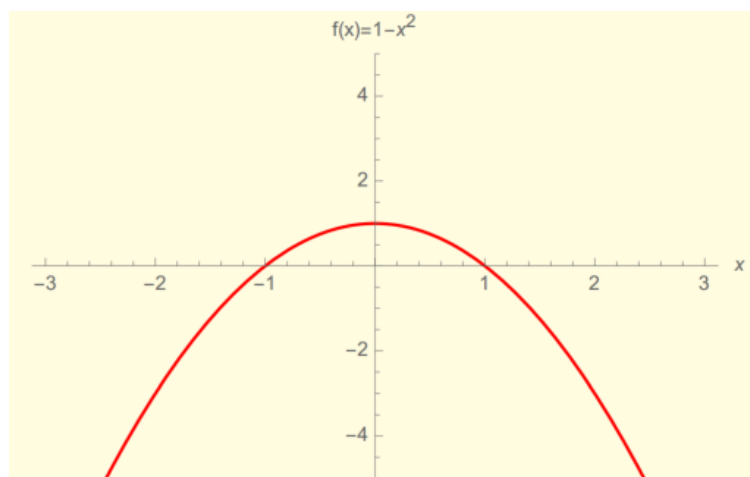
- X 是失业率， $F_1(X)$ 是对货币基金组织的影响， $F_2(X)$ 是对退休老年人的影响（薄尾）
- X 是股票价格， $F(X)$ 是你买的期权赔付（有利肥尾）
- X 是财富变化， $F(X)$ 是财富变化对你行为的影响（薄尾）

在真实世界中我们往往无法控制或是估计随机变量 X 的表现，但是我们可以在一定程度上选择 $F(X)$ ，有时甚至可以完全决定 $F(X)$ 。其要点在于选择函数 $F(X)$ 尾部的凹凸性，下图为同样的高斯分布经过尾部凹凸性不同的 $F(X)$ 变换后

成为薄尾分布 (1) (2) 和肥尾分布 (3)



在金融市场中的一个例子是方差互换合约 (Variance Swap, 波动率衍生品之一), 押注市场的波动率稳定就是一个脆弱状态, 每次的赔付可以简化为 $F(X) = 1 - X^2$, 如下图所示:



假设 7 次连续押注下的两种结果:

- X 为 1, 1, 1, 1, 1, 0, 0, X 均值为 0.71, 赔付 $F(X)$ 为 2 (盈利)
- X 为 0, 0, 0, 0, 0, 0, 5, X 均值也为 0.71, 但赔付 $F(X)$ 为 -18 (破产)

从期望的角度上来说, 两者都押注了正确的方向, 但是由于薄尾/肥尾分布假设和非线性尾部赔付, 导致了交易结果的巨大差异。

破产和路径依赖

因为存在破产事件, 财富这一变量在时间维度上存在路径依赖。对于赌场来说, 每个人在场内下注是独立的, 大数定律最终会起作用, 而对于赌客来说, 财富会依赖过去的路径, 哪怕期望为正, 一旦破产游戏就截断了。人们往往只重视灾难单次发生的小概率, 但忽略了在重复游戏中风险会不断累积, 只有在将破产风险严格控制为零, 才能让自身在长期得以存活。如巴菲特左轮手枪的例子, 子弹只一发, 在头上开一枪可以获得 100 万美元, 是否值得参与这样的游戏呢? 可能有人会愿意为了钱冒险, 但是对于成功的风险管理者来说, 只要有归零风险就不应该参与游戏。

决策方法论

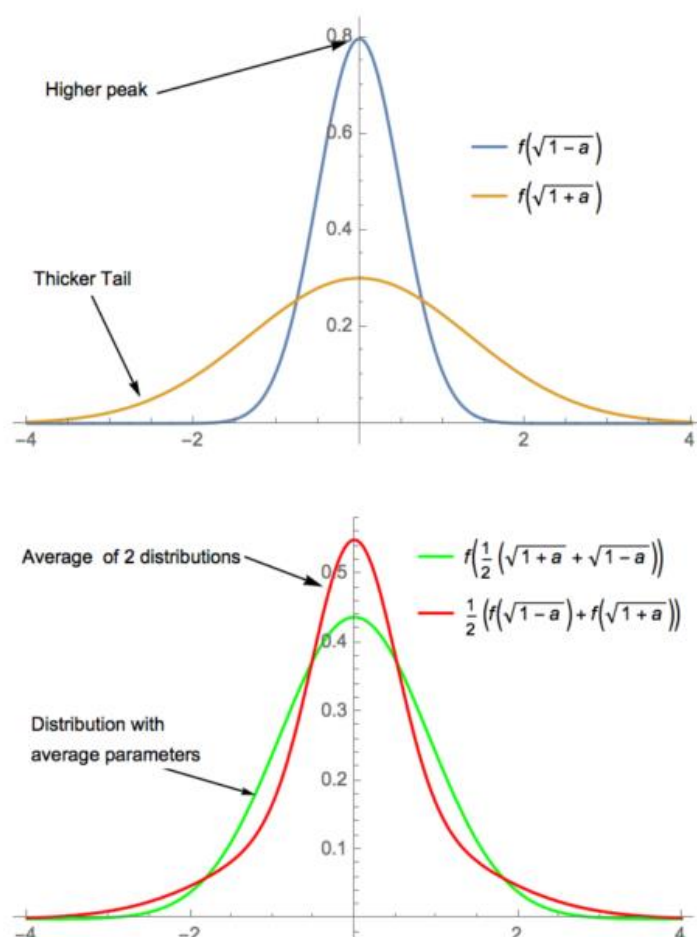
决策的三要素 (1) 观察我们所处的环境是平均斯坦还是极端斯坦 (2) 观察我们面对的事件是否存在路径依赖 (3) 考量该事件对我们的影响，脆弱或是反脆弱。

肥尾分布简介

任何的肥尾分布都可以被表示为一个方差随机的高斯分布，比如时间序列分析中的异方差误差项就属于肥尾分布。假设我们将一个分布的 $E(x^2)$ 固定，增大 $E(x^4)$ 项使其肥尾，相当于将方差项“随机化”了，比如假设随机变量 X 满足：

$$\begin{cases} \mathcal{N}(0, \sigma\sqrt{1-a}) & p = \frac{1}{2} & 0 \leq a < 1 \\ \mathcal{N}(0, \sigma\sqrt{1+a}) & p = \frac{1}{2} & 0 \leq a < 1 \end{cases}$$

从特征函数可以求得该分布的方差为 σ^2 ，四阶矩为 $3(a^2 + 1)\sigma^4 > 3$ ，因此为厚尾分布。下图中直观展示了其厚尾特性，其中红线分布（两分布平均）的尾部厚于绿线分布（方差平均）的尾部：

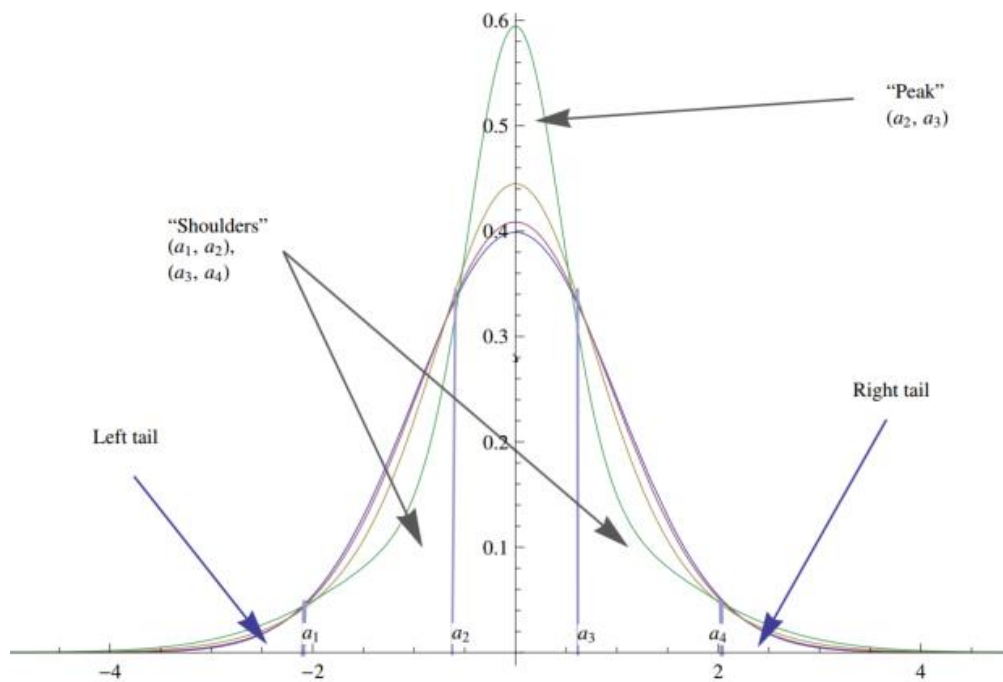


肥尾的隧穿效应

对于任意一个分布来说，对尾部加厚必然会导致尖峰，而加厚部分的概率也自然来源于非尾部部分，导致分布的隧穿。对于随机变量 X 来说，概率密度函数为 $p(x)$ ，如果该分布是双尾分布，对于方差随机化的分布：

$$p^\delta(x) \triangleq \frac{1}{2}(p(x, \sigma + \delta) + p(x, \sigma - \delta))$$

存在一个 δ 扰动相关的尖峰使 $p^\delta(x) \geq p(x), x \in (a_2, a_3)$ ，同时也存在隧穿尾部 $p^\delta(x) \geq p(x), x \in (-\infty, a_1)(a_4, +\infty)$ 。



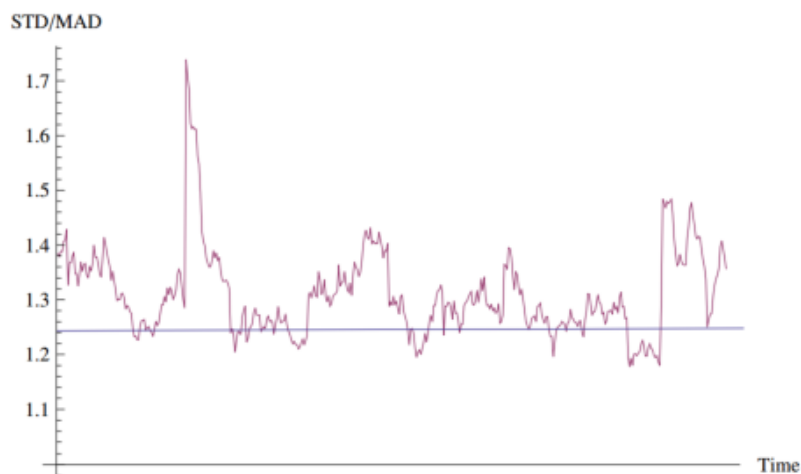
这里可以通过 $\{x: \frac{\partial^2 p(x)}{\partial \sigma^2} |_{\sigma} = 0\}$,求解 a_1, a_2, a_3, a_4 , 其中 a_1, a_4 代表了肥尾的起点。

分布	$\{a_1, a_2, a_3, a_4\}$
高斯分布	$\{\mu - \sqrt{\frac{1}{2}(5 + \sqrt{17})}\sigma, \mu - \sqrt{\frac{1}{2}(5 - \sqrt{17})}\sigma, \mu + \sqrt{\frac{1}{2}(5 - \sqrt{17})}\sigma, \mu + \sqrt{\frac{1}{2}(5 + \sqrt{17})}\sigma\}$
对数正态分布	$\{e^{\frac{1}{2}(2\mu - \sqrt{2}\sqrt{5 - \sqrt{17}}\sigma)}, e^{\frac{1}{2}(2\mu - \sqrt{2}\sqrt{5 + \sqrt{17}}\sigma)}, e^{\frac{1}{2}(2\mu + \sqrt{2}\sqrt{5 - \sqrt{17}}\sigma)}, e^{\frac{1}{2}(2\mu + \sqrt{2}\sqrt{5 + \sqrt{17}}\sigma)}, \}$
学生 T 分布	$\{-\frac{\sqrt{5\alpha - \sqrt{(\alpha+1)(17\alpha+1)+1}}}{\sqrt{2(\alpha-1)}}S, -\frac{\sqrt{5\alpha + \sqrt{(\alpha+1)(17\alpha+1)+1}}}{\sqrt{2(\alpha-1)}}S, \frac{\sqrt{5\alpha - \sqrt{(\alpha+1)(17\alpha+1)+1}}}{\sqrt{2(\alpha-1)}}S, \frac{\sqrt{5\alpha + \sqrt{(\alpha+1)(17\alpha+1)+1}}}{\sqrt{2(\alpha-1)}}S\}$

一般来说分布的肥尾 (a_1, a_4) 开始于 2-3 个标准差处, 在实际数据中由于小样本效应, 尾部开始的位置可能更大。

平均差和标准差

和标准差 $STD = \sqrt{\frac{1}{n} \sum x_i^2}$ 相比, 让我们对数据离散程度更直观的是平均差 $MAD = \frac{1}{n} \sum |x_i|$ 。假设问每日温度的波动有多大, 比如 $\{-2, 3, 4, -5, 0, 1\}$, 直接用绝对值的均值比平方加起来再开方要直观得多。标准差会比平均差要大一些, 肥尾分布的标准差会比平均差更大, 对于高斯分布 $\frac{STD}{MAD} = \sqrt{\frac{\pi}{2}} \approx 1.253$, 如果用这个标准检验 SP500 历史数据, 会发现日回报率基本都在肥尾范围。

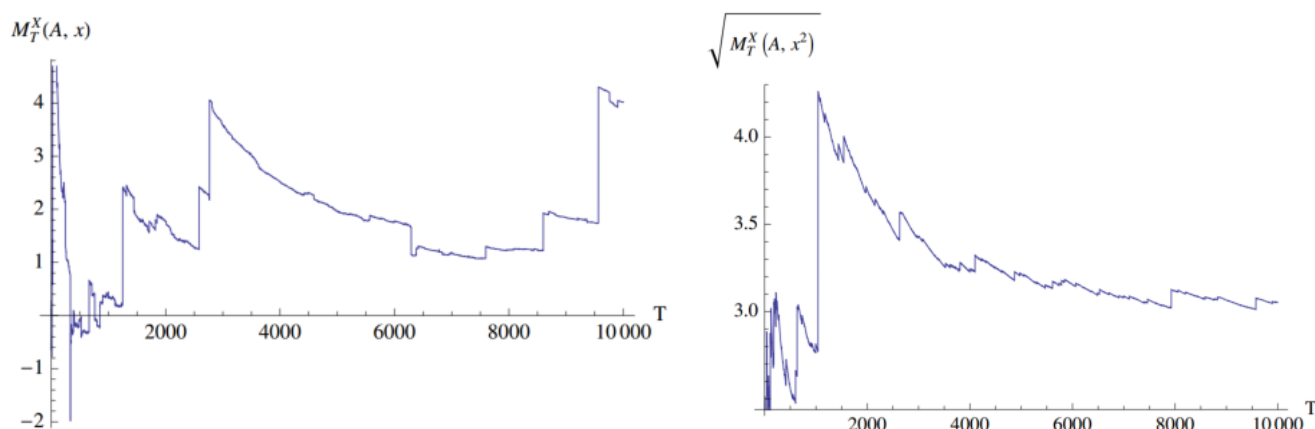


平均差相对标准差有如下几个优点：

- 平均差在衡量样本离散程度上更加稳定，标准差相当于用自身数据进行了一次加权
- 在金融应用（期权定价）中平均差的意义更大
- 对于部分分布方差无限，不存在标准差，但是存在平均差

无限矩

对于每个有限量样本来说，我们总是可以求得有限矩，比如样本均值和方差，但是对于分布本身来说矩并不一定有限，比如柯西分布的均值和方差并不存在。在这种情况下，每个不同的样本都会得到不同的矩，且增大样本并不会收敛。如下图所示，均值和方差在下次跃变前会出现“伪收敛”。



亚指数分布和幂律分布

在分布由极度薄尾（伯努利）到极度肥尾的过程中，我们通过矩的收敛情况将肥尾程度做如下分类。

类别	描述
极度薄尾	离散紧支分布（伯努利分布，二项分布）
薄尾	高斯分布（由薄尾分布的中心极限定理逼近）
常规薄尾	高斯分布（由自然现象分布逼近）
厚尾	峰度稍高于高斯分布，但是在求和以后迅速逼近高斯分布
亚指数	对数正态分布
超立方 α	对于 $t > 3$ 克萊姆条件失效, $\int e^{-tx} d(Fx) = +\infty$
方差无限	列维稳定分布 $\alpha < 2$, $\int e^{-tx} d(Fx) = +\infty$
一阶矩无限	别整了没啥好研究的

之前我们所提到的平均斯坦和极端斯坦的界限就出现在亚指数地带，假定如下条件：

$$\lim_{x \rightarrow +\infty} \frac{1 - F^{*2}(X)}{1 - F(X)} = 2$$

其中 $\bar{F}(X) = 1 - F(X)$ 为生存函数, $X = X_1, X_2 \dots X_n$ 为独立同分布随机变量, F^{*2} 是 $X_1 + X_2$ 的累积分布, 也就是说 $X_1 + X_2$ 超过一个给定值 x 两倍于每一个单独 X 超过给定 x 的概率, 因此对于 X 趋于无穷时, 如果两者和大于 x , 大概率会来自单变量的贡献。该结论可以推广到 n 维, 满足条件时随机变量的和主要由最大的样本贡献:

$$\lim_{x \rightarrow +\infty} \frac{P(S_n > x)}{P(X > x)} = n \rightarrow \lim_{x \rightarrow +\infty} \frac{P(S_n > x)}{P(M_n > x)} = 1$$

其中 $S_n = \sum_{i=1}^n x_i$, $M_n = \max_{1 \leq i \leq n} x_i$

亚指数的尾部比指数分布的尾部更厚, 而且亚指数分布无法通过矩母函数求矩, 对于所有的 $\epsilon > 0$:

$$\int_0^\infty e^{\epsilon x} dF(x) = +\infty$$

临界概率分布

根据尾部的极限特征, 对于随机变量 X 我们可以进行如下定义:

$$\lim_{K \rightarrow \infty} \frac{1}{K} \mathbb{E}(X|_{X>K}) = \lambda$$

- 如果 $\lambda = 1$, X 属于薄尾分布类别
- 如果 $\lambda > 1$, X 属于无特征尺度的肥尾分布类别
- 如果 $\lim_{K \rightarrow \infty} \mathbb{E}(X|_{X>K}) - K = \mu$, $\mu > 0$ 则 X 属于临界指数分布类

幂律分布（帕累托分布）和尾部 α

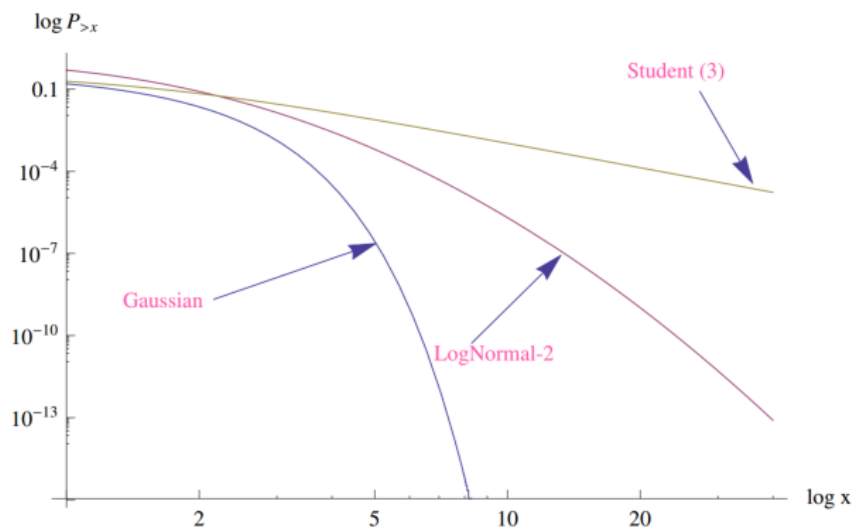
对于矩有限的分布来说, 从尾部特性来看 $\frac{P_{X>nK}}{P_{X>K}}$ 同时依赖于 n 和 K , 而对于尺度无限的分布来说, 当 K 很大时, $\frac{P_{X>nK}}{P_{X>K}}$ 只依赖于 n , 最终呈现出帕累托特征, 也即 $P_{X>x} = Cx^{-\alpha}$, 或者可以更严谨的定义为 $P_{X>x} = L(x)x^{-\alpha}$ 且 $L(x)$ 满足:

$$\lim_{x \rightarrow +\infty} \frac{L(tx)}{L(x)} = 1$$

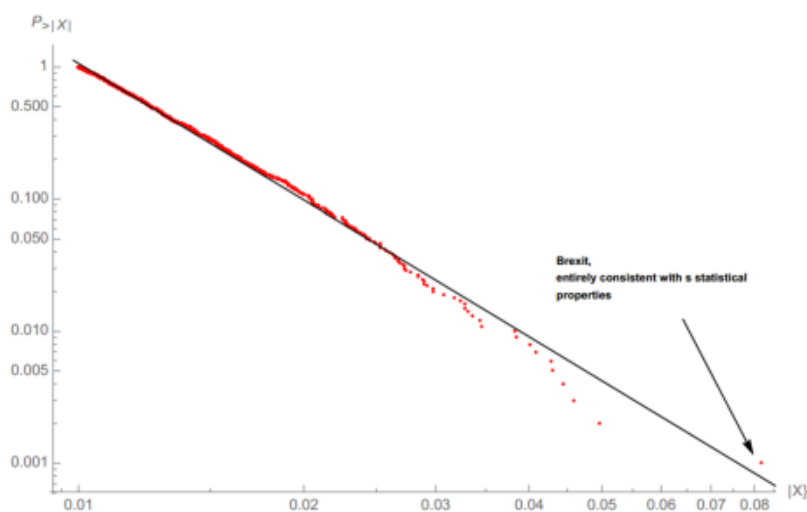
下表展示了高斯分布, 学生 T 分布和帕累托分布的尾部特性, 可以看到高斯分布的尾掉的很快, **如果我们观测到一个很大的随机变量, 几乎不可能再观测到该值 2 倍的结果**。而对于学生 T 和帕累托分布则不然, 翻倍的尾部值永远会以固定概率出现, 不断打破人们对尾部风险的认知。

k	$\mathbb{P}(X > k)^{-1}$ (Gaussian)	$\frac{\mathbb{P}(X>k)}{\mathbb{P}(X>2\ k)}$ (Gaussian)	$\mathbb{P}(X > k)^{-1}$ Student(3)	$\frac{\mathbb{P}(X>k)}{\mathbb{P}(X>2\ k)}$ Student (3)	$\mathbb{P}(X > k)^{-1}$ Pareto(2)	$\frac{\mathbb{P}(X>k)}{\mathbb{P}(X>2\ k)}$ Pareto (2)
2	44	720	14.4	4.9	8	4
4	31600.	5.1×10^{10}	71.4	6.8	64	4
6	1.01×10^9	5.5×10^{23}	216	7.4	216	4
8	1.61×10^{15}	9×10^{41}	491	7.6	512	4
10	1.31×10^{23}	9×10^{65}	940	7.7	1000	4
12	5.63×10^{32}	fughedaboudit	1610	7.8	1730	4
14	1.28×10^{44}	fughedaboudit	2530	7.8	2740	4
16	1.57×10^{57}	fughedaboudit	3770	7.9	4100	4
18	1.03×10^{72}	fughedaboudit	5350	7.9	5830	4
20	3.63×10^{88}	fughedaboudit	7320	7.9	8000	4

由帕累托尾部我们可以得到 $\frac{\log P_{X>x}}{\log x}$ 收敛于尾部常数 $-\alpha$, 也即 log-log 图的尾部斜率。对高斯分布我们可以做类似的求解, 结果是尾部常数趋于负无穷, 在图上可以直观的比较尾部下降速度。



由此出发我们可以通过 log-log 图估计金融市场分布的尾部情况，此处要注意灰天鹅（对分布估计的不足），比如下图右下角的点为英国脱欧事件，一个极值点会极大地改变尾部 α 值。



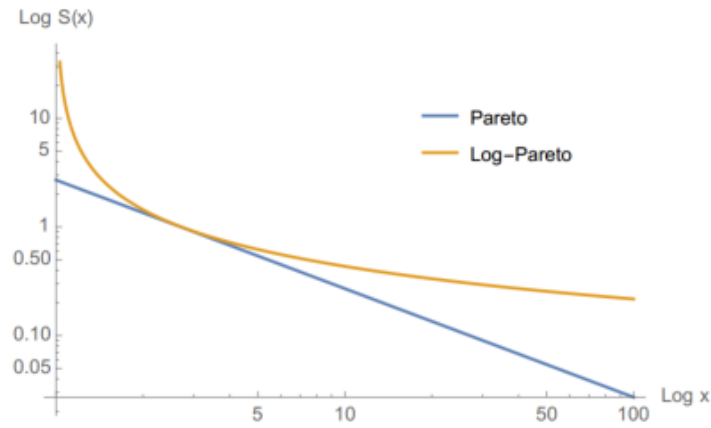
幂律分布的特性：

- 不同随机变量混杂：幂律尾+薄尾=幂律尾，几个分布组合的尾部取决于最厚的尾
- 如果随机变量 X 尾部指数为 α ，则 X^p 的尾部指数为 $\frac{\alpha}{p}$

超级肥尾：对数帕累托分布

对于普通的幂律尾部，随着 x 不断增大，尾部概率在 log-log 图上最终会以直线的形式下降（斜率为 $-\alpha$ ），而对数帕累托尾部更加可怕，在 log-log 图中极限尾部斜率为 0（不再下降），对数帕累托分布概率密度函数如下：

$$f(x) = \frac{\alpha L^\alpha \log^{-\alpha-1}(x)}{x}$$



对于这种超级肥尾分布，自然所有矩都不存在。

多变量肥尾分布

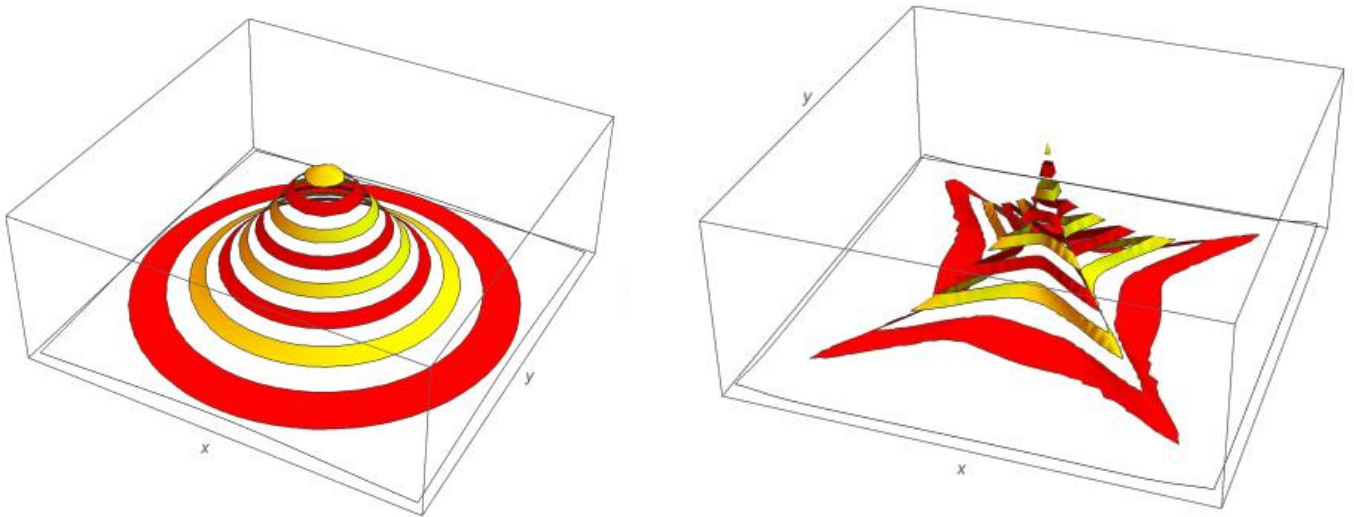
对于多元分布来说，可以用协方差矩阵 Σ 将概率密度函数写成向量化的形式，比如对于 n 维高斯分布，其概率密度函数定义为：

$$f(\vec{x}) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1}(\vec{x} - \vec{\mu})\right)$$

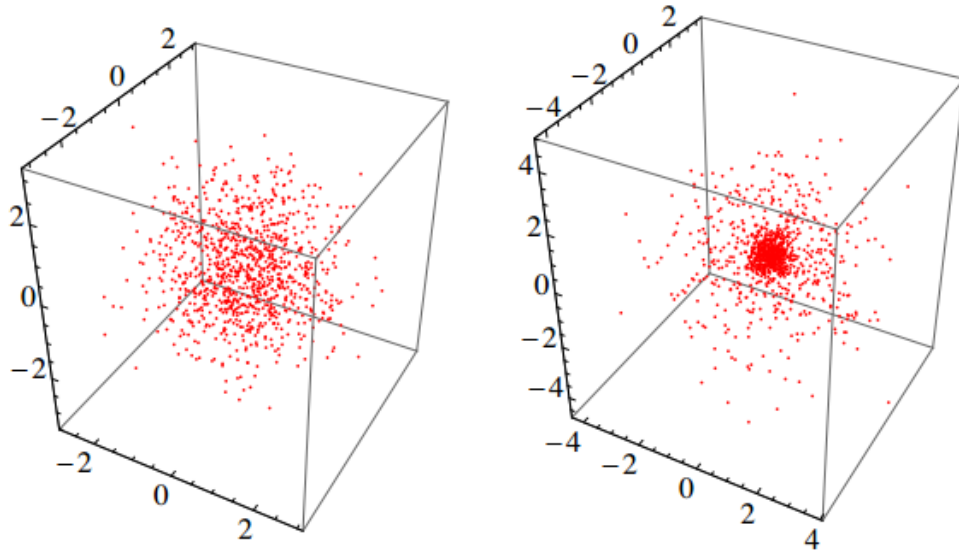
其中 Σ 为对称正定矩阵，和一维高斯分布类似，我们可以通过一定的概率使协方差矩阵漂移来获得多变量肥尾分布，即：

$$\Sigma = \begin{cases} \Sigma_0(1+a) & p \\ \Sigma_0(1-a) & 1-p \end{cases}$$

上述分布属于椭圆分布，椭圆分布要求分布的特征函数能被表示成 $\exp(it'\mu)\Psi(t\Sigma t')$ 的形式，其中 Σ 为非负定矩阵，直观上看就会呈现出“椭圆”的特征，下图显示了学生 t 椭圆分布（左）和随机相关性下的非椭圆分布（右）。



在肥尾分布方面一些单变量的结论仍然成立，比如分布的隧穿，肥尾分布的“尾部”和“中间”会更密集，而“肩部”的概率降低，如下图中随机变量分别满足三维薄尾分布（左）和肥尾分布（右）。



肥尾多变量之间的相关性比较特殊，数学上判断两随机变量 X 和 Y 独立的条件为：

$$\frac{f(x, y)}{f(x)f(y)} = 1$$

对于高斯分布我们由相关系数 $\rho = 0$ 可以推出变量独立不相关，但是对于肥尾分布则不然，比如 $\rho = 0$ 时二元柯西分布的概率密度函数如下：

$$f_0(x, y) = \frac{1}{2\pi(x^2 + y^2 + 1)^{3/2}}$$

而两柯西分布的变量乘积为：

$$f(x)f(y) = \frac{1}{\pi^2(x^2 + 1)(y^2 + 1)}$$

并不满足独立性条件，因此两肥尾变量独立（共同信息为 0）时线性相关系数 ρ 不为 0。

另外，对于方差无限的随机变量来说，在多变量联合分布中，其相关系数依然存在于 $[-1, 1]$ 之间，不过收敛很慢，对样本进行估计会呈现出极大误差。

线性回归中的肥尾误差

在很多经济学研究中都会用到线性回归，并通过一个相对较高的 R^2 来表明模型具备解释性。但是在肥尾分布的世界中，如果我们用肥尾随机变量相对薄尾随机变量做回归，在小样本下很容易得到高 R^2 ，带来统计上的失真。

比如对于随机变量 $Y = aX + b + \epsilon$ ，其中 X 满足高斯分布 $\mathcal{N}(0, 1)$ ，而 ϵ 满足幂律分布且 $\mathbb{E}(\epsilon) = 0$ ， $\mathbb{E}(\epsilon^2) < \infty$ ，此时我们可以求 R^2 如下：

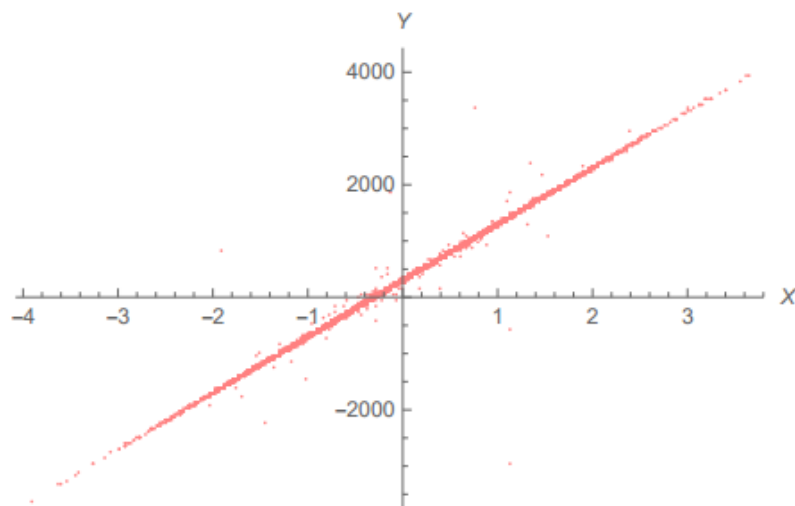
$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^n (y_i - (ax_i + b + \epsilon_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

当样本量 n 很大时：

$$R^2 = \frac{a^2}{a^2 + \mathbb{E}(\epsilon_i^2)} + O\left(\frac{1}{n^2}\right)$$

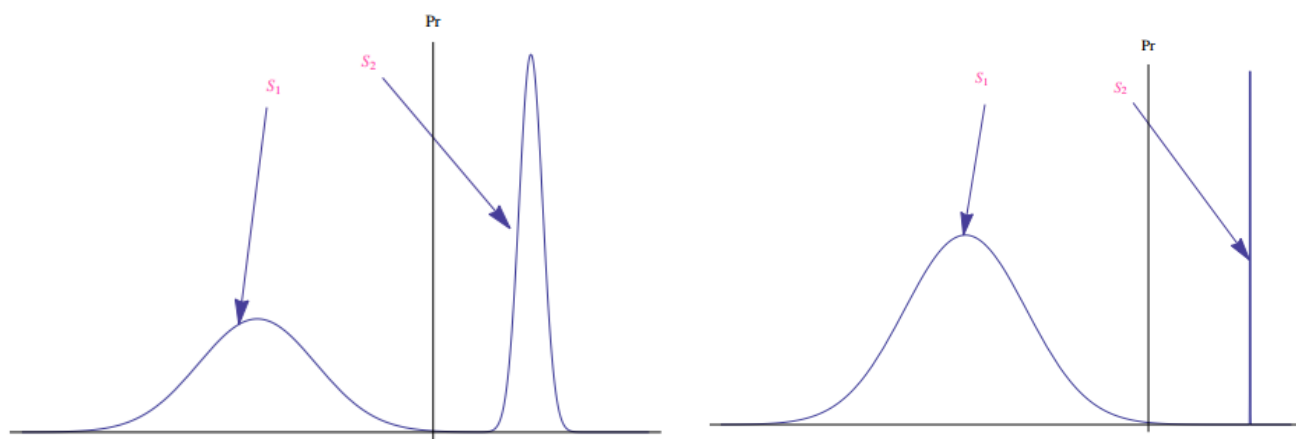
也即样本 R^2 依赖于误差项 ϵ 的样本方差，如果肥尾变量 ϵ 的方差无限大，则 $R^2 = 0$

而 ϵ 即使方差有限，其收敛速度较慢也会导致小样本下 $\mathbb{E}(\epsilon_i^2)$ 被低估，真正的尾部尚未出现从而使 R^2 偏高。下面是 ϵ 在柯西分布下的随机变量散点图，小样本下呈现出非常强的解释性（ $R^2 = 0.985$ ），而实际 $R^2 = 0$



特殊肥尾分布

我们之前讨论的都是单一模型的分布，并通过“随机化”方差实现了肥尾。如果把均值也随机化会如何呢？也即将完全不同的概率分布组合起来。在这样的多模型分布中，不同状态之间是阶跃式变化的，比如一个咖啡杯的状态分布，只有完好和破碎两种，并不存在“稍有破碎”这样的中间状态。在现实世界中，我们经常会看到类似的状态模型，比如战争-和平模型，两种状态下经济变量的分布截然不同（左），或者是债券的赔付模型，固定的利息支付可以看作是狄拉克函数，而一旦违约则赔付不确定（右）。



对于多模型分布我们同样可以定量求解合并分布的峰度，假设随机变量满足：

$$X \sim \begin{cases} \mathcal{N}(\mu_1, \sigma_1^2) & p = \frac{1}{2} \\ \mathcal{N}(\mu_2, \sigma_2^2) & p = \frac{1}{2} \end{cases}$$

组合分布的峰度为：

$$\text{峰度} = 3 - \frac{2((\mu_1 - \mu_2)^4 - 6(\sigma_1^2 - \sigma_2^2)^2)}{((\mu_1 - \mu_2)^2 + 2(\sigma_1^2 + \sigma_2^2))^2}$$

可以看到均值的拉开会降低峰度，而方差的拉开会提高峰度，两者互相竞争，此时通过峰度大小定义肥尾分布的传统方法不再有效。

双状态模型的例子有：

- 货币汇率（固定汇率制）：维系或者脱钩，一旦脱钩就会发生大幅波动
- 企业并购：成功或者失败
- 职业选择：不同路径选择下的不同结果
- 冲突：战争与和平

中数定律

在数学上，大数定律和中心极限定理共同解决的问题是：无穷个随机变量相加时其和的渐进性质如何。由此出发，一个更具现实意义的问题是“中数定律”：如果随机变量数目不到无穷，其和的性质会如何？

大数定律

大数定律描述的是当 n 趋于无穷时， n 个独立同分布随机变量取平均值收敛于分布均值。这里收敛的模式有两种，一种是以分布形式收敛，对应弱大数定律，另一种是必然收敛，对应强大数定律。

弱大数定律：对于独立同分布随机变量 $X_1, X_2, X_3 \dots$ ，分布均值 $E(X_n) = \mu$ ，当 $n \rightarrow \infty$ 时，

$$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n) \xrightarrow{P} \mu$$

也就是说对于任意正数 $\varepsilon > 0$ ：

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) = 0$$

强大数定律：同样对于独立同分布随机变量 $X_1, X_2, X_3 \dots$ ，分布均值 $E(X_n) = \mu$ ，当 $n \rightarrow \infty$ 时，

$$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n) \xrightarrow{a.s.} \mu$$

也即：

$$\mathbb{P}(\lim_{n \rightarrow \infty} \bar{X}_n = \mu) = 1$$

中心极限定理

按照传统中心极限定理，独立同分布随机变量的和趋向正态分布，这里的分布需要满足方差有限的条件，也即对于 $E(X_i) = \mu, V(X_i) = \sigma^2 < \infty$,

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

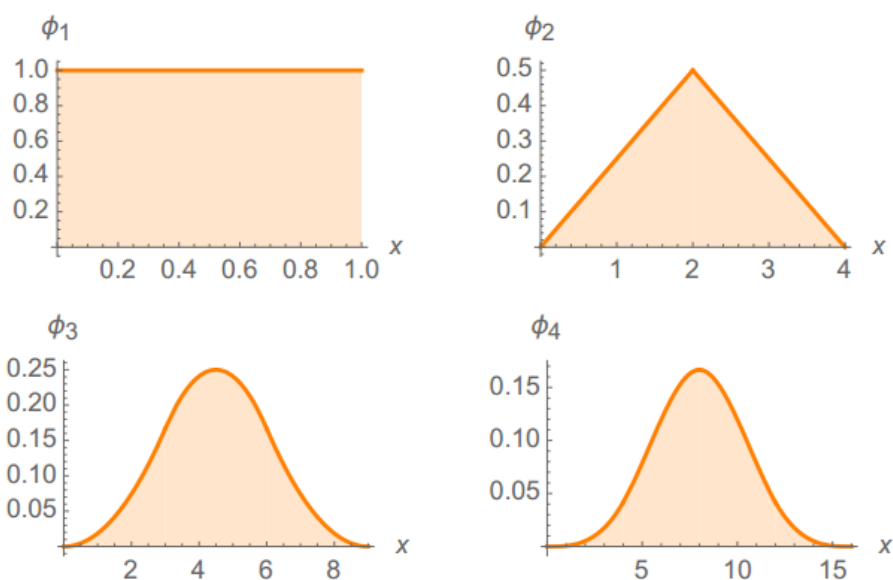
如果放宽方差有限的条件，广义中心极限定理描述独立同分布随机变量的和分布满足稳定分布。

对于大数定律和中心极限定理，我们知道无论如何当 $n \rightarrow \infty$ 时会收敛，但是当 $n < \infty$ 时不同分布收敛的速度可能有很大差异，因此需要对收敛速度进一步研究。

快速收敛：均匀分布

对于均匀分布来说，概率密度函数为一条直线，下方总面积为 1。求和时收敛极快，只需要 3 个变量就会出现钟型曲线，假设均匀分布 $X \in [L, H]$ ，其和分布为：

$$\phi_n(x) = \sum_{k=0}^n (-1)^k \binom{n}{k} \left(\frac{x-L}{H-L} - k \right)^{n-1} \text{sgn}\left(\frac{x-L}{H-L} - k\right)$$

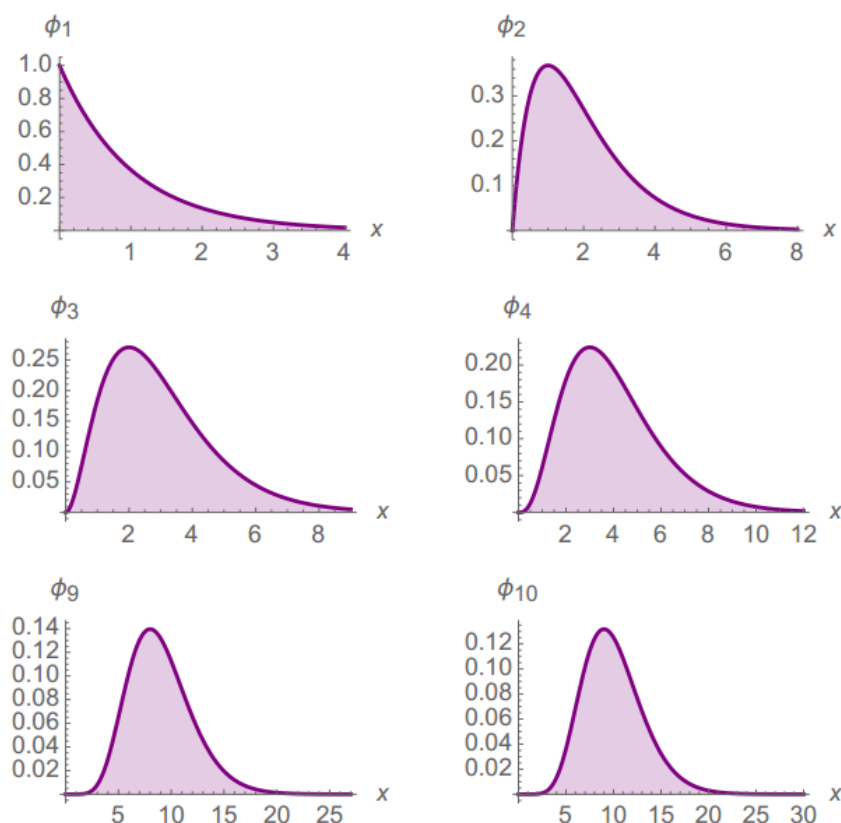


中速收敛：指数分布

指数分布概率密度函数满足 $\phi_1(x) = \lambda e^{-\lambda x}$ ，属于不对称分布，尾部以指数速度衰减，和分布为：

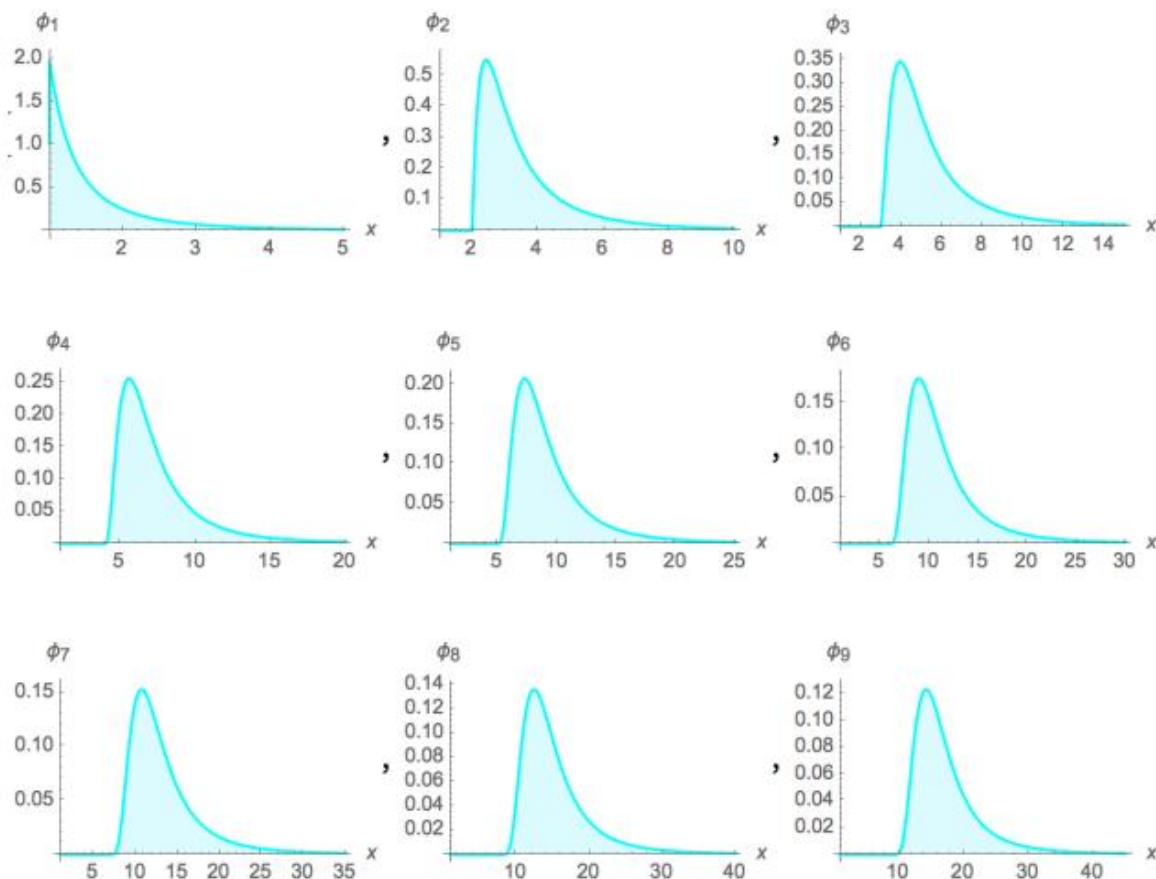
$$\phi_n(x) = \left(\frac{1}{\lambda}\right)^n \frac{x^{n-1} e^{-\lambda x}}{\Gamma(n)}$$

在 $n \rightarrow \infty$ 时该分布趋近于均值 $\frac{n}{\lambda}$ ，方差 $\frac{n}{\lambda^2}$ 的正态分布，但是速度就会慢很多，如下图所示：



缓慢收敛：帕累托分布

帕累托分布尾部满足 $\mathbb{P}(X > x) = cx^{-\alpha_p}$ ，对于 $\alpha_p \leq 2$ ，其和分布收敛于稳定分布，而对于 $\alpha_p > 2$ ，和分布收敛于正态分布。虽然 $\alpha_p > 2$ 最终会逐渐逼近正态分布，但是收敛速度会慢上很多，下图中可以看到多变量加和后分布还是存在不小的偏度，



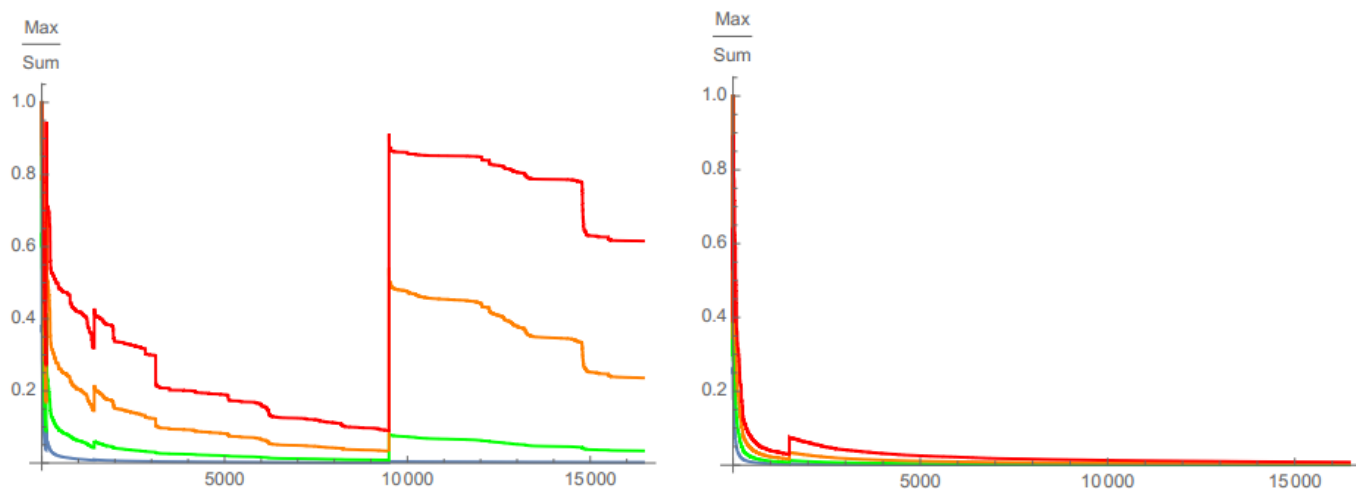
高阶矩的大数定律

测试分布肥尾的一个办法是通过大数定律观察高阶矩的收敛性。如果分布的高阶矩存在，通过不断增加数据量，可以看到样本矩逐渐向理论值收敛。如果矩不存在则每组样本的值都会不同，增加数据量依然会看到值出现跃变。另外我们还可通过 MS 图（Max to Sum）来观察，即样本内矩极大值比所有矩之和，定义如下：

$$Max\ to\ Sum_n^p = \frac{M_n^p}{S_n^p} = \frac{Max(X_1^p, \dots, X_n^p)}{\sum_{i=1}^n X_i^p}$$

其中 X_n^p 为 n 样本的 p 阶矩，直观来说当 $n \rightarrow \infty$ 时该比值会趋近于 0。

下图为 $p=1,2,3,4$ 时的 MS 图，左边基于 SP500 近 60 年数据（截止于 2018），右边基于正态分布数据。可以看到 SP500 的回报属于明显的肥尾分布，其高阶矩极大程度依赖出现于金融危机的尾部值。



肥尾指标 Kappa

之前我们讨论了不同程度的肥尾分布，而决策者通常面临的一个问题是：如果已有一个样本集，比如金融市场收益率，看上去是肥尾分布，我们需要多大的数据量才能得出有意义的统计结论？如果是高斯分布，我们用少量的

数据就能有效的估计分布参数，但是如果是一个尾部 $\alpha = 2.1$ 的帕累托分布呢，又或者 2 个自由度的学生 T 分布，这时多少数据量合适？

一般人们在比较肥尾程度时会采用高阶矩，比如峰度（如果存在的话），用其超过某个值的概率作为指标（尺度调整后），但是对于峰度不存在的极度厚尾则无能为力。这里提出一个新方法：**考量在给定分布下额外数据能在多大程度上增加均值观测的稳定性？**由此产生肥尾指标 κ (*Kappa*)。 κ 可以有如下用途：

- 对不同分布的和分布进行比较，或是对同一分布不同数量的和分布比较
- 衡量一个分布在列维 α 稳定分布中的位置（正态分布是其中的一个特例）
- 定量比较大数定律下不同分布收敛的速度
- 比较两个完全不同分布的肥尾程度（只要分布一阶矩存在）
- 在研究中帮助推出蒙特卡洛模拟所需次数

κ 定义如下：

对于独立同分布的随机变量 X_1, X_2, \dots, X_n ，假设均值有限 $E(X) < \infty$ ，和为 $S_n = \sum X_i$ ， n 个变量和的平均差为：

$$M(n) = E(|S_n - E(S_n)|)$$

定义 κ 为从 n_0 到 n 的平均差收敛速率：

$$\kappa_{n_0,n} = \min \left\{ \kappa_{n_0,n} : \frac{M(n)}{M(n_0)} = \left(\frac{n}{n_0} \right)^{\frac{1}{2-\kappa_{n_0,n}}}, n_0 = 1, 2 \dots \right\}$$

其中 $n > n_0 \geq 1$ ，因此：

$$\kappa_{n_0,n} = 2 - \frac{\log(n) - \log(n_0)}{\log\left(\frac{M(n)}{M(n_0)}\right)}$$

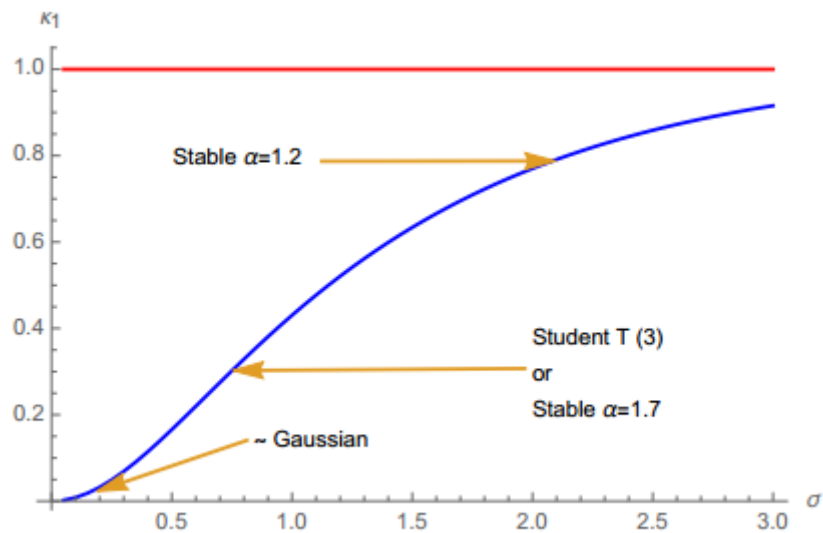
对于 $n = n_0 + 1$ ，我们直接用 κ_{n_0} 表示。下表为不同尾部 α 下，帕累托分布和学生分布的 κ 比较：

α	Pareto	Pareto	Pareto	Student	Student	Student
	κ_1	$\kappa_{1,30}$	$\kappa_{1,100}$	κ_1	$\kappa_{1,30}$	$\kappa_{1,100}$
1.25	0.829	0.787	0.771	0.792	0.765	0.756
1.5	0.724	0.65	0.631	0.647	0.609	0.587
1.75	0.65	0.556	0.53	0.543	0.483	0.451
2.	0.594	0.484	0.449	0.465	0.387	0.352
2.25	0.551	0.431	0.388	0.406	0.316	0.282
2.5	0.517	0.386	0.341	0.359	0.256	0.227
2.75	0.488	0.356	0.307	0.321	0.224	0.189
3.	0.465	0.3246	0.281	0.29	0.191	0.159
3.25	0.445	0.305	0.258	0.265	0.167	0.138
3.5	0.428	0.284	0.235	0.243	0.149	0.121
3.75	0.413	0.263	0.222	0.225	0.13	0.10
4.	0.4	0.2532	0.211	0.209	0.126	0.093

对于列维稳定分布 $X \sim S(\alpha, \beta, \mu, \sigma)$ 来说，

$$\kappa_{n_0,n} = \begin{cases} 2 - \alpha & \alpha \in [1,2) \\ 0 & \alpha \geq 2 \end{cases}$$

因此正态分布 κ 为 0，如果分布的 κ 值超过 0.15，常规的正态近似就不适用了，而 κ 越高，所谓的置信区间就越不可靠。 κ 同时也体现了大数定律的收敛速率， κ 越接近 0，分布的中心极限行为越接近正态分布。

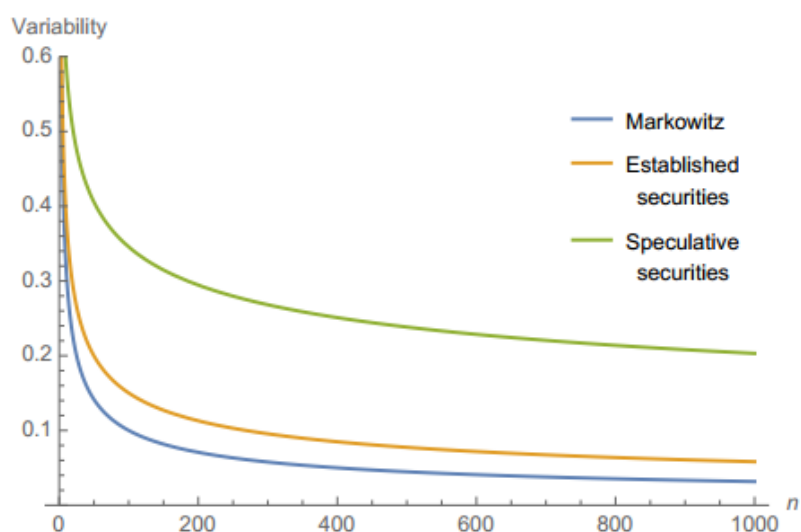


由此出发，假设正态分布随机变量 $X_{g,1}, X_{g,2}, \dots, X_{g,n_g}$ ，有未知的某分布随机变量 $X_{v,1}, X_{v,2}, \dots, X_{v,n_v}$ ，并将两个分布的尺度调为一致，也即 $M^v(1) = M^g(1) = \sqrt{\frac{2}{\pi}}\sigma$ ，这样我们可以比较两者在同样平均差下 n_v 和 n_g 的关系：

$$n_v = n_g^{-\frac{1}{\kappa_1 n_g - 1}} \approx n_g^{-\frac{1}{\kappa_1 - 1}}$$

在这个式子中我们可以得到均值收敛到相同精度所需的样本数，比如对于 3 个自由度的学生 T 分布，两者比值为 4，也就是说同样得到有效均值，正态分布需要 30 个数据点，而这样的学生 T 分布需要 120 个数据点。而对于 $\alpha = 3$ 的单尾帕累托分布，则需要 543 个数据点。如果是 $\alpha \approx 1.14$ 的帕累托 80/20 分布，需要 $> 10^9$ 个数据点，也就是说，拟合这样的肥尾分布有限的几千上万个数据远远不够。

在金融领域，肥尾分布由于高 κ 带来大数定律和中心极限定理收敛较慢，会影响很多的现有的学术结论。比如对于 **Markowitz 投资组合理论**，分散化带来的好处就被大大降低了。投资组合理论中假设各个证券的收益率符合正态分布，在收益率互相独立时建立投资组合，组合波动率会随着证券加入飞快的降低。但是在收益率肥尾分布的假设下，根本无法用原来数量的证券来控制组合波动，投资者需要远远比理论更多的证券量。因此在实际市场中很多组合的波动远远高于理论，其关系如下图所示，为控制组合波动率实际需要的证券数量（绿线）远远高于投资者以为安全的数量（橙线），和 Markowitz 预测所需证券数量（蓝线）。



分布极端值和隐藏风险

在研究某分布的过程中，因为能得到的样本量总是有限，在尾部总会有一段分布遗失。这样一来过去的极值并不是一个预测未来极值的好指标。我们可以研究随机变量极大值的分布，来进一步理解肥尾分布样本下隐含的风险。

假设对于独立同分布随机变量 X_1, X_2, \dots, X_n , 随机变量 $X_{max} = \max(X_1, X_2, \dots, X_n)$ 满足:

$$\mathbb{P}(X_{max} \leq x) = \mathbb{P}(X_1 \leq x) \mathbb{P}(X_2 \leq x) \dots \mathbb{P}(X_n \leq x) = F(x)^n$$

对于帕累托分布,

$$F(x) = 1 - \left(\frac{L}{x}\right)^\alpha$$

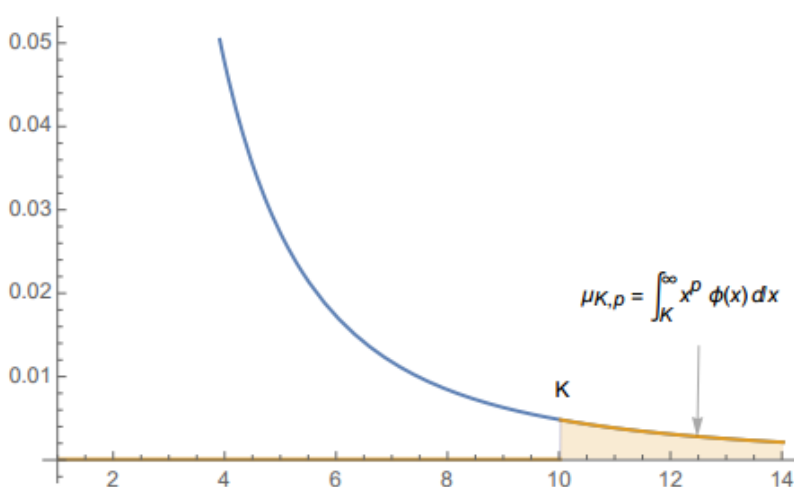
因此累计最大值的概率密度函数为:

$$\psi(x) = \frac{\alpha n \left(\frac{L}{x}\right)^\alpha \left(1 - \left(\frac{L}{x}\right)^\alpha\right)^{n-1}}{x}$$

当 x 很大时该分布和 Frechét 分布的尾部性质类似, 尾部指数也为 α , 也即和原分布 X 尾部相同。

看不见的尾

对于以上最大值随机变量 $K_n = X_{max}$, 我们可以将分布的矩分为两块, 看得见的尾部 $\mu_{0,p}$ (到 K 为止) 和看不见的尾部 $\mu_{K,p}$ (从 K 开始), 如下图所示。

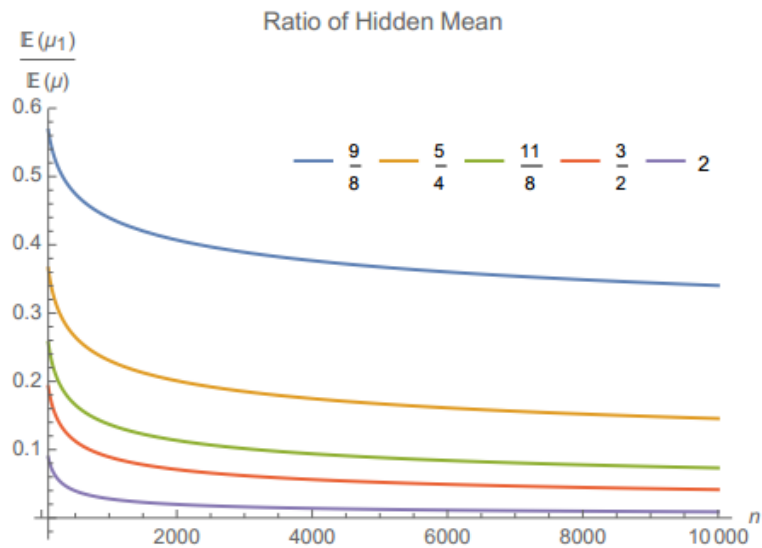


$$E(X^p) = \underbrace{\int_L^{K_n} x^p \phi(x) dx}_{\mu_{0,p}} + \underbrace{\int_{K_n}^\infty x^p \phi(x) dx}_{\mu_{K,p}}$$

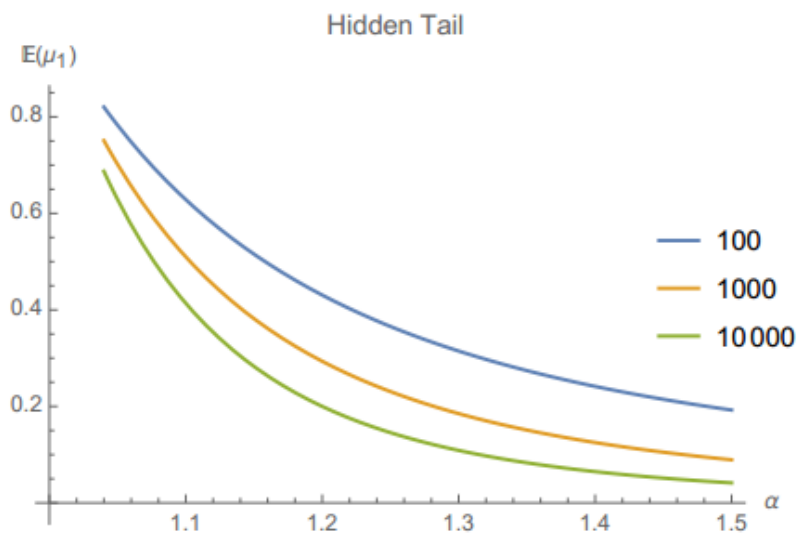
在帕累托分布下,

$$E(\mu_{K,p}) = \frac{\alpha(L^p - L^\alpha K^{p-\alpha})}{\alpha - p}$$

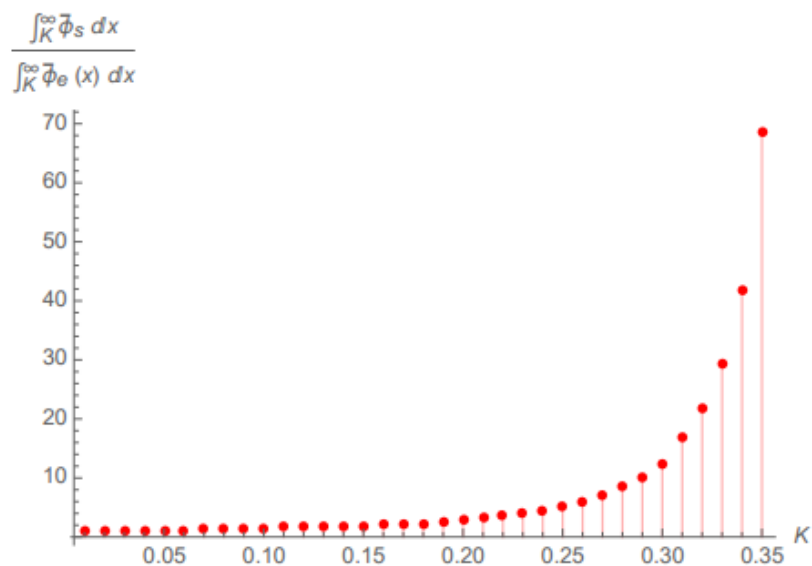
我们可以在 n 增大的过程中, 比较 $\mu_{K,p}$ 和整个分布矩的比值。随着样本的扩大, 越来越多的尾部值出现, 看不见的尾比重会逐渐降低。但是对于 α 较小的帕累托分布来说其下降速度并不快, 下图为均值中“看不见的尾” $\mu_{K,1}$ 占比, 在 10000 次事件后肥尾分布的均值依然可能被低估 30% 以上。



该比值随着 α 降低而升高，也就是说对越肥尾的分布，我们在有限样本下获得的信息偏差越大，“看不见的尾”甚至可以完全起主导作用。



从这个角度出发，**金融市场中看不见的尾会导致我们系统性的低估风险**。一个典型的例子是 CVAR 这样的尾部风险指标（条件风险值）。实践中的 CVAR 往往由历史数据得出，但是历史数据会大大低估尾部均值，哪怕有“大量的”历史数据。对于美国股票市场来说，随着风险条件的提高，基于历史的 CVAR 会被低估 5-70 倍。



大偏差理论

大偏差理论着眼于分布均值大于给定值的概率 $\mathbb{P}(M_N > x)$ ，这里通过伯努利分布的大偏差我们可以看到薄尾分布极强的收敛性。

对于伯努利分布 $\{0,1\}$ ， $\mathbb{P}(X = 1) = p$ ， S_n 为变量和 $\sum X$ ，期望 $\mathbb{E}(S_n) = np = \mu$ ，由切尔诺夫界对于任意的 $\delta > 0$ ：

$$\mathbb{P}(S \geq (1 + \delta)\mu) \leq \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu$$

在 $0 < \delta \leq 1$ 时：

$$\mathbb{P}(S \geq (1 + \delta)\mu) \leq 2e^{-\frac{\mu\delta^2}{3}}$$

因此假设一个抛硬币游戏，正反各 50% 概率，正面盈利 1，反面为 0，玩 n 次以后盈利均值为 $\frac{n}{2}$ ，可以看到实际盈

利高于 50% 盈利均值的概率为 $e^{-\frac{n}{24}}$ ，在 $n=1000$ 时，观测 1.24×10^{18} 次才可能看到一次“大偏差”。因此对于薄尾分布来说统计规律极为强势，你几乎看不到大偏差出现。

帕累托分布参数拟合

如果样本满足帕累托分布，要求解分布统计量时先进行参数拟合会极大提高精度，即先通过样本拟合尾部 α 值，假定分布满足：

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

对 α 的估计满足倒 gamma 分布且迅速收敛到高斯分布，因此不需要大量观测就可以对分布进行有效逼近。从下图的蒙特卡洛模拟中可以看出，求解分布均值时，极大似然估计 α 后求解相对于直接求样本均值更有效。（每个分布代表 10000 次模拟，黄色为样本均值分布，蓝色为极大似然估计法均值分布，均值采样样本量分别为 100 和 1000，黑色线为实际均值）

