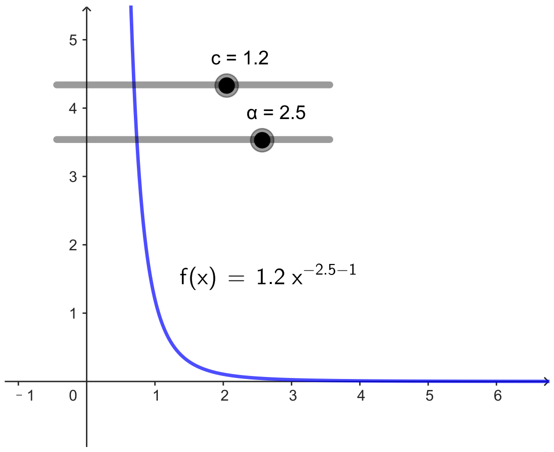
目录

文件名

第一部分 幂律分布 Power law distribution

1 幂律分布：在随机变量中，越小的数值，出现的概率越大; 越大的数值，出现的概率则越小

幂律分布的 “概率 (密度) 函数 f(x)” 是: $\boxed{f(x) = cx^{-\alpha-1}, \quad x \to \infty}$ ← α 是参数.

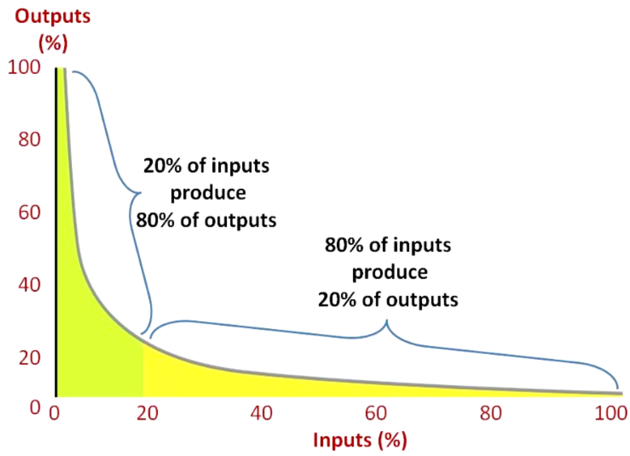


幂律分布的 “互补累积分布函数 F(x)”(CCDF) 是: $\boxed{P(X \geq x) = cx^{-\alpha}, \quad x \to \infty}$

注意: CDF 和 CCDF 的区别:

累积函数 CDF (Cumulative Distribution Function)	$F_X(x) = P(X \leq x)$
互补累计函数 CCDF (Complementary Cumulative Distribution Function)	$F(a) = P(x > a)$. 注意, 这里是大于号 (>)!

在统计学中，幂律 power law 表示的是两个量之间的函数关系: 其中一个量的相对变化, 会导致另一个量的相应 “幂次比例” 的变化，且与初值无关.



曲线的横坐标，代表随机变量的取值; 纵坐标，代表发生的概率.
幂律分布曲线的含义非常明确：在随机变量中，越小的数值，出现的概率越大；越大的数值，出现的概率则越小.

2 性质

2.1 尖峰肥尾

- 尖峰: 说明有些 x 的值很小 (赚 1000 元/月), 但其数量规模大到超乎想象 (6 亿人).
- 肥尾: 说明很多极小概率事件 (世界首富), 依然有可能发生.

如果人的身高, 是符合 "幂率分布" 的话, 则就会有极少数 "身高能长到数公里" 的人存在.

2.2 无标度 – 分形效果

无标度，也叫 "无尺度" , "尺度无关" . 意思是: 在任何观测尺度下，"幂律分布" 都呈现同样的分布特征. 即, 无论你从曲线上截取哪一段, 是长是短, 它都含有二八定律存在 (虽然曲率不同). 就相当于 "分形" 效果.
一般的分布, 都会有个尺度范围，在这个范围内服从这个分布，超过这个尺度可能就不服从这种分布了。而 "幂律分布" 没有尺度的限制，不管截取任何一个部分，都仍然呈现幂律分布的特征.

比如，图书销量是服从 "幂律分布" 的:
- 最畅销那本书的销量, 在前 10 名销量中占的比例，
- 和前 10 名的销量, 在前 100 名的销量中占的比例，
- 和前 100 名, 在前 1000 名的总销量中占的比例，
大体都是相同的. 这就是 "幂律分布" 唯一的数学特征——无标度.

2.3 幂律分布让 "均值 μ" 失去意义

"正态分布" 是一种均匀对称分布, 大多数数据都集中在 "均值 μ" 附近, 所以均值非常有用, 因为它代表大多数.
而 **"幂律分布" 呢？它的数据变化幅度非常大，平均值毫无意义.** 比如个人收入，有穷人，也有富豪，把这两群人的资产平均 (人均收入)，毫无意义.

2.4 幂律分布中的 "波动性 (方差 σ)" 失去意义

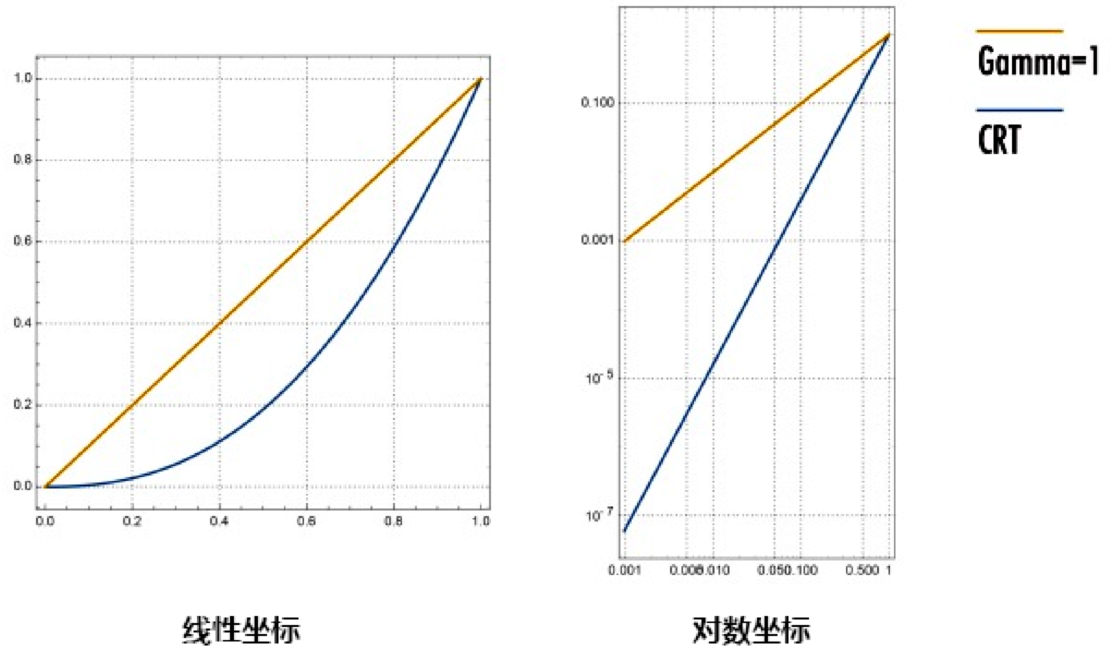
幂律分布，随机变量波动的范围非常大，常用的 "平均值"、"标准差" 到这里都没用了.

2.5 在 "双对数坐标系" 下，幂律分布表现为一条斜率为 "幂指数的负数" 的直线

自然界中大多数被识别的幂律的指数是这样的：平均值是明确的，但方差不是. 这意味着它们能够出现黑天鹅行为.

关于坐标系:
(1) 算术坐标系统 (笛卡儿坐标) : 横, 纵的刻度, 都是是等距的.

(2) "对数" 坐标系统: **坐标轴是按照 "相等的指数增长变化" 表示的.** 举例来说：如果每 1cm 代表 10 的 1 次方增加，则坐标轴刻度的表示依次为: 1，10，100，1000，10000 ...



<u>什么时候, 要使用到 “对数” 坐标系呢?</u>
→ 如果所研究的函数的 y 值, 和自变量 x, 在数值上均变化了几个数量级. 比如，已知 x 和 y 的数据为：x= 10, 20, 40, 60, 80, 100, 1000, 2000, 3000, 4000   y= 2, 14, 40, 60, 80, 100, 177, 181, 188, 200. 则,   在 “直角坐标系” 上, 就很难作图. 而换用 “对数坐标系”, 就能够画出来.
→ **当需要变换某种 “非线性关系” 为 “线性关系” 时.** (比如, “幂率分布” 的概率函数图上.)

(3) “双对数” 坐标: 指两个坐标轴, 都是 “对数坐标”. 即假如对应于 x、y 轴，则两轴等刻度情况下，其值 “以相应底数, 成次方增长”. (注意: 在各自坐标轴上的是 “真数”，不是求对数后的值.)

2.5.1 真数: 即 $\boxed{\text{底}^{\text{指(真数)}} = \text{幂}}$ 中的 “指数”

什么是 “真数 natural(number); antilogarithm”?

即: $\underbrace{2}_{\text{底}}{}^{\overbrace{3}^{\text{指, 对数}}} = \overbrace{8}^{\text{幂, 真数}}$, 即 $\log_{\underbrace{2}_{\text{底}}} \overbrace{8}^{\text{幂, 真数}} = \overbrace{3}^{\text{指, 对数}}$)