

维基百科

自由的百科全书

中心极限定理

维基百科，自由的百科全书

中心极限定理（英语：central limit theorem，简作 CLT）是概率论中的一组定理。在概率论中，中心极限定理（CLT）确认，在许多情况下，对于独立并同样分布的随机变量，即使原始变量本身不是正态分布，标准化样本均值的抽样分布也趋向于标准正态分布。这组定理是数理统计学和误差分析的理论基础，指出了大量随机变量之和近似服从正态分布的条件。

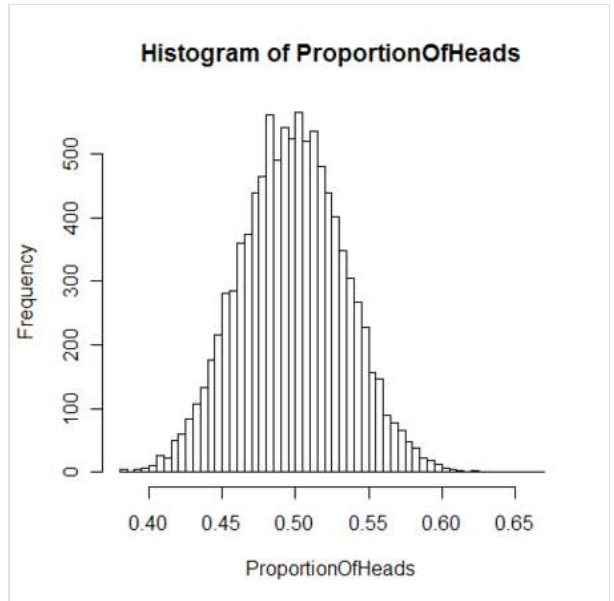
历史

Tijms（2004, p.169）写到：

“ 中心极限定理有着有趣的历史。这个定理的第一版被法国数学家棣莫弗发现，他在1733年发表的卓越论文中使用正态分布去估计大量抛掷硬币出现正面次数的分布。这个超越时代的成果险些被历史遗忘，所幸著名法国数学家拉普拉斯在1812年发表的巨著 *Théorie Analytique des Probabilités*中拯救了这个默默无闻的理论。

拉普拉斯扩展了棣莫弗的理论，指出二项分布可用正态分布逼近。但同棣莫弗一样，拉普拉斯的发现在当时并未引起很大反响。直到十九世纪末中心极限定理的重要性才被世人所知。1901年，俄国数学家里雅普诺夫用更普通的随机变量定义中心极限定理并在数学上进行了精确的证明。如今，中心极限定理被认为是（非正式地）概率论中的首席定理。

”



10,000 次抛掷硬币实验中出现正面的平均比率，每次抽样（实验）的样本数为 200（抛掷 200 次硬币）

棣莫佛-拉普拉斯定理

棣莫佛-拉普拉斯定理（De Moivre - Laplace theorem）是中央极限定理的最初版本，讨论了服从二项分布的随机变量序列。它指出，参数为n, p的二项分布以np为均值、np(1-p) 为方差的正态分布为极限。

内容

若 $X \sim B(n, p)$ 是 n 次伯努利实验中事件 A 出现的次数，每次试验成功的概率为 p ，且 $q = 1 - p$ ，则对任意有限区间 $[a, b]$ ：

令 $x_k \equiv \frac{k - np}{\sqrt{npq}}$ ，当 $n \rightarrow \infty$ 时

$$(i) P(X = k) \rightarrow \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_k^2}$$

$$(ii) P\left(a \leq \frac{X - np}{\sqrt{npq}} \leq b\right) \rightarrow \int_a^b \varphi(x) dx, \text{ 其中}$$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (-\infty < x < \infty).$$

在高尔顿板问题上的应用

棣莫弗-拉普拉斯定理指出二项分布的极限为正态分布。高尔顿板可以看作是伯努利试验的实验模型。如果我们把小球碰到钉子看作一次实验，而把从右边落下算是成功，从左边落下看作失败，就有了一次 $p = \frac{1}{2}$ 的伯努利试验。小球从顶端到底层共需要经过 n 排钉子，这就相当于一个 n 次伯努利试验。小球的高度曲线也就可以看作二项分布随机变量的概率密度函数。因此，中央极限定理解释了高尔顿板小球累积高度曲线为什么是正态分布独有的钟形曲线。

林德伯格-莱维定理

林德伯格-莱维 (Lindeberg-Levy) 定理，是棣莫弗-拉普拉斯定理的扩展，讨论独立同分布随机变量序列的中央极限定理。它表明，独立同分布 (i. i. d., 即 independent and identically distributed)、且数学期望和方差有限的随机变量序列的标准化和以标准正态分布为极限：

内容

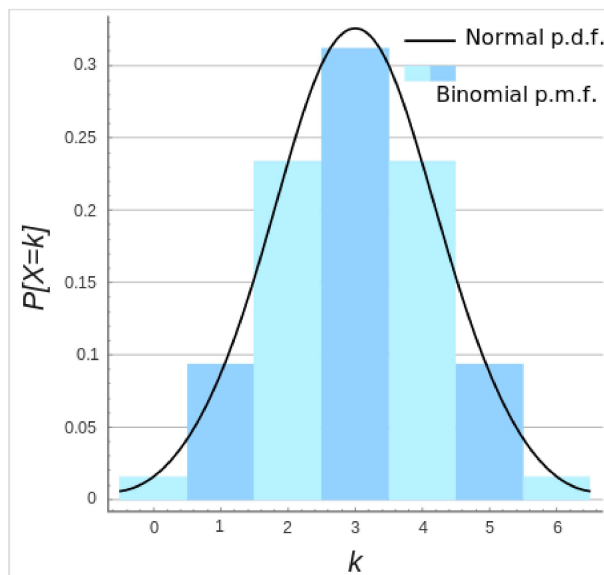
设随机变量 X_1, X_2, \dots, X_n 独立同分布，且具有有限的数学期望和方差 $E(X_i) = \mu$ ， $D(X_i) = \sigma^2 \neq 0 (i = 1, 2, \dots, n)$ 。记

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \zeta_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}, \quad \text{则}$$

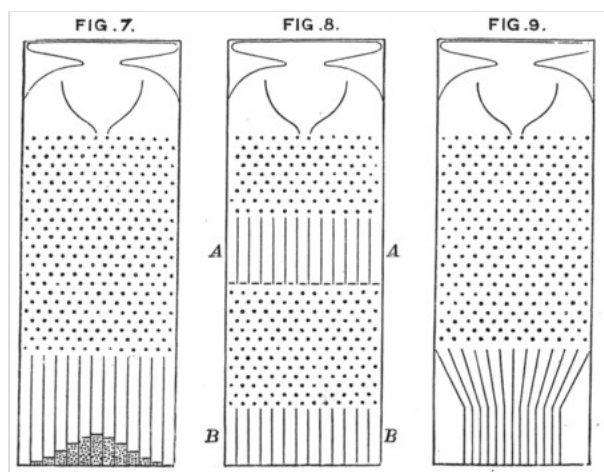
$$\lim_{n \rightarrow \infty} P(\zeta_n \leq z) = \Phi(z)$$

其中 $\Phi(z)$ 是标准正态分布的分布函数。

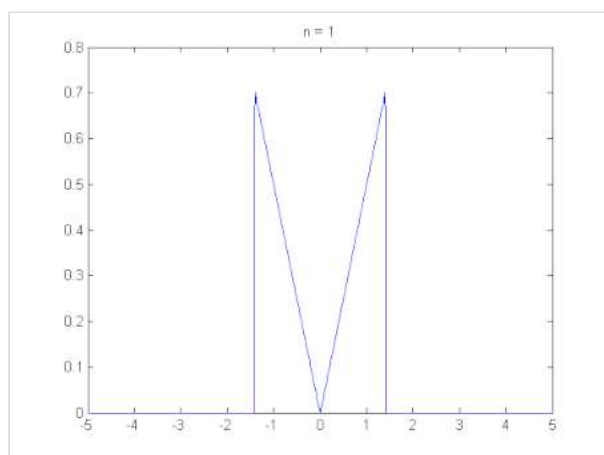
证明



用正态分布逼近二项分布



高尔顿绘制的高尔顿板模型，其中的小球显出钟形曲线。



中央极限定理的动态展示，独立同分布随机变量之和趋近正态分布。

记 $X_k - \mu$ 的特征函数为 $\varphi(t)$ ，根据傅里叶变换，样本空间中的褶积在特征函数空间变为乘积，因此 ζ_n 的特征函数为 $\left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$ 。由于 $E(X_k) = \mu$, $D(X_k) = \sigma^2$ 故 $\varphi'(0) = 0$, $\varphi''(0) = -\sigma^2$ 。因此

$$\varphi(t) = 1 - \frac{1}{2}\sigma^2 t^2 + o(t^2)$$

所以

$$\left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n = \left[1 - \frac{1}{2n}t^2 + o\left(\frac{t^2}{n}\right)\right]^n \rightarrow e^{-t^2/2}$$

由于 $e^{-t^2/2}$ 是连续函数，它对应的分布函数为 $\Phi(Z)$ ，因此由逆极限定理知

$$\lim_{n \rightarrow \infty} P(\zeta_n \leq z) \rightarrow \Phi(z)$$

定理证毕。

林德伯格-费勒定理

林德伯格-费勒（Lindeberg-Feller）定理，是中心极限定理的高级形式，是对林德伯格-莱维定理的扩展，讨论独立的，但不同分布的情况下的随机变量和。它表明，满足一定条件时，独立的，但不同分布的随机变量序列的标准化和依然以标准正态分布为极限：

内容

记随机变量序列 X_i （ X_i 独立但不一定同分布， $E[X_i] = 0$ 且有有限方差）部分和为

$$S_n = \sum_{i=1}^n X_i$$

记

$$s_i^2 = \text{Var}(X_i)$$

$$\sigma_n^2 = \sum_{i=1}^n s_i^2 = \text{Var}(S_n).$$

如果对每个 $\epsilon > 0$ ，序列满足

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma_n^2} \sum_{i=1}^n E[X_i^2; \{|X_i| > \epsilon\sigma_n\}] = 0$$

则称它满足林德伯格（Lindeberg）条件。

满足此条件的序列趋向于正态分布，即

$$S_n / \sigma_n \xrightarrow{d} N(0, 1)$$

同时，该条件也是期望为零、方差有限的独立变量之和趋于正态分布的必要条件。

与之相关的是李亚普诺夫 (Lyapunov) 条件：

$$E[|X_i|^3] < \infty, \lim_{n \rightarrow \infty} \frac{1}{\sigma_n^3} \sum_{i=1}^n E[|X_i|^3] = 0$$

满足李亚普诺夫条件的序列，必满足林德伯格条件。

证明

在此只对较强的李亚普诺夫条件给出证明。

以下证明对每一实数 t ，特征函数满足 $\varphi_{S_n/\sigma_n}(t) \rightarrow e^{-t^2/2}$ 。

$$\left| \varphi_{S_n/\sigma_n}(t) - e^{-t^2/2} \right| = \left| \prod_{k=1}^n \varphi_{X_k}(t/\sigma_n) - \prod_{k=1}^n e^{-t^2 s_k^2 / 2\sigma_n^2} \right| \leq \sum_{k=1}^n \left| \varphi_{X_k}(t/\sigma_n) - e^{-t^2 s_k^2 / 2\sigma_n^2} \right|$$

泰勒展开，上式可近似为

$$\sum_{k=1}^n \left| \frac{i^3 t^3 E[X_k^3]}{6\sigma_n^3} + \frac{t^4 s_k^4}{8\sigma_n^4} \right| \leq \frac{|t|^3}{6\sigma_n^3} \sum_{k=1}^n E[|X_k|^3] + \frac{t^4}{8\sigma_n^4} \sum_{k=1}^n s_k^4 \leq \frac{|t|^3}{6\sigma_n^3} \sum_{k=1}^n E[|X_k|^3] + \frac{t^4}{8} \max_{1 \leq k \leq n} \frac{s_k^2}{\sigma_n^2}$$

由李亚普诺夫条件，当 $n \rightarrow \infty$ 时，第一项收敛于零。

令 $k_n = \arg \max_{1 \leq k \leq n} s_k^2 / \sigma_n^2$ ，则由李亚普诺夫不等式，

$$(s_{k_n} / \sigma_n)^{3/2} \leq E[|X_{k_n} / \sigma_n|^3] \leq \frac{1}{\sigma_n^3} \sum_{k=1}^n E[|X_k|^3]$$

因此第二项也收敛于零。

证毕。

广义的定理

中心极限定理指出，随着随机变量数量的增加，许多具有有限方差的独立的且相同分布的随机变量的总和将趋于正态分布。

参阅

- [渐进分布](#)
- [大数定律](#)
- [重对数律](#)
- [鞅中心极限定理](#)
- [本福特定律](#)

- [欧文-贺尔分布](#)

参考文献

- 李贤平，[概率论基础](#)（第二版），高等教育出版社
- Olav Kallenberg，[现代概率论基础](#)（第二版），Springer（2002）。

外部链接

- [中央极限定理的动态例示（英文）](http://www.statisticalengineering.com/central_limit_theorem.htm) (http://www.statisticalengineering.com/central_limit_theorem.htm) ([页面存档备份](https://web.archive.org/web/20080913114236/http://www.statisticalengineering.com/central_limit_theorem.htm) (https://web.archive.org/web/20080913114236/http://www.statisticalengineering.com/central_limit_theorem.htm), 存于互联网档案馆)

取自 “<https://zh.wikipedia.org/w/index.php?title=中心极限定理&oldid=79535436>”

-