

基于蒙特卡洛模拟的CLT正态近似收敛速度 量化研究

宋超敏

淮北理工学院教育学院, 安徽 淮北

收稿日期: 2026年8月1日; 录用日期: 2026年8月22日; 发布日期: 2026年9月2日

摘要

中心极限定理是统计推断的理论基础, 该定理的收敛速率并非固定不变, 与总体分布形态密切相关。本研究选取均匀分布、指数分布、对数正态分布、双峰混合正态分布四类总体分布作为研究对象, 借助蒙特卡洛随机模拟, 在多个样本量条件下对标准化后的样本均值的抽样分布进行大量重复试验。研究采用Shapiro-Wilk检验、超额峰度、Q-Q图等多种诊断工具, 系统评估各分布向标准正态逼近的收敛速度。研究表明: 均匀分布收敛最快; 双峰混合正态分布虽形态异常, 其收敛速度优于右偏的指数分布; 对数正态分布在样本量2至10区间偏度与峰度出现反常上升, 显示CLT收敛过程存在非单调特征。上述发现呈现了不同总体分布下CLT收敛速度的差异, 为大样本统计推断中的样本量选取提供了参考依据。

关键词

中心极限定理, 蒙特卡洛模拟, 收敛速度, Shapiro-Wilk检验

Quantitative Study on the Convergence Rate of CLT Normal Approximation Based on Monte Carlo Simulation

Chaomin Song

School of Education, Huaibei Institute of Technology, Huaibei Anhui

Received: August 1, 2026; accepted: August 22, 2026; published: September 2, 2026

Abstract

The central limit theorem is the theoretical foundation of statistical inference. The convergence rate of this theorem is not fixed and is closely related to the overall distribution pattern. This study

selected four types of population distributions, namely uniform distribution, exponential distribution, logarithmic normal distribution, and bimodal mixed normal distribution, as the research objects. With the help of Monte Carlo random simulation, large-scale repeated experiments were conducted on the sampling distribution of the standardized sample mean under multiple sample size conditions. The study used various diagnostic tools such as the Shapiro-Wilk test, excess kurtosis, and Q-Q plot to systematically evaluate the convergence speed of each distribution towards the standard normal. Research shows that uniform distribution converges the fastest. Although the bimodal mixed normal distribution has an abnormal shape, its convergence speed is better than the right skewed exponential distribution. The logarithmic normal distribution shows an abnormal increase in skewness and kurtosis in the sample size range of 2 to 10, indicating the existence of non monotonic characteristics in the CLT convergence process. The above findings present differences in the convergence speed of CLT under different population distributions, providing a reference for sample size selection in large sample statistical inference.

Keywords

Central Limit Theorem, Monte Carlo Simulation, Convergence Rate, Shapiro-Wilk Test

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

中心极限定理(Central Limit Theorem, CLT)是概率论与数理统计核心基础,是大样本假设检验、区间估计、抽样推断的理论支撑。该定理表明:设 $X_1, X_2, \dots, X_n, \dots$ 是独立同分布的随机变量序列,且数学期望 $E(X_i) = \mu$, 方差 $Var(X_i) = \sigma^2 > 0$ 存在,则当样本容量 n 充分大时,样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 的抽样分布近似服从正态分布 $N(\mu, \sigma^2/n)$ 。等价地,标准化统计量

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

依分布收敛于标准正态分布 $N(0,1)$ [1]。

中心极限定理关键在于放宽了统计推断对总体分布形态的要求,只要样本容量 n 充分大,无论原始数据服从何种分布,样本均值的分布均近似服从正态分布,这为参数估计和假设检验提供可行的理论依据[2]。然而,中心极限定理本质上属于渐近性定理,其严格成立的条件需要样本量趋近于无穷。在现实研究场景中,研究者所面对的样本量总是有限的,抽样分布与正态分布之间必然存在一定偏差[3]-[5]。日常教学中常以样本容量 $n \geq 30$ 作为经验阈值,这一结论的普适性从未得到严格的论证。事实上,中心极限定理的收敛速度并非是一个固定值,受到总体分布内在属性偏度与峰度的影响。原始分布越不对称、尾部越厚向正态收敛的速度就越慢;反之,对称且轻尾的分布收敛就越快。此外,当总体方差不存在时,中心极限定理根本不适用。因此,对不同类型总体分布下中心极限定理收敛速度的系统性比较,不仅具有理论意义,更具有重要的实践指导价值。

利用蒙特卡洛模拟方法验证中心极限定理的收敛性是数理统计中一种极具代表性的研究方法。该方法的核心思想是从选定的总体中反复生成随机样本,接着从生成的随机样本中计算每次抽样所得到的标准化均值,通过观察其经验分布向标准正态分布的逼近程度来展示中心极限定理的收敛过程[6]-[8]。在已有的研究中,大多数学者们所选取的研究对象为均匀分布、指数分布等简单的总体分布。例如,吴婷等

借助 R 语言环境中随机模拟了几种常见分布的中心极限定理[9] [10]。Magsalay 进一步研究偏态总体分布, 通过蒙特卡洛模拟研究了中心极限定理收敛所需的最小样本量[11]。

然而, 上述研究在分布的选取上多集中于均匀分布和指数分布, 未涉及极端形态的总体分布。基于此, 本文选取多类典型分布进行模拟仿真, 对比不同样本量下标准化样本均值向标准正态分布逼近的收敛速度。

2. 模拟设计

2.1. 总体分布选择

本研究选取均匀分布、指数分布、对数正态分布、双峰混合正态分布四类具有代表性的母体分布来比较不同分布特征对中心极限定理收敛速度的影响。

2.1.1. 均匀分布

均匀分布是连续型分布中最简单的形式, 设随机变量 $X \sim U(0,1)$, 则其概率密度函数为

$$f(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{其他.} \end{cases}$$

该分布的数学期望 $\mu = 0.5$, 方差 $\sigma^2 = 1/12 \approx 0.0833$, 偏度为 0, 超额峰度为 -1.2。均匀分布具有对称、有界、轻尾的特征, 理论上收敛速度极快。选取该分布作为“最快收敛基准”, 用于检验在最优情境下 CLT 的逼近效果。

2.1.2. 指数分布

指数分布具有无记忆性, 是概率论中常用的分布之一, 其概率密度函数为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \lambda > 0, \\ 0, & \text{其他.} \end{cases}$$

本研究取速率参数 $\lambda = 1$, 此时期望 $\mu = 1$, 方差 $\sigma^2 = 1$, 偏度为 2, 超额峰度为 6。指数分布具有右偏、轻尾的特征, 是模拟偏态分布的经典代表。选取该分布作为“中等收敛基准”, 用于考察偏度对收敛速度的拖累效应。

2.1.3. 对数正态分布

对数正态分布是金融、生物医学等领域中常见的数据分布模型。若 $Y \sim N(\mu, \sigma^2)$, 则 $X = e^Y$ 服从对数正态分布, 其概率密度函数为

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], x > 0.$$

本研究取 $\mu = 0, \sigma = 1$, 此时期望 $E(X) = e^{0.5} \approx 1.6487$, 方差 $Var(X) = e(e-1) \approx 4.67$ 。该分布的核心特征在于具有极强的右偏性, 其偏度系数为 $(e+2)\sqrt{e-1} \approx 6.185$, 超额峰度高达 $e^4 + 2e^3 + 3e^2 - 6 \approx 110.936$ 。

2.1.4. 双峰混合正态分布

双峰混合正态分布的概率密度函数为两个正态分布的加权平均, 设随机变量 X 服从双峰混合正态分布, 由两个正态总体 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ 加权组合, 权重系数 $\omega_1, \omega_2 > 0$, 且满足 $\omega_1 + \omega_2 = 1$, 则其概率密度函数为

$$f(x) = \omega_1 \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right] + \omega_2 \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right], -\infty < x < +\infty.$$

本研究选取 $\omega_1 = \omega_2 = 0.5$ ，此时期望 $E(X) = 0$ ，方差 $Var(X) = \mu^2 + 1$ 。其偏度始终为 0，但峰度 μ 随的增大而增大，具体表达式为

$$\text{Kurtosis} = 3 + \frac{\mu^4 + 6\mu^2}{\mu^2 + 1}.$$

当 $\mu \rightarrow 0$ 时双峰混合正态分布的峰度为 3，与标准正态分布一致；当 $\mu \rightarrow \infty$ 时，峰度为 $3 + \mu^2$ ，趋于无穷。由于该分布的图形具有对称性，峰度便成为了衡量其偏离正态程度的唯一指标，所以选取该分布作为研究对象可单独考察形态异常对收敛速度的影响，而不受偏度因素的干扰。

当 $\mu = 2.0$ 时该分布形态呈现明显的双峰分离特征，因此，本文选取 $\mu = 2.0$ 作为基准参数，足以代表态显著异常的情形。

为便于后续参考，将上述四类分布的关键理论特征汇总于表 1。

Table 1. Summary of theoretical characteristics of each population distribution
表 1. 各总体分布的理论特征汇总

分布	参数	期望	方差	偏度	超额峰度	选取定位
均匀分布	$a=0, b=1$	0.5	0.0833	0	-1.2	最快基准(对称轻尾)
指数分布	$\lambda=1$	1	1	2	6	偏态基准(中等偏态)
对数正态分布	$\mu=0, \sigma=1$	1.6487	4.6708	6.185	110.936	极端偏态
双峰混合正态	$\mu=2.0$	0	5	0	3.5	形态异常(中等双峰)

2.2. 样本量与模拟次数设定

本文用 R 语言统计软件进行所有蒙特卡洛模拟，设定样本量为 $n = 2, 5, 10, 20, 30, 50, 100, 200$ 。样本量序列的选取遵循三条原则。其一，在 $n < 30$ 的小样本区间上加密取值 (2, 5, 10, 20)，旨在精细捕捉 CLT 收敛初期的演化轨迹；其二，纳入 $n = 30$ 经典阈值，以便检验经典经验法则的有效性；其三，取 $n = 50, 100, 200$ 是用于考察大样本下的收敛行为。

设定蒙特卡洛模拟次数为 $R = 5000$ 。选择依据为：5000 次模拟足以保证标准化样本均值 Z 的分布特征(均值、标准差、分位数等)具有足够的稳定性，模拟均值的标准误约为 $1/\sqrt{5000} \approx 0.014$ ；兼顾计算效率， $R = 5000$ 可计算软件上于数分钟内完成全部模拟。

2.3. 模拟流程

针对每一种总体分布和每一个样本量 n ，执行以下模拟流程。

步骤一：从指定的总体分布中独立抽取 n 个随机样本 $X_1, X_2, \dots, X_n$ 。

步骤二：计算样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ ，并将其标准化

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

其中 μ 和 σ 为总体真实的期望与标准差。

步骤三：将步骤一至步骤二独立重复 $R = 5000$ 次，得到标准化均值的序列 $\{Z_1, Z_2, \dots, Z_{5000}\}$ 。

步骤四：对上述 5000 个 Z 值进行正态性诊断，计算相关量化指标。

步骤五：遍历所有样本量 $n \in \{2, 5, 10, 20, 30, 50, 100, 200\}$ ，重复步骤一至步骤四。

步骤六：遍历所有总体分布，重复步骤一至步骤五。

2.4. 收敛速度的量化指标体系

为避免仅依赖直方图视觉判断的主观性，本研究采用多维度量化指标综合评估收敛速度。

2.4.1. Shapiro-Wilk 正态性检验

Shapiro-Wilk 检验是检验小样本正态性的常用方法，其原假设为样本来自正态分布。为评估不同样本量下标准化均值向正态的收敛程度，本研究采用分组检验策略。将全部 $R = 5000$ 次模拟结果分为 $B = 100$ 组，每组包含 $M = 50$ 个标准化均值 Z 。对每一组独立实施 Shapiro-Wilk 检验，记录 p 值。若 $p > 0.05$ ，则在 5% 显著性水平下不能拒绝正态性假设，认为抽样分布已近似正态。最终统计 100 组中拒绝正态性假设的组数比例，定义为拒绝率。拒绝率越低，表明该样本量下抽样分布与正态分布的偏离程度越小，收敛效果越好。

分组数 $B = 100$ 与每组样本量 $M = 50$ 的设定基于以下考虑：Shapiro-Wilk 检验在小样本条件下仍具有较高的检验功效，每组 50 个观测值足以保证检验的稳定性；同时 100 组的划分使拒绝率的标准误控制在较小范围内，确保统计结果的可靠性。在收敛性的主要判断上，以拒绝率为核心依据，当拒绝率低于 5% 时，认为该样本量下的正态近似已具备可接受的水平。

2.4.2. 偏度系数

偏度衡量分布的不对称性，其定义为

$$S = E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right],$$

标准正态分布的偏度为 0。计算每组 5000 个 Z 值的偏度系数，其绝对值越小，说明分布越对称，收敛越好。

2.4.3. 超额峰度

峰度衡量分布的尾部厚度，其定义为

$$K = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right],$$

标准正态分布的峰度为 3，超额峰度为 $K - 3$ ，标准正态的超额峰度为 0。计算每组 5000 个 Z 值的超额峰度，其绝对值越接近 0，说明尾部行为越接近正态。

2.4.4. Kolmogorov-Smirnov 检验

KS 检验是比较两个分布的非参数检验方法。将每组 5000 个 Z 值的经验分布函数与标准正态分布函数 $N(0,1)$ 进行比较，计算 D 统计量

$$D = \sup_x |F_n(x) - \Phi(x)|,$$

其中 F_n 为经验分布函数， Φ 为标准正态分布函数。 D 值越小，说明抽样分布越接近标准正态分布。

上述的四项指标各有侧重，四者相互印证，确保结论的有效性。

3. 模拟结果

3.1. Shapiro-Wilk 正态性检验结果

为刻画各分布下的收敛速度，本研究采用 Shapiro-Wilk 检验进行正态性诊断。具体操作如下：将 $R = 5000$ 次模拟分为 $B = 100$ 组，每组包含 $M = 50$ 个 Z 值，在显著性水平 $\alpha = 0.05$ 下进行统计拒绝正态性假设组数的比例。结果见表 2。

Table 2. Shapiro Wilk test rejection rate (%) for different sample sizes in various distributions
表 2. 各分布不同样本量下 Shapiro-Wilk 检验拒绝率(%)

样本量 n	均匀分布 $U(0,1)$	指数分布 $Exp(1)$	对数正态分布 $LogN(0,1)$	双峰混合正态 $BMN(\mu = 2)$
2	3	98	100	99
5	2	58	99	92
10	6	36	98	86
20	4	15	91	70
30	3	19	86	74
50	8	13	75	69
100	5	3	54	51
200	5	4	41	48

注：表中拒绝率表示在显著性水平下拒绝正态性假设的模拟次数比例。

Shapiro-Wilk 检验结果表明，均匀分布在 $n = 2$ 时拒绝率低于 5%，验证了极小样本量下对称轻尾分布可达到良好的正态近似。

对于右偏的指数分布，拒绝率从 $n = 2$ 时的 98%单调下降至 $n = 100$ 时的 3%，但首次通过正态性检验所需的最小样本量约为 $n = 100$ ，说明经典的 $n \geq 30$ 经验法则对于右偏分布而言并不充分。

在这四类分布中，对数正态分布收敛速度最缓慢，其拒绝率从 $n = 2$ 时的 100%仅降至 $n = 200$ 时的 41%，在本文所研究选取的所有样本量下均未进入可接受区间，这一结果清晰展现了极端偏态与重尾特征对中心极限定理收敛构成了严重的影响。

对于双峰混合正态分布而言，尽管该分布具有对称性，但其拒绝率却从 $n = 2$ 时的 99%缓慢降至 $n = 200$ 时的 48%，这一现象说明 Shapiro-Wilk 检验对分布形态中双峰结构具有极强的敏感度，即使偏度和峰度指标均已接近正态水平，但其双峰结构仍会影响收敛进程。

3.2. 偏度与超额峰度分析

偏度衡量分布的对称性，正态分布偏度为 0；超额峰度衡量分布的尾部厚度，正态分布超额峰度为 0。对标准化均值 Z 计算偏度系数与超额峰度，结果如表 3 所示。

Table 3. Skewness and excess kurtosis of standardized means under different sample sizes for each distribution
表 3. 各分布不同样本量下标准化均值的偏度与超额峰度

样本量 n	均匀分布 $U(0,1)$	指数分布 $Exp(1)$	对数正态分布 $LogN(0,1)$	双峰混合正态 $BMN(\mu = 2)$
2	-0.020/-0.545	1.520/3.830	3.060/11.340	-0.017/2.470
5	0.026/-0.256	0.809/0.804	2.850/7.910	-0.056/1.740
10	0.029/-0.085	0.663/0.667	3.950/12.670	0.055/0.682
20	-0.008/-0.079	0.404/0.285	1.130/3.070	0.013/0.171
30	-0.024/-0.026	0.343/0.092	1.000/1.880	0.029/0.012
50	-0.025/0.038	0.313/0.197	0.760/1.820	0.017/0.030
100	0.015/0.002	0.201/-0.037	0.616/1.590	0.010/-0.057
200	-0.038/-0.050	0.109/-0.073	0.550/0.700	0.046/-0.114

注：表中数值格式为“偏度/超额峰度”。

结果表明均匀分布的偏度与超额峰度在所有样本量下均接近于 0，进一步验证了其快速收敛特性。指数分布的偏度从 $n=2$ 时的 1.520 单调递减至 $n=200$ 时的 0.109，超额峰度从 3.830 快速降至 -0.073，其在 $n=30$ 时超额峰度已降至 0.092，但偏度仍为 0.343，表明偏度是指数分布收敛速度的主要制约因素。而对数正态分布的偏度在 $n=2$ 至 $n=10$ 区间从 3.060 反常上升至 3.950，超额峰度从 11.340 上升至 12.670。这表明在极端偏态分布下，中心极限定理的收敛过程并不具有单调性，当 $n=200$ 时偏度与超额峰度仍未收敛到 0。双峰混合正态分布的偏度始终接近于 0，超额峰度从 $n=2$ 时的 2.470 快速下降到 $n=30$ 时的 0.012，此后便在 0 附近波动。然而，在相同样本量下双峰混合正态分布的 Shapiro-Wilk 检验的拒绝率为 74%，这反映了 Shapiro-Wilk 检验对分布形态的敏感度远超于偏度与峰度指标。

3.3. Q-Q 图直观诊断

Q-Q 图将样本分位数与理论正态分位数进行对比，若数据来自正态分布，则各点应大致落在参考直线 $y=x$ 上。为了进一步展示四种分布在不同样本量下向正态逼近的程度，本研究在 Q-Q 图中引入 95% Bootstrap 置信带，其原理如下：对每个样本量设定下的 5000 个标准化均值 Z ，在 99 个等间距概率点上计算样本分位数，通过 500 次 Bootstrap 重采样估计各分位数的置信区间。置信带越窄、越贴近参考线，表明该样本量下抽样分布的收敛越稳定、正态近似越可靠。研究绘制四种分布在样本量 $n=2,10,30,100$ 下的 Q-Q 图，其中蓝色区域为置信带，红色虚线为参考线，见图 1。

图 1 的 Q-Q 图呈现了四种分布在不同样本量下的收敛过程，不同分布之间的差异显著。均匀分布表现最为平稳，在 $n=2$ 极端小样本下 Q-Q 图中的置信带虽有一定宽度且两端略有扩张，但整体已包络参考线，点云走势已大致呈线性；在 $n=10$ 时置信带明显收窄，完全贴附于参考线两侧。此后随着样本量的增大置信带几乎缩为一条线，与参考线高度重合，反映出对称轻尾分布的极速收敛特性。

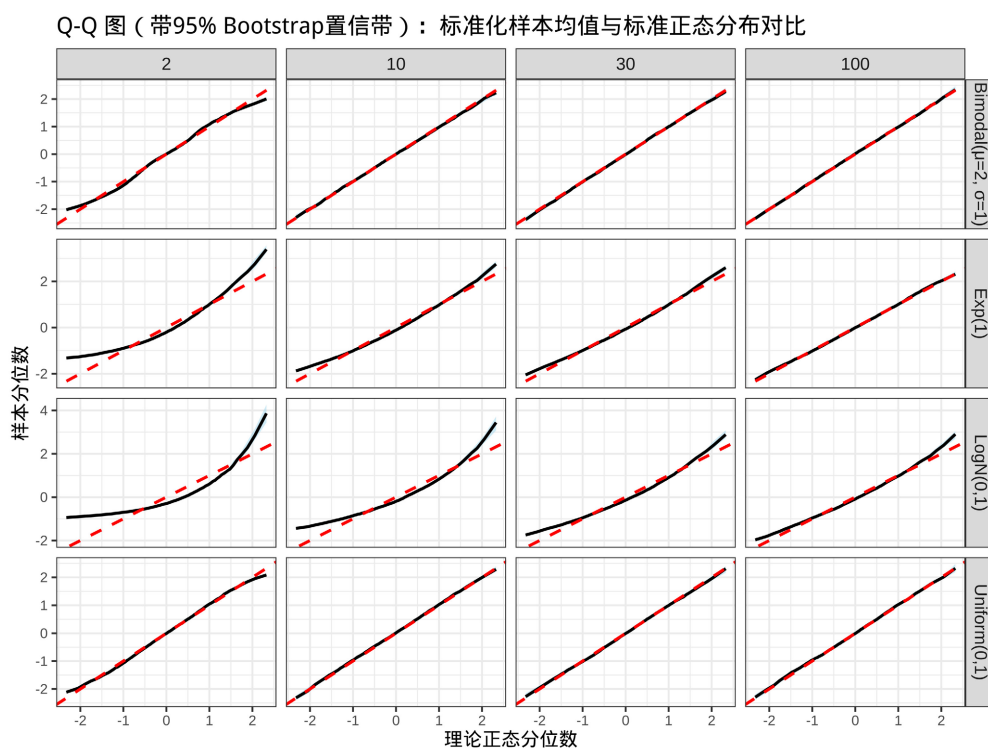


Figure 1. Q-Q plots of standardized means under four representative sample sizes $n=2,10,30,100$

图 1. 标准化均值在 $n=2,10,30,100$ 四个代表性样本量下的 Q-Q 图

图中指数分布在 $n=10$ 时置信带整体偏向右上方,尤其在右尾区域偏离参考线较为明显,反映出右偏分布的残余影响;在 $n=30$ 时置信带大幅收窄,大部分区域已贴近参考线,但右尾仍存在轻微的系统性偏离,直至 $n=100$ 置信带完全包络参考线,表明该样本量下正态近似已足够可靠。

对数正态分布与上述两种分布存在明显差异,在 $n=2$ 时 Q-Q 图中置信带已呈现极端上凸形态,右尾大幅偏离参考线;当样本量从 2 增加到 10 时,偏离程度不仅没有改善反而加剧,这与表 3 中偏度与峰度的反常上升相互吻合;此后,在 $n=30$ 时置信带上凸程度开始下降,但右尾依旧高于参考线;当 $n=100$ 时右尾末端仍有轻微上翘的趋势。这一结果清晰地展示了极端偏态分布对中心极限定理收敛速率有持久性的阻碍作用。

双峰混合正态分布在 $n=2$ 时置信带呈现出 S 形曲线,中间区域斜率较为陡峭,两端趋于平缓,这正是两个正态成分混合后在分位数尺度上产生的突变特征;然而,这一异常形态随样本量增加而迅速消退,在 $n=10$ 时 S 形大幅减弱,而在 $n=30$ 时置信带基本呈线性趋势,仅中心区域可见极其微弱的波动。

3.4. 超额峰度收敛路径

为刻画各分布超额峰度随样本量的衰减过程,绘制了超额峰度绝对值随样本量 n 变化的收敛路径图见图 2,横轴采用对数刻度。图 2 展示了四种分布超额峰度绝对值的收敛路径。

图 2 展示了四种分布超额峰度绝对值随样本量增加而衰减的收敛路径。均匀分布作为对称轻尾的代表,其超额峰度在 $n=2$ 时超额峰度仅为 0.545,到 $n=10$ 时已降至 0.085,此后稳定在 0 附近,是四条曲线中最为平直的收敛路径。

双峰混合正态分布的超额峰度从 $n=2$ 时的 2.47 快速下降到 $n=20$ 时的 0.17,在样本容量 20 附近逐渐稳定在零轴。这一现象表明尽管双峰分布在直观形态上偏离钟形曲线,但其轻尾本质并不影响峰度收敛的速度。

指数分布的超额峰度从 $n=2$ 时的 3.83 快速下降, $n=10$ 时约为 0.67, $n=30$ 时降至 0.09,在 $n=30$ 时进入 0 轴附近,其收敛路径与双峰分布相似,但略慢于双峰混合正态分布。

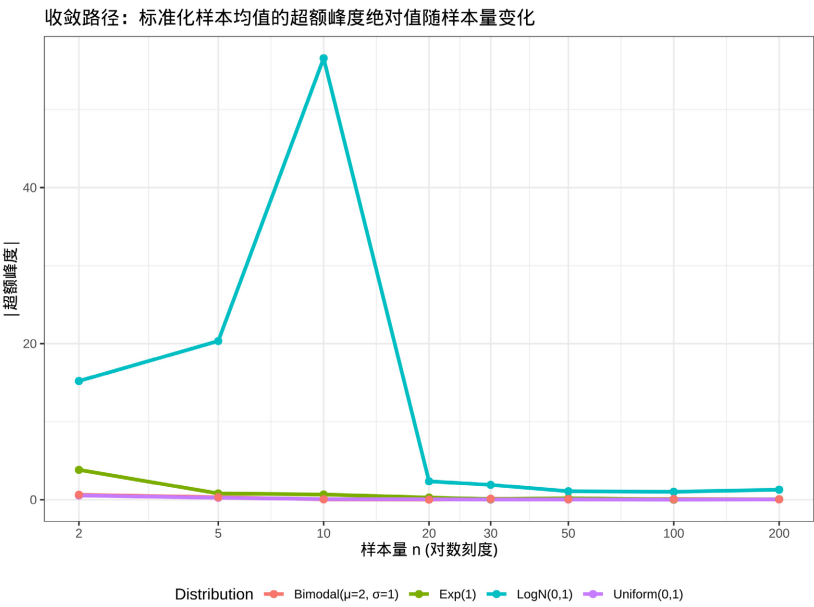


Figure 2. Convergence path of excess kurtosis absolute value with sample size n
图 2. 超额峰度绝对值随样本量 n 变化的收敛路径图

对数正态分布呈现出最为极端的收敛特征，其超额峰度绝对值在 $n=2$ 时为 11.34， $n=5$ 时短暂回落至 7.91，但 $n=10$ 时反而上升至 12.67，形成整条曲线中的峰值，随后以极为缓慢的速率下降，直至 $n=200$ 时仍维持在 0.70，远未收敛至 0 轴。这条曲线初始值极高、中途反常上升、其后持续缓慢衰减的特征清晰地揭示出中心极限定理在极端偏态分布下的收敛路径并非具有单调性，其缓慢性也远超基于常规分布的经验预期。

3.5. 收敛速度综合比较

综合 Shapiro-Wilk 检验、偏度、超额峰度及 Q-Q 图诊断，四种分布的收敛速度排序如表 4 所示。

Table 4. Comprehensive comparison of convergence speed of various distributions
表 4. 各分布收敛速度综合比较

收敛速度排序	分布	主要制约因素
1 (最快)	均匀分布	无
2	双峰混合正态分布	双峰形态异常
3	指数分布	右偏偏态
4 (最慢)	对数正态分布	极强偏态 + 重尾

通过综合诊断发现：

- (1) 偏度的影响大于形态异常。双峰分布虽形态异常，但因其图形具有对称性，所以超额峰度收敛速度快于指数分布，这表明在 CLT 收敛中，偏度所起的制约作用比分布的形态异常更为重要。
- (2) CLT 收敛存在非单调性。对数正态分布的偏度与超额峰度在 $n=2$ 至 $n=10$ 区间不降反升，表明增大样本容量并不总是改善正态近似。
- (3) 教科书上的 $n \geq 30$ 经验法则并不适合所有的抽样分布。均匀分布选取极小样本量即可、指数分布所需样本量应至少大于 100、对数正态分布样本容量选取 200 仍不够。这表明不存在普适的最小样本量阈值，在实际应用中应根据总体分布的偏态与峰度特征灵活判断。
- (4) 多指标联合诊断的必要性。双峰分布的超额峰度在 $n=30$ 时已接近 0，在数值上已接近完美正态，但 Shapiro 检验拒绝率仍高达 74%，这一发现说明在评估 CLT 收敛程度时，不宜依赖任何单一诊断指标，需结合正态性检验、偏度、超额峰度与 Q-Q 图进行多维度诊断。

4. 结论

本文通过蒙特卡洛模拟，对比了均匀分布、指数分布、对数正态分布和双峰混合正态分布在 CLT 中的收敛速度，主要发现如下：均匀分布收敛速度大于双峰混合正态大于指数分布大于对数正态分布；偏度的影响大于形态异常，双峰分布虽形状异常但对称，收敛快于右偏的指数分布；对数正态分布的偏度和峰度在 $n=2$ 至 $n=10$ 区间不降反升，说明增大样本量不一定立即改善近似； $n \geq 30$ 经验法不普适，应因分布制宜。基于模拟结果，在实践过程中可将总体分布的偏度与峰度作为判断依据。当偏度接近于 0 且超额峰度低于 1 时样本量选取 $n \geq 5$ 即可获得良好的正态近似；当偏度绝对值在 0.5 至 1.5 之间、超额峰度低于 3 时建议样本量选取应不低于 50，若提升到 $n=100$ 则稳定性更佳；当偏度绝对值超过 2 或超额峰度显著高于 3 时选取 $n=200$ 可能仍不足以保证可靠的近似，优先考虑数据变换或非参数方法；当偏度接近于 0 但分布呈现双峰或多模态特征时建议选取 $n \geq 100$ ，并联合多种诊断工具验证。

参考文献

- [1] 茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程[M]. 第2版. 北京: 高等教育出版社, 2011.
- [2] Serfling, R.J. (2009) *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.
- [3] 江海峰. 中心极限定理的蒙特卡罗模拟分析[J]. 统计教育, 2007(2): 6-7.
- [4] 农吉夫, 黄文宁. 随机模拟实验在概率极限理论教学中的应用[J]. 广西民族大学学报(自然科学版), 2008, 14(1): 101-104.
- [5] 肖胜, 李应求, 危渝绍. 随机环境中多型分枝过程均值估计量的中心极限定理和 Berry-Esseen 界[J/OL]. 数学学报(中文版), 1-15. <https://link.cnki.net/urlid/11.2038.O1.20260330.1510.012>, 2026-07-24.
- [6] Kolokoltsov, V., Lin, F. and Mijatović, A. (2021) Monte Carlo Estimation of the Solution of Fractional Partial Differential Equations. *Fractional Calculus and Applied Analysis*, **24**, 278-306. <https://doi.org/10.1515/fca-2021-0012>
- [7] Robert, C.P., Casella, G. and Casella, G. (2004) *Monte Carlo Statistical Methods*. Springer.
- [8] 张立华. 蒙特卡洛统计方法衡量股票市场风险——研究误区与本源反思[J]. 中国统计, 2017(5): 65-67.
- [9] 吴婷, 普映娟. R 语言在概率统计课堂教学中的应用——中心极限定理随机模拟[J]. 保山学院学报, 2022, 41(5): 69-76.
- [10] 赵渊, 周家启, 刘志宏. 大电网可靠性的序贯和非序贯蒙特卡洛仿真的收敛性分析及比较[J]. 电工技术学报, 2009, 24(11): 127-133.
- [11] Magsalay, R.J.M. (2025) Quantifying Central Limit Theorem Convergence: A Monte Carlo Simulation Approach to Minimum Sample Size. *International Journal of Research and Innovation in Social Science*, **9**, 639-644. <https://doi.org/10.47772/ijriss.2025.908000050>