

基于大数定律和中心极限定理的高尔顿钉板试验结果分析

孙慧静¹ 马启建² 杜彬彬¹

(1. 海军航空大学, 航空基础学院, 山东 烟台 264001;

2. 烟台文化旅游职业学院, 山东 烟台 264003)

摘要: 在 MATLAB 环境下实现高尔顿钉板试验的模拟, 并对试验结果进行了总结: 当钉子的层数 n 较多时, 小球落入位置近似服从正态分布; 当小球的数量 m 较大时, 小球堆成的形状具有稳定性且近似于二项分布分布律折线形状; 当小球数量 m 较大、钉子的层数 n 较多时, 小球堆成的形状具有稳定性且近似于正态分布的密度曲线形状。基于 (大数定律和中心极限定理) 分析了高尔顿钉板试验结果出现的原因, 并纠正了人们对试验结果以及出现原因的模糊认识, 同时数值仿真的结果验证了大数定律和中心极限定理的正确性。

关键词: 高尔顿钉板试验; 大数定律; 中心极限定理; 正态分布; MATLAB

高尔顿钉板试验是由英国生物统计学家高尔顿设计。许多学者对高尔顿钉板试验做了分析与研究, 但往往停留在描述性说明而没有严格的数学证明 [2-4], 或者仅仅利用中心极限定理来解释试验结果 [5-6]。目前人们往往对高尔顿试验结果存在一些模糊认识。例如:

当放入大量小球, 则其最后呈现的曲线几乎总是类似的。这一试验结果仅用中心极限定理就可作解释吗? 事实上, 要解释该试验结果, 不仅需要中心极限定理, 而且还需要大数定律。

严格地说, 小球的落入位置服从二项分布而非正态分布, 但是当满足一定的条件时, 可以说“小球落入位置近似服从正态分布”以及“大量小球落入格子后形成的形状近似于正态分布的密度曲线形状”。这里的条件仅仅需要“小球的数目足够多”吗? 事实上, 条件既需要“小球的数目足够多”, 又需要“钉子的层数足够大”。

鉴于此, 本文将利用 MATLAB 编程实现高尔顿钉板试验的模拟, 将实际中耗时耗力重复的试验, 利用计算机得以实现, 大大提高了效率; 基于严格的数学理论分析了试验结果出现的原因以及所得结论的正确性, 并纠正了人们对试验结果以及结果产生原因的模糊认识。

一、高尔顿钉板试验的仿真及实验结果分析

(一) 高尔顿钉板试验的结果

设高尔顿钉板有 n 层钉, 第 n 层钉共有 n 个, 两个钉之间一个空, 则有 $(n+1)$ 个空。把这 $(n+1)$ 个空由左到右依次编号 $i=0, 1, \dots, n$ 共 $(n+1)$ 个空。设 Y 表示小球的落入位置, 则 Y 的所有可能取值为 $0, 1, \dots, n$, 并设小球碰到钉子向左落下的概率为 p , 向右落下的概率为 $q=1-p$ 。

在 MATLAB 环境下, 高尔顿钉板试验的模拟。图 1-3 展示了当钉子的层数 $n=8$, 小球的个数分别为 $m=10, 100$, 和 1000 时的模拟结果。这里小球碰到钉子向左、向右落下的概率分别为 $p=\frac{1}{2}$, $q=\frac{1}{2}$ 。

由试验结果可以得到如下结论:

结论 (1): 小球的落入位置 $Y \sim B(n, p) B(n, p)$; 当钉子的层数 n 较多时, 小球落入位置近似服从正态分布;

结论 (2): 当小球数量 m 较小时, 小球堆成的形状不具有

稳定性, 也不具备“正态密度曲线形状”的特点; 当小球数量 m 较大时, 小球堆成的形状才具有稳定性, 具备“二项分布分布律折线”形状;

结论 (3): 当小球数量 m 较大, 且钉子的层数 n 较多时, 小球堆成的形状具有稳定性, 且小球落入格子后形成的形状近似于“正态分布的密度曲线”形状。

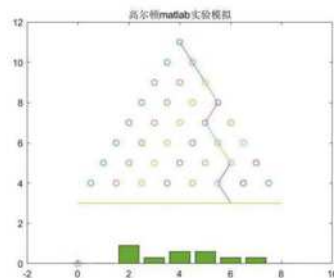


图 1 高尔顿钉板试验的 Matlab 模拟结果, 小球的个数 $m=10$, 钉子的层数 $n=8$

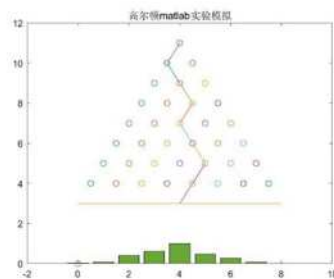


图 2 高尔顿钉板试验的 Matlab 模拟结果, 小球的个数 $m=100$, 钉子的层数 $n=8$

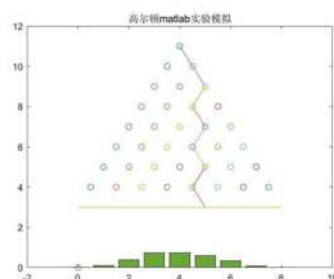


图 3 高尔顿钉板试验的 Matlab 模拟结果, 小球的个数 $m=1000$, 钉子的层数 $n=8$

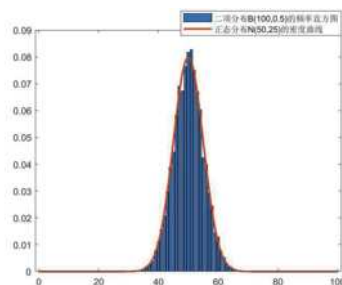


图 6 二项分布 $B(100, 0.5)$ 的频率直方图和正态分布 $N(50, 25)$

(二) 结论分析

1. 结论 1 的分析说明:

$$X_i = \begin{cases} \frac{1}{2}, & \text{若小球第 } i \text{ 次碰钉后向右落下,} \\ -\frac{1}{2}, & \text{若小球第 } i \text{ 次碰钉后向左落下,} \end{cases} \quad i=1, 2, \dots, n,$$

其中 n 代表钉子的层数.

$$X_i \text{ 的分布律为: } P\left\{X_i = -\frac{1}{2}\right\} = 0.5, \quad P\left\{X_i = \frac{1}{2}\right\} = 0.5.$$

若引入直角坐标系, 设第 0 个格子的坐标为 0, 第一个格子的坐标为 1, 依次类推, 第 n 个格子的坐标为 n , 则小球最终的落入位置为 $Y = \frac{n}{2} + \sum_{i=1}^n X_i$, 且 Y 服从二项分布, 即

$$Y = \sum_{i=1}^n X_i \sim B(n, 0.5). \text{ 故}$$

$$P\left\{\text{小球落入第 } k \text{ 个格子}\right\} = P\{Y = k\} = \binom{n}{k} (0.5)^k (0.5)^{n-k}, k = 0, 1, 2, \dots, n.$$

根据棣莫弗-拉普拉斯中心极限定理可得, 当 $n \rightarrow \infty$ 时, 二项分布依分布收敛到正态分布. 故当钉子的层数 n 较多时, 小球落入位置 Y 近似服从正态分布. 故结论 (1) 成立. 显然试验结果验证了中心极限定理的正确性.

2. 结论 2 的分析说明:

假设共有 m 个小球, 令 Y_i 表示第 i 个小球的最终落入位置, 显然 $Y_1, Y_2, \dots, Y_m$ 可看作是来自总体 Y 的一个样本. $Y_1, Y_2, \dots, Y_m$ 相互独立, 且 $Y_i \sim B(n, 0.5), i=1, 2, \dots, m$.

据伯努利大数定律可得, 当 $m \rightarrow \infty$ 时, 小球落入第 i 个格子的频率依概率收敛到落入第 i 个格子的概率. 当小球的数量 m 较小时, 不满足大数定律的条件, 则频率就不具备稳定性. 小球的数量 m 较大时, 满足大数定律的条件, 则频率依概率收敛到概率. 因此, 我们有: 当小球数量 m 较小时, 小球堆成的形状不具有稳定性, 也不具备落入中间格子的较多, 落入两边格子的较少的特点; 当小球数量 m 较大时, 小球堆成的形状具有稳定性. 故结论 (2) 成立.

图 4 (a) 展示了二项分布 $B(8, 0.5)$ 的分布律, 图 4 (b) 展示了当小球的数目为 $m=10000$, 钉子的层数 $n=8$ 时落入每个格子的小球频数直方图, 结果显示落入第 5 个格子的频数最多. 图 4 (c) 展示了 4 (b) 图相应的落入每个格子的小球频率直方图. 根据图 4 (a) 和图 4 (c), 证实伯努利 (Bernoulli) 大数定律的正确性, 即, 当小球的数目 $m \rightarrow \infty$, 小球落入每个格子的频率依概率收敛到小球落入每个格子的概率.

3. 结论 3 的分析说明:

根据中心极限定理, 当钉子的层数 n 较多时, 小球落入位置 Y 近似服从正态分布, 根据大数定律, 当 $m \rightarrow \infty$ 时, 小球落入第 i 个格子的频率依概率收敛到落入第 i 个格子的概率. 所以, 当小球数量 m 较大, 且钉子的层数 n 较多时, 小球堆成的形状具有稳定性, 且小球落入格子后形成的形状近似于正态分布的密度曲线形状. 故结论 (3) 成立.

图 5 (a) 展示了二项分布 $B(100, 0.5)$ 的分布律, 图 5 (b) 展示了当小球的数目为 $m=10000$, 钉子的层数 $n=8$ 时落入每个格子的小球频数直方图. 图 5 (c) 展示了图 5 (b) 图相应的落入每

个格子的小球频率直方图. 图 5 (a)、图 5 (c) 及图 6, 不仅证实了伯努利大数定律的正确性, 而且也证实了棣莫弗-拉普拉斯中心极限定理的正确性.

比较图 4 和图 5, 虽然都是大样本的情况下, 当钉子的层数较少时, 小球落入格子后最后呈现的轮廓形状与正态分布差别较大; 当钉子的层数较多时, 小球落入格子后最后呈现的轮廓形状与正态分布比较吻合, 从而也证明了中心极限定理只有在钉子层数较多的情形下成立.

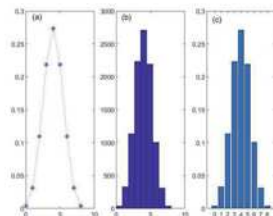


图 4 (a) 二项分布 $B(8, 0.5)$ 的分布律; (b) 二项分布 $B(8, 0.5)$ 的频数直方图, $m=10000$; (c) 二项分布 $B(8, 0.5)$ 的频率直方图, $m=1000$.

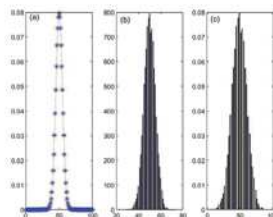


图 5 (a) 二项分布 $B(100, 0.5)$ 的分布律; (b) 二项分布 $B(100, 0.5)$ 的频数直方图, $m=10000$; (c) 二项分布 $B(100, 0.5)$ 的频率直方图, $m=10000$.

二、总结

本文在 MATLAB 环境下实现了对高尔顿钉板试验的模拟, 对试验结果进行总结得出一般结论:

基于大数定律和中心极限定理, 分析试验结果出现的原因, 解释了所得结论的合理性和准确性, 纠正了人们长期以来对于高尔顿钉板试验的结果以及结果出现原因的模糊认识.

参考文献:

- [1] 李贤平. 概率论基础 [M]. 第三版. 北京: 高等教育出版社, 2010: 4-5.
- [2] 王维, 向瀚淋, 龚雯丽, 等. 常见分布中心极限定理适用样本量研究 [J]. 高师理科学刊, 2021, 41 (7): 20-25.
- [3] 徐彬. 中心极限定理的形象化模拟及其应用 [J]. 湖北汽车工业学院学报, 2016, 30 (2): 77-80.
- [4] 侯臣平, 矫媛媛. Matlab 在《概率论与数理统计》教学中的应用 [J]. 教育教学论坛, 2019 (5): 156-157.
- [5] 武新乾, 张翠霞, 轩凤霞. 高尔顿钉板试验动态图形软件的设计与制作 [J]. 洛阳师范学院学报, 2015, 34 (2): 4.
- [6] 聂燕. 高尔顿钉板试验的算法实现及分析 [J]. 中国民航飞行学院学报, 2008 (3): 3.

基金项目: 海军航空大学教育训练理论研究 (13), 海军航空大学课程思政专项研究立项课题 (05), 海军航空大学教学成果培育立项 (30)

作者简介: 孙慧静 (1982-), 女, 副教授, 博士, 从事大学数学教育教学研究、数学理论与应用方法研究. Email: 14118420@bjtu.edu.cn