

文章编号: 1001-4217(2005)02-0012-07

用 MATLAB 模拟大数定律和中心极限定理

林小苹, 吴文杰

(汕头大学数学系, 广东 汕头 515063)

摘 要: 利用 MATLAB 软件对概率论中的大数定律和中心极限定理进行随机模拟, 并通过图形将模拟结果直观、动态地演示出来, 从而对这两个定理所反映的本质内容给出了直观的解释和说明, 达到了传统教学所不能达到的目的和效果.

关键词: 随机模拟; 大数定律; 中心极限定理; MATLAB

中图分类号: O 211; O 242.1; G 434 **文献标识码:** A

0 引 言

大数定律和中心极限定理是概率论中两类具有广泛实用意义的重要定理. 其中, 大数定律是数理统计中“算术平均值法则”和矩法估计的理论基础; 中心极限定理则是正态分布得以广泛应用的理论基础. 由于这两类定理的理论性较强, 结论奇特, 不易被学生深刻理解和掌握, 教学中常出现的问题是, 无论教师对这两类定理解释得如何通俗, 学生还是对定理的结论似懂非懂. 为了解决这一突出的教学难题, 本文尝试用强大的数学软件 MATLAB 作为开发工具, 开发了一个交互式的大数定律和中心极限定理的计算机随机模拟教学辅助软件, 给讲授、学习和理解这两类定理提供了一个界面友好的计算机辅助教学环境, 使学生可以形象、直观地“看”到大数定律和中心极限定理的“真面目”.

1 相关定理

1.1 大数定律和中心极限定理

考虑到随机模拟过程的易操作性, 本文只对大量独立同分布的随机变量进行模拟, 即只对辛钦大数定律、贝努利大数定律、林德贝尔格-勒维中心极限定理和德莫佛-拉普拉斯中心极限定理进行随机模拟^[1].

辛钦大数定律 设 ξ_1, ξ_2, \dots 是一列独立同分布的随机变量, 且具有数学期望

收稿日期: 2005-02-04

作者简介: 林小苹(1964~), 女, 广东汕头人, 讲师.

$E(\xi_i) = \mu$, ($i = 1, 2, \dots$). 作前 n 个随机变量的算术平均 $\eta_n = \frac{1}{n} \sum_{i=1}^n \xi_i$, 则序列 $\eta_n = \frac{1}{n} \sum_{i=1}^n \xi_i$ 依概率收敛于 μ . 即对任意的 $\epsilon > 0$, 有:

$$\lim_{n \rightarrow \infty} P\left\{\left|\eta_n - \mu\right| < \epsilon\right\} = \lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n \xi_i - \mu\right| < \epsilon\right\} = 1 \quad (1)$$

贝努利大数定律 设 μ_n 是 n 重贝努利试验中事件 A 出现的次数, p 是事件 A 在每次试验中出现的概率, 即 $P(A) = p$, 则对任意的 $\epsilon > 0$, 有:

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{\mu_n}{n} - p\right| < \epsilon\right\} = 1 \quad (2)$$

显然, 当辛钦大数定律中的 ξ_i 服从 0-1 分布时, 它就是贝努利大数定律.

林德贝尔格-勒维中心极限定理 若 ξ_1, ξ_2, \dots 是一列独立同分布的随机变量, 且具有数学期望 $E(\xi_i) = \mu$ 和方差 $D(\xi_i) = \sigma^2 (\sigma^2 > 0)$, $i = 1, 2, \dots$, 令

$$\zeta_n = \frac{\sum_{i=1}^n \xi_i - \sum_{i=1}^n E(\xi_i)}{\sqrt{\sum_{i=1}^n D(\xi_i)}} = \frac{\sum_{i=1}^n \xi_i - n\mu}{\sqrt{n\sigma^2}},$$

则 ζ_n 的分布函数 $F_n(x) = P\{\zeta_n \leq x\}$ 当 $n \rightarrow \infty$ 时, 趋于标准正态分布 $N(0, 1)$, 即:

$$\lim_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} P\left\{\frac{\sum_{i=1}^n \xi_i - n\mu}{\sqrt{n\sigma^2}} \leq x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad (3)$$

德莫佛-拉普拉斯中心极限定理 在 n 重贝努利试验中, 事件 A 在每次试验中出现的概率为 p , μ_n 为 n 次试验中事件 A 出现的次数, 则:

$$\lim_{n \rightarrow \infty} P\left\{\frac{\mu_n - np}{\sqrt{np(1-p)}} \leq x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad (4)$$

显然, 当林德贝尔格-勒维中心极限定理中的 ξ_i 服从 0-1 分布时, 它就成了德莫佛-拉普拉斯中心极限定理.

1.2 格列汶科定理

格列汶科定理 当 $n \rightarrow \infty$ 时, 经验分布函数 $F_n(x)$ 以概率 1 关于 x 一致地收敛于总体的分布函数 $F(x)$, 即: $P\left\{\lim_{n \rightarrow \infty} \sup_{-\infty < x < +\infty} |F_n(x) - F(x)| = 0\right\} = 1^{[2]}$.

此定理说明, 当样本容量足够大时, 对所有 x , 经验分布函数 $F_n(x)$ 充分接近于总体的分布函数 $F(x)$. 显然, 格列汶科定理为我们提供了可以用样本来推断总体的理论基础.

2 随机模拟算法及其软件实现

2.1 随机模拟算法

为了对大数定律和中心极限定理进行模拟, 根据格列汶科定理, 可以将大量独立随机变量的算术平均 $\frac{1}{n} \sum_{i=1}^n \xi_i$ 及大量独立随机变量的和 $\sum_{i=1}^n \xi_i$ 分别看作总体, 并对总体进

行抽样, 通过研究抽样样本的分布来实现对总体本身分布的推断. 因此, 要实现大数定律和中心极限定理的计算机模拟, 关键在于尽可能多地进行抽样试验. 为解决这一问题, 本文利用 MATLAB 软件产生服从某种分布的一组随机数来代替现实中的一次抽样试验.

对于公式(1), 设 $\eta_n = \frac{1}{n} \sum_{i=1}^n \xi_i$, n 为随机变量的个数, 则大数定律的随机模拟算法为:

- 1) 设置循环的跳跃步长 $step = 5$;
- 2) 选择随机数 ξ_i 的分布类型;
- 3) 设置 n 的上限值 m ;
- 4) 给出第一次抽样的样本容量初始值 n_0 , 即 $n = n_0$;
- 5) 利用计算机产生一组(n 个)服从同一分布的随机数 ξ_i , ($i = 1, 2, \dots, n$);
- 6) 计算 η_n 的值;
- 7) $n = n + step$, 如 $n > m$, 则转 8), 否则, 回到 5);
- 8) 以 x 轴代表样本容量 n , y 轴代表每次抽样所得的样本平均值 η_n , 画出整个试验的过程(静态显示或动态显示均可).

根据大数定律, 随着随机变量的个数 n 逐步增大, 抽样值的平均值 η_n 会越来越趋向于 μ . 显然, 如果在算法 2) 中选择随机数服从 0-1 分布, 则该算法模拟的是公式(2), 即贝努利大数定律.

对于公式(3), 设 $\zeta_n = \frac{\sum_{i=1}^n \xi_i - n\mu}{\sqrt{n\sigma}}$, n 仍为随机变量的个数, m 为总体 ζ_n 的样本

容量, 则中心极限定理的随机模拟算法如下:

- 1) 选择随机数 ξ_i 的分布类型;
- 2) 设置试验总次数 m 的值;
- 3) 设置随机变量个数 n 的值;
- 4) 利用计算机产生一组(n 个)服从同一分布的随机数 ξ_i , ($i = 1, 2, \dots, n$);
- 5) 将 $\sum_{i=1}^n \xi_i$ 的值规范化, 即计算 ζ_n 的值(总共有 m 个), 记为 $(\zeta_n^{(1)}, \zeta_n^{(2)}, \dots, \zeta_n^{(m)})$;
- 6) 画出这 m 个 ζ_n 值的频率直方图.

根据中心极限定理, 当随机变量的个数 n 越来越大时, 这 m 个 ζ_n 值的分布情况将越来越趋于标准正态分布. 同样, 根据格列汶科定理, m 值的大小也会影响模拟效果. 在 n 一定的情况下, m 越大, ζ_n 的分布越接近于标准正态分布.

显然, 若在算法 1) 中选择随机数服从 0-1 分布, 则本算法模拟的是公式(4), 即德莫佛-拉普拉斯中心极限定理.

2.2 软件实现

本文的计算机随机模拟软件全部由 MATLAB 数学软件实现, 软件的设计过程主要包

括两大部分: 1) 图形用户界面(GUI)设计; 2) 回调程序设计。

1) 图形用户界面设计

本文采用 MATLAB 的低级函数设计 GUI, 使用起来灵活方便。界面如图 1 所示。整个界面划分为三大部分: 菜单项区, 图形显示区, 用户操作区。其中, 菜单项区位于整个界面的最上端, 主要有 6 项功能: 辛钦大数定律的随机模拟, 贝努利大数定律的随机模拟, 林德贝尔格-勒维中心极限定理的随机模拟, 德莫佛-拉普拉斯中心极限定理的随机模拟, 缩放, 显示。缩放功能可以对图形显示区的图形进行放大或缩小, 显示功能则可以对图形的显示功能进行修改(如修改图线的颜色、线型态和数据点的标志等)。图形显示区主要用于显示大数定律和中心极限定理的随机模拟结果。用户操作区则用于放置各种用户操作控件, 负责完成数据输入、分布类型选择和图形显示形式的选择等。

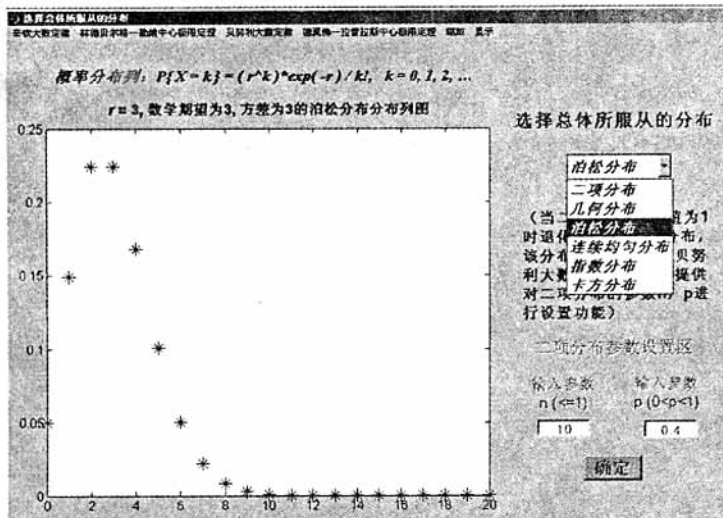


图 1 大数定律和中心极限定理随机模拟软件的主界面

在 GUI 设计中, 主要使用图形句柄函数 `uicontrol` 创建各种含图形对象的图形用户界面。该函数能建立按钮、选择框、列表框以及文本框等对象。其调用格式为:

`h = uicontrol(hfig, 属性名, 属性值, ...)`

其中, `hfig` 参数是某个图形窗口的句柄值, 缺省时为当前窗口。

设计图形用户界面时, 在界面上每创建一个对象, 就需要调用该函数一次, 并根据不同对象的需要, 选择不同的属性名和属性值。

例如, 创建如图 2 中的“动态显示”命令按钮, 其程序代码段主要为:

```
h = uicontrol('style','pushbutton', ... %指定控件类型, 缺省为“按钮”
            'units', 'characters', ... %指定 Position 的度量单位
            'position', [147 14 22 3], ... %定义按钮所在的位置和外形大小
            'string', '动态显示', ... %定义按钮的名称
            'callback', DtxsPrg, ... %定义按钮的调用事件
            'tag', 'pushbutton1');
```

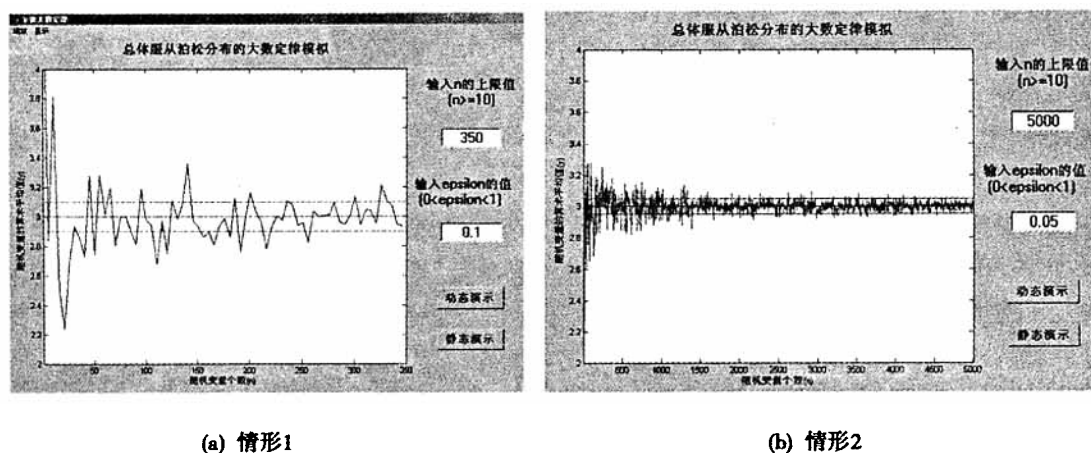


图 2 总体服从泊松分布的大数定律的模拟界面

需要注意的是,随着所创建图形对象的不同, `uicontrol` 函数的属性值也不同,且对同一种图形对象,属性也较多,可有选择地使用^[3].

2) 回调程序设计

要实现对大数定律和中心极限定理的动态随机模拟,还需要利用 MATLAB 软件编写相应的随机模拟程序^[4],并通过 GUI 所提供的回调(Callback)处理模式,把各个随机模拟程序嵌入到用户操作界面中,以实现各个模拟功能.限于文章的篇幅,这里仅给出图 2 中“动态演示”按钮的部分回调程序.

```

.....
x = (1:5:n);           % n 为随机变量个数
y = (1:5:n);
for k = 1:5:n
    a = poissrnd(3, k, 1); %生成 k × 1 形式的服从泊松分布的随机数矩阵
    a = sum(a)/k;
    y(count) = a;
    count = count + 1;
end
axis([1, n, 2, 4]);
hold on;
plot((1:n), 3 + e, 'r:', (1:n), 3 - e, 'r:', (1:n), 3, 'k-');
comet(x, y);           % 动态显示点(x, y)的运动轨迹
.....

```

3 随机模拟及其假设检验

3.1 大数定律的随机模拟

选取泊松分布(其数学期望为3)作为本次大数定律的模拟对象,即 ξ_i 服从泊松分

布. 当 $n = 350$, $\epsilon = 0.1$ 时, $\frac{1}{n} \sum_{i=1}^n \xi_i$ 的随机模拟曲线几乎位于带状区域 3 ± 0.1 内, 见图 2(a), 这里的“几乎”用概率的语言说, 就是“大概率”; 相反, 模拟曲线落在带状区域以外的那部分属于小概率情形. 这是因为, 此时的 n 是有限的 ($n \leq 350$), 并非无穷大. 如果继续给定更小的 $\epsilon (= 0.05)$, 并且取足够大的 $n (= 5\,000)$, 则这次的模拟曲线几乎全部落入带状区域内, 见图 2(b).

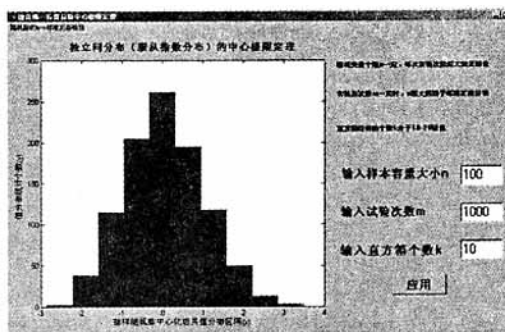
因此, 由图 2(a)和(b)的图形变化可以看出, 随着随机变量个数 n 的增加, $\frac{1}{n} \sum_{i=1}^n \xi_i$ 的取值逐渐密集在其数学期望值 3 的附近, 这正是大数定律所反映出的统计规律性.

3.2 中心极限定理的随机模拟及其假设检验

选取指数分布作为本次模拟所用的随机数的分布类型(其数学期望为 2, 方差为 4). 如图 3(a)所示, 当随机变量个数 $n = 10$, 试验总次数 $m = 1\,000$ 时, 得到 $(\zeta_{10}^{(1)}, \zeta_{10}^{(2)}, \dots, \zeta_{10}^{(1000)})$ 的直方图形状已经有点标准正态分布的“味道”(中间高, 两边低). 继续增大 n 值进行模拟, 当 $n = 100$ 时, 1 000 个随机数 $(\zeta_{100}^{(1)}, \zeta_{100}^{(2)}, \dots, \zeta_{100}^{(1000)})$ 的直方图更加接近于标准正态分布的图形, 见图 3(b). 显然, 由图 3(a)→(b)的这种演变过程, 正是中心极限定理所要表达的客观事实.



(a) 情形1



(b) 情形2

图3 总体服从指数分布的林德贝尔格-勒维中心极限定理的模拟界面

当然, 这只是直观上的推断, 并不严密. 为了从理论上验证总体 $\zeta_n = \frac{\sum_{i=1}^n \xi_i - n\mu}{\sqrt{n\sigma}}$

是否真的服从标准正态分布, 本软件还提供了 k-s 检验(即 Kolmogorov-Smirnov 检验^[5])功能, 如图 4 所示.

在图 4 中, 取显著性水平 $\alpha = 0.05$, 点击“k-s 检验”按钮进行 k-s 检验. 该检验的

原假设是总体 $\zeta_n = \frac{\sum_{i=1}^n \xi_i - n\mu}{\sqrt{n\sigma}} \sim N(0,1)$, 返回值为 H , P , $Ksstat$, Cv , 它们分别表示检验结果、 p 值、检验统计量和截断值. 当 $H = 0$ 时, 表示接受原假设; $H = 1$ 时,

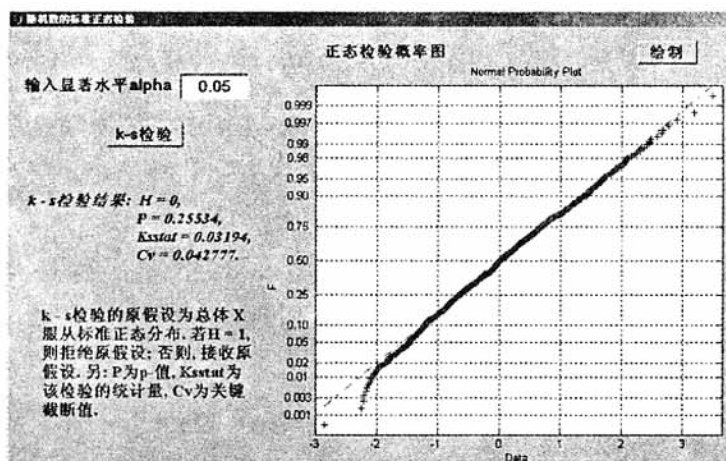


图4 随机数的标准正态假设检验的操作界面

表示拒绝原假设. 从图4可以看出, 本次的 k-s 检验结果是 $H=0$. 因此, $\zeta_n = (\sum_{i=1}^n \xi_i - n\mu)/(\sqrt{n}\sigma)$ 服从标准正态分布的原假设成立. 图4右边是正态概率图, 这是一种直观的假设检验方法^[2]. 若图上的点近似地落在一条直线附近(直线两端附近允许偏差大一些), 接受原假设, 否则拒绝原假设. 从图中可以看出, 正态概率图的结果与 k-s 检验的结果吻合.

4 结 语

概率论中的大数定律和中心极限定理是理解和学习数理统计的难点和重点. 本文所介绍的对这两类定理的计算机随机模拟软件已投入实际应用. 由于它能在短时间内对随机现象进行成千上万次模拟实验, 并直观形象地显示实验结果, 因而使传统课堂教学中无法实现的大量实验及动态演示在瞬间成为现实, 从而把原本抽象难懂的知识变得直观形象, 容易理解, 使教学过程形象生动, 富于趣味性, 达到了用计算机辅助教学以提高课堂教学效果的目的.

参考文献:

- [1] 盛聚, 谢式千, 潘承毅. 概率论与数理统计[M]. 北京: 高等教育出版社, 1979.
- [2] 华东师范大学数学系. 概率论与数理统计教程[M]. 北京: 高等教育出版社, 1983.
- [3] 王学辉, 张明辉. MATLAB 6.1 最新应用详解[M]. 北京: 中国水利水电出版社, 2002.
- [4] 陈桂明, 戚红雨, 潘伟. MATLAB 数理统计(6.x)[M]. 北京: 科学出版社, 2002.
- [5] 陆璇. 数理统计基础[M]. 北京: 清华大学出版社, 1998.

(下转第80页)

- [21] 邵伟波, 饶江. 高压氧治疗对脑卒中后抑郁状态和神经功能康复的影响[J]. 中国康复理论与实践, 2002, 12: 755.
- [22] 杨明明. 脑卒中后抑郁及相关因素分析[J]. 中国康复理论与实践, 2001, 4: 160.
- [23] Takahashi S, Crana A M, Jehle J, et al. Role of the cerebellar fastigial nucleus in the physiological regulation of cerebral blood flow[J]. J Cereb Blood Flow Metab, 1995, 15: 128
- [24] 许长春, 王黎荔, 路常玲. 低能量 He-Ne 激光治疗卒中后抑郁症[J]. 实用医药杂志, 2002, 11: 839.

Diagnosis and Treatment of Poststroke Depression

LIU Yi-xing¹, ZHENG Shao-xiong², LIU Lian¹, JIN Qin-hua¹

(1. Department of Neurology, Second Affiliated Hospital, Shantou University Medical College, Shantou 515041, Guangdong, China; 2. Mental Health Centre, Shantou University Medical College, Shantou 515063, Guangdong, China)

Abstract: To counter the high rate of missed diagnosis of post-stroke depression (PSD), the clinical symptom, diagnostic norm and treatment of PSD are introduced in this article. The curative effect is analyzed and compared with the side effect. Combined treatment is more suitable for senile patients with PSD.

Key words: stroke; post-stroke depression; diagnosis; treatment

(上接第 18 页)

Simulating of Large Numbers Law and Central Limit Theorem with MATLAB Software

LIN Xiao-ping, WU Wen-jie

(Department of Mathematics, Shantou University, Shantou 515063, Guangdong, China)

Abstract: In this paper, law of large numbers and central limit theorem in probability theory were simulated randomly with MATLAB software, and the simulating results were demonstrated audio-visually and dynamically with graphs. Thus the essence of the above-mentioned theorems could be expounded audio-visually. It was proved that the use of this software in teaching could achieve the goals and effects that could not be gained by the traditional teaching.

Key words: random simulation; law of large numbers; central limit theorem; MATLAB