Advances in Applied Mathematics 应用数学进展, 2024, 13(5), 2209-2220
Published Online May 2024 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2024.135210

Hans汉斯

基于Matlab的概率论与数理统计问题的探究与算法

李浩清[1]，胡新豪[2]，付鹏辉[2]，廖　秀[3*]

[1]桂林信息科技学院商学院，广西　桂林
[2]桂林信息科技学院电子工程学院，广西　桂林
[3]桂林信息科技学院数学教研部，广西　桂林

收稿日期：2024年4月29日；录用日期：2024年5月22日；发布日期：2024年5月31日

摘　要

本文主要探讨了如何使用Matlab对概率论与数理统计中的一些问题进行探讨。结合Matlab强大的计算和可视化优势，主要从极大似然估计法、线性回归、大数定律及聚类分析通过Matlab实现计算。通过应用Matlab，学生可以更深入地领悟概率论与数理统计里的概念，提高学习兴趣和实际应用能力。同时，Matlab的实践性和可视化特点也使得研究更加便捷和直观。通过本文的探讨，可以为学生和教育者提供有益的指导和支持，促进概率论与数理统计的学习和研究。

关键词

Matlab，概率论与数理统计，可视化，实际应用

Exploration and Algorithms of Probability Theory and Mathematical Statistics Problems Based on Matlab

Haoqing Li[1], Xinhao Hu[2], Penghui Fu[2], Xiu Liao[3*]

[1]Business School, Guilin Institute of Information Technology, Guilin Guangxi
[2]College of Electronic Engineering, Guilin Institute of Information Technology, Guilin Guangxi
[3]Department of Mathematics Teaching and Research, Guilin Institute of Information Technology, Guilin Guangxi

Received: Apr. 29th, 2024; accepted: May 22nd, 2024; published: May 31st, 2024

*通讯作者。

文章引用: 李浩清, 胡新豪, 付鹏辉, 廖秀. 基于 Matlab 的概率论与数理统计问题的探究与算法[J]. 应用数学进展, 2024, 13(5): 2209-2220. DOI: 10.12677/aam.2024.135210

Abstract

This paper mainly explores how to use Matlab to investigate some problems in probability theory and mathematical statistics. Leveraging the powerful computational and visualization advantages of Matlab, computations are mainly implemented through Matlab for maximum likelihood estimation, linear regression, law of large numbers, and cluster analysis. Through the use of Matlab, students can better understand the concepts of probability theory and mathematical statistics, enhance their interest in learning, and improve their practical application abilities. Furthermore, Matlab's practicality and visualization features make research more convenient and intuitive. Through the discussion in this paper, beneficial guidance and support can be provided to students and educators, promoting learning and research in probability theory and mathematical statistics.

Keywords

Matlab, Probability Theory and Mathematical Statistics, Visualization, Practical Application

Copyright © 2024 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

概率论与数理统计是数学领域中非常重要的分支，其为我们提供了研究随机现象和数据分析的理论基础，也是研究生入学考试的重要内容之一；概率论主要研究随机事件、随机变量和随机过程的数学模型，而数理统计则关注如何从数据中获取有用的信息和推断出结论。概率论与数理统计的方法被广泛应用于各个领域，例如，在金融、医疗、工业、自然灾害等领域。

许多学生却因其理论的抽象性和计算的复杂性而感到困扰。传统的教育方式往往侧重于理论讲解和公式推导，这使得学生难以将这些理论与实际问题相结合，从而影响了学习的兴趣和效果。为了解决这一问题，我们引入了 Matlab 这一强大的数学工具，Matlab 是一种高级编程语言和交互式环境，它在算法开发、数据分析、图像处理、数值计算以及可视化等多个领域享有广泛应用。它提供了丰富的函数库和工具箱，使得用户可以方便地进行数据处理、统计分析、图像处理等工作。基于 Matlab 对概率论与数理统计的问题进行计算，使该课程相关问题的计算和结果更加便捷和直观，帮助学生更好地理解概率论与数理统计的概念与解决实际问题，提高学习兴趣和实际应用能力。

2. 概率论与数理统计问题引入 Matlab 的意义

2.1. 对学生而言

1) 提升学习效率：概率论与数理统计涉及大量的数值计算和模型建立，使用 Matlab 可以大大简化这些过程。

2) 增强实践能力：Matlab 的引入使学生有机会接触和使用先进的数学软件，锻炼他们的实际操作能力。通过解决具体的概率论与数理统计问题，学生可以培养将理论知识应用于实际问题的能力，为未来的科学研究或工作实践打下基础。

3) 拓宽视野：通过学习和使用 Matlab，学生可以接触到更多的学科领域和应用场景，从而拓宽他们

的视野，增强跨学科学习的能力。

2.2. 对教育者而言

1) 提升教学质量：Matlab 的引入可以使教育者更容易地展示复杂的数学模型和计算过程，从而帮助学生更好地理解和掌握知识点。此外，教育者还可以利用 Matlab 设计更多的实践项目和案例分析，提升教学质量和效果。

2) 推动课程创新：教育者可以根据 Matlab 的特点和优势，结合概率论与数理统计的课程内容，开发新的教学资源和课程模式。例如，可以设计基于 Matlab 的在线课程、实验课程或项目式课程，以满足不同学生的学习需求和发展方向。

3. Matlab 在概率论与数理统计问题中的应用

在概率论与数理统计领域，Matlab 作为一款强大的数学计算软件，广泛应用于各种统计分析中。本文主要通过 Matlab 探讨了方差分析、大数定律、极大似然回归、聚类分析、参数估计等方法的实际应用与计算，展示其在数据处理和分析中的强大功能。以下应用分析基于 Matlab2023b 来实现，安装 Matlab2023b 建议最低配置：内存 16G+，处理器：3.0GHz+。

3.1. 极大似然估计法

最大似然估计法的原理是基于最大似然原理的一种统计方法，其目的是利用已知的样本数据来估计模型参数。在统计学中，给定一组观测数据，假设这些数据服从某种概率分布，但其中的参数是未知的。极大似然估计的目标就是通过观测数据，找到能使得观测数据出现的概率最大的参数值，即估计出最可能的参数值[1]。

极大似然估计的优点在于通过最大化似然函数来估计参数，方法较为直观，并且在估计量的性质上有一些很好的性质。但是，在实际应用中需要注意估计量的一致性和有效性等问题。

案例 1：

假设某电子元件的寿命服从指数分布，现在我们有一组观测数据，表示了这些元件的寿命。假设我们观测到的寿命数据为：{100, 150, 200, 250, 300, 350, 400}小时。这组数据表示了不同元件的寿命。

设其概率密度函数为：

$$f\left(x|\lambda\right)=\lambda*\exp\left(-\lambda x\right)\quad \text{for}\quad x\geq 0$$

其中 λ 为指数分布的参数，x 为元件的寿命。

使用极大似然估计来估计参数 λ。假设有 n 个观测数据 $\left\{x_1,x_2,\cdots,x_n\right\}$，则似然函数为：

$$L\left(\lambda\right)=\prod_{i=1}^{n}\lambda*\exp\left(-\lambda x_i\right)$$

对似然函数取对数，得到对数似然函数：

$$l\left(\lambda\right)=\sum_{i=1}^{n}\left[\log\left(\lambda\right)-\lambda x_i\right]$$

为了找到使得观测数据出现的概率最大的参数值，我们需要求解对数似然函数的导数，并令其等于 0，从而得到极大似然估计值。求导后得到：

$$\frac{\partial l\left(\lambda\right)}{\partial\lambda}=\sum_{i=1}^{n}\left(\frac{1}{\lambda}-x_i\right)=0$$

解上述方程得到极大似然估计值为：

$$\hat{\lambda} = n \Big/ \sum_{i=1}^{n} x_i$$

用 Matlab 进行指数分布的极大似然估计时[2]，可以使用以下代码：

```
% 定义观测数据
data = [100, 150, 200, 250, 300, 350, 400];
% 使用极大似然估计来估计指数分布的参数
lambda = 1 / mean(data);
% 输出估计得到的参数值
disp(['估计的指数分布参数 lambda 为: ', num2str(lambda)]);
% 绘制指数分布的概率密度函数图
x = 0:10:500;
pdf = exppdf(x, 1/lambda); %使用自带的指数分布概率密度函数
plot(x, pdf, 'LineWidth', 2);
xlabel('寿命(小时)');
ylabel('概率密度');
title('指数分布的概率密度函数');
```

结果

```
>> Untitled02
估计的指数分布参数 lambda 为：0.004。
```

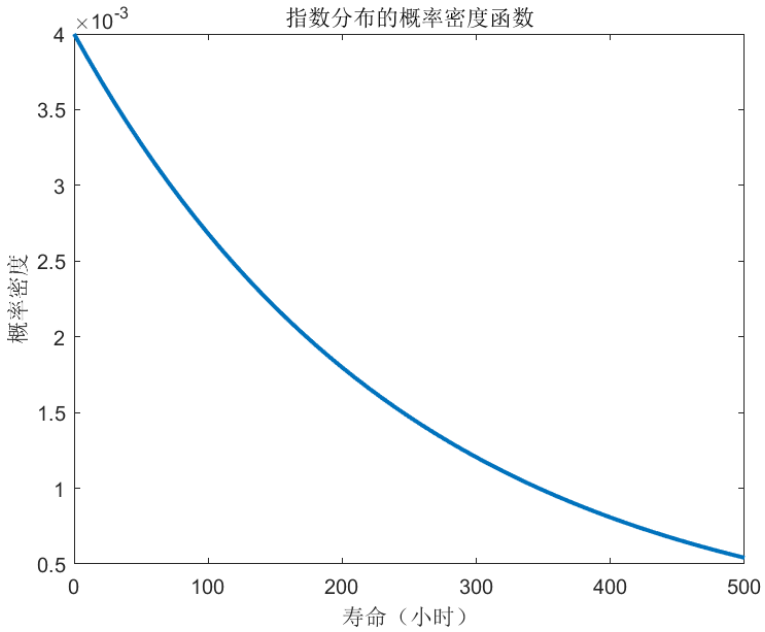

Figure 1. Probability density function of exponential distribution
图 1. 指数分布的概率密度函数图

结果如图 1 显示估计的指数分布参数 lambda 为 0.004。这表明，对于所观测的元件寿命数据，利用极大似然估计得到的指数分布参数 λ 的估计值为 0.004。并且通过绘制的概率密度函数图可以看出，该参数估计对应的指数分布曲线的形状。

3.2. 线性回归

在现实生活中，一个变量通常并非单一因素所决定，而是多个变量共同作用的结果，这些变量的最优组合往往能更有效地预测或估计因变量的变化。在拥有两个及两个以上的自变量时，就应该进行多元线性回归。

多元线性回归模型的一般形式为：

$$Y=\beta_0+\beta_1 x+\cdots+\beta_{p-1}x+u_i$$

式中 β_0 为常数项，$\beta_1,\cdots,\beta_{p-1}$ 为回归系数，x 为解释变量，u_i 为随机误差。

•观测数据

多元线性模型就是有多个未知数 β。

$$y=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix},\quad X=\begin{bmatrix}1&x_{11}&\cdots&x_{1,p-1}\\1&x_{21}&\cdots&x_{2,p-1}\\\vdots&\vdots&\ddots&\vdots\\1&x_{n1}&\cdots&x_{n,p-1}\end{bmatrix},\quad \beta=\begin{bmatrix}\beta_0\\\beta_1\\\vdots\\\beta_{p-1}\end{bmatrix},\quad e=\begin{bmatrix}u_1\\u_2\\\vdots\\u_n\end{bmatrix}$$

•确定回归系数

•求经验方程

设 $\hat{\beta}=\left(\hat{\beta}_0,\hat{\beta}_1,\cdots,\hat{\beta}_{p-1}\right)'$ 为 β 的一种估计，则经验方程是：

$$Y=\hat{\beta}_0+\hat{\beta}_1 X_1+\cdots+\hat{\beta}_{p-1}X_{p-1}$$

调用 Matlab 中 regress 函数：[b, bint, r, rint, stats] = regress (Y, X, alpha)，其本质为最小二乘法。

案例 2：我们想研究影响中国新能源汽车的销量的因素有哪些，根据现有资料得出以下可能因素：充电桩数量、新能源汽车续航里程、人均可支配收入。收集了 2012~2022 年的数据，其中 x_1 为充电桩数量，x_2 为新能源汽车续航里程，x_3 为人均可支配收入；

```
x1=[0.5 1.2 2.0 3.8 8.0 15.0 21.4 33.0 51.8 71.5 261.7 520.0]';
x2=[100.00 110.00 120.36 146.93 168.86 190.52 231.00 315.34 364.63 396.63 416.17 447.72]';
x3=[15412 17366 18311 20167 21966 23821 25974 28228 30733 32189 35128 36883]';
y=[0.8 1.2 1.8 7.5 33.1 50.7 77.7 125.6 120.6 136.7 352.1 688.7]';
X=[ones(size(x1)),x1,x2,x3];
[b,bint,r,rint,stats]=regress(y,X);
b,bint,stats
rcoplot(r,rint)
```

结果

```
b=
−65.5254
1.1260
0.0415
0.0037
bint=
−215.9462      84.8954
```

1.0015　　1.2506

−0.5984　　0.6814

−0.0083　　0.0157

stats=

0.9939　　433.9865　　0.000　　339.44

即 $\beta_0 = -65.5254$　$\beta_1 = 1.1260$　$\beta_2 = 0.0415$　$\beta_3 = 0.0037$。

β_0 的置信区间为[−215.9462, 84.8954]，β_1 的置信区间为[1.0015, 1.2506]，β_2 的置信区间为[−0.5984, 0.6814]，β_3 的置信区间为[−0.0083, 0.0157]。

R^2=0.9939　　F=433.9865　　p=0.0000

由 p < 0.05，可知回归模型 $y = -65.5254 + 1.126x_1 + 0.0415x_2 + 0.0037x_3$ 成立；R^2 = 0.9939 接近 1，说明回归拟合效果好。

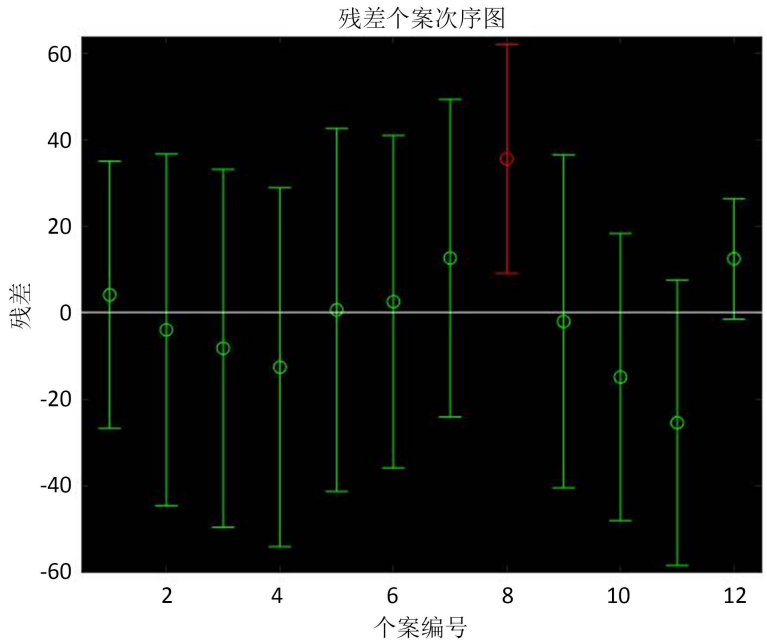

Figure 2. Linear regression residuals
图 2. 线性回归残差图

如图 2 所示，基本所有的数据都在 0 轴上下，说明我们的预测值与真实值都比较接近，第 8 组数据脱离了范围，可以看作为异常值，整体拟合程度好。

3.3. 大数定律

大数定理指的是在独立同分布的随机变量序列中，随着样本量的增加，样本均值会趋近于总体均值的稳定性质。简单来说，大数定理告诉我们，随着试验次数的增加，样本平均值会逐渐接近总体均值[3]。

案例 3：假设有一个罐子里装有红色和蓝色的球，红色球的比例为 0.6，蓝色球的比例为 0.4。我们进行了 100 次抽样，并记录了每次抽样得到的红色球的比例。

假设随机变量 X 表示抽样得到的红色球比例，X 的取值范围为 0 到 1。因为每次抽样都是独立的，所以可以将 $X_1, X_2, \cdots, X_{100}$ 看作是独立同分布的随机变量序列。根据大数定理，随着抽样次数的增加，这 100 次抽样得到的红色球比例的样本均值会趋近于总体的红色球比例 0.6。

假设抽样次数为 n，那么随机变量序列 $X_1, X_2, \cdots, X_n$ 的样本均值可以表示为：

$$\bar{X}_n = \left(X_1 + X_2 + \cdots + X_n\right)/n$$

根据大数定理，随着 n 的增加，$\bar{X}_n$ 会逐渐接近总体的红色球比例 0.6。这个模型可以帮助我们理解大数定理在实际问题中的应用，以及随机变量序列的稳定性质。

我们使用 Matlab 进行模拟实验并计算抽样得到的红色球比例的平均值。

```
%模拟实验
num_samples=100;%?总共进行 100 次抽样
red_ratio=0.6;%?红色球比例
blue_ratio=0.4;%?蓝色球比例
sample_results=binornd(1,red_ratio,1,num_samples);
%二项分布模拟每次抽样结果
red_ratio_mean=cumsum(sample_results)./(1:num_samples);
%计算每次抽样得到的红色球比例的平均值
%绘制红色球比例的平均值随抽样次数的变化
figure
plot(1:num_samples,red_ratio_mean,'b','LineWidth',2)
xlabel('抽样次数')
ylabel('红色球比例的平均值')
title('红色球比例的平均值随抽样次数的变化')
```

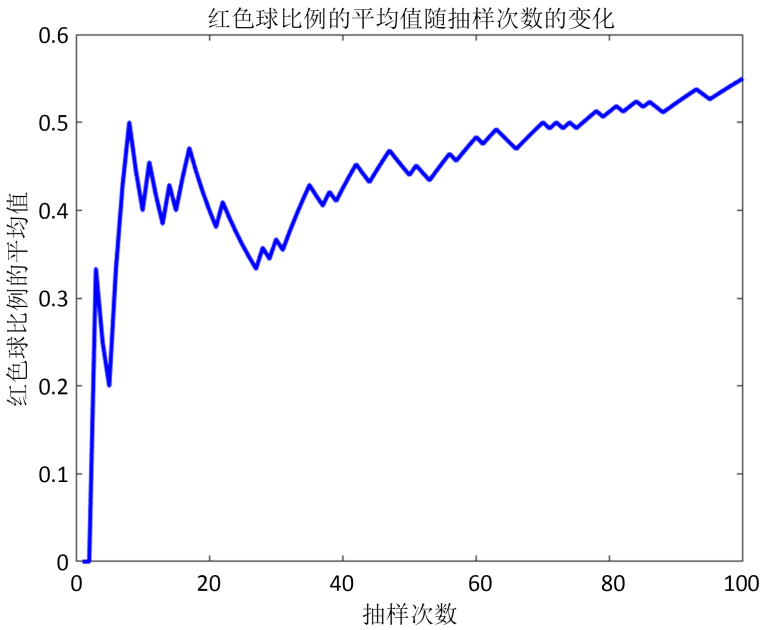

Figure 3. Variation of the mean value of the proportion of red balls with sampling times

图 3. 红色球比例的平均值随抽样次数的变化图

如图 3，可以观察到红色球比例的平均值随着抽样次数的增加而变化的趋势。结果可能会显示随着抽样次数的增加，红色球比例的平均值逐渐接近 0.6，即红色球的真实比例。这可以帮助我们了解抽样实

验在统计推断中的应用，以及红色球比例随着抽样次数变化的趋势。这符合大数定律的预期，即随着实验次数的增加，样本平均值会收敛于总体平均值。

3.4. 聚类分析

3.4.1. 层次聚类(简单连接法)

合理的聚类方式应使得同一族群内的观测尽可能地“相似”，但不同族群之间有明显区分。相似度通常使用欧式距离来度量。

在进行层次聚类时，可以从凝聚法和分离法两个方向入手。

凝聚法的主要思路在于，初始时将每个样本点视为独立的聚类，随后不断迭代地将距离最近的两个聚类进行合并，这便是凝聚的含义。此过程将持续进行，直至满足设定的迭代终止条件。

分离法：所有数据点都被视为一个类别，然后根据它们之间的相似性逐步分裂成更小的类别，直到每个数据点都成为一个单独的类别。

简单连接法：定义族群之间的距离为两族群中相隔最近的两个体间的距离。

案例 4：设有 5 个公司职员 a_1,a_2,a_3,a_4,a_5，它们的销售业绩有二维变量 (t_1,t_2) 描述，见下表 1。

Table 1. Statistical table of employee sales performance
表 1. 员工销售业绩统计表

公司职员	t_1(销售量)/十件	t_2(回收款项)/千元
a_1	2	1
a_2	2	2
a_3	4	3
a_4	5	4
a_5	3	6

记公司职员 $a_i\,(i=1,2,3,4,5)$ 的销售业绩为 (t_{i1},t_{i2})。使用绝对值距离来测量点与点之间的距离，使用简单连接法来测量类与类之间的距离，即

$$d(a_i,a_j)=\sum_{k=1}^{2}\left|t_{ik}-t_{jk}\right|,\ D(G_p,G_q)=\min_{\substack{a_i\in G_p\\ a_j\in G_q}}\left\{d(a_i,a_j)\right\}.$$

有距离公式 $d(a_i,a_j)$，可以算出距离矩阵：

$$\begin{array}{c} \\ a_1\\ a_2\\ a_3\\ a_4\\ a_5 \end{array}\begin{array}{c} \begin{array}{ccccc} a_1 & a_2 & a_3 & a_4 & a_5 \end{array}\\ \begin{bmatrix} 0 & 1 & 4 & 6 & 6\\ & 0 & 3 & 5 & 5\\ & & 0 & 2 & 2\\ & & & 0 & 0\\ & & & & 0 \end{bmatrix} \end{array}$$

首先，将全部元素归为一类 $K_1=\{a_1,a_2,a_3,a_4,a_5\}$。这样每一个类平台高度都为 0，这时

$$D(G_p,G_q)=d(a_p,a_q)。$$

然后，取平台高度为 1，将 a_1,a_2 重新组成一个新的类 k_6，这时分类如下：

$$K_2=\{k_6,a_3,a_4,a_5\}$$

接着，取平台高度为 2，将 a_3, a_4 重新组成一个新的类 k_7，这时分类如下：

$$K_3 = \{k_6, k_7, a_5\}$$

然后，取平台高度为 3，将 k_6, k_7 重新组成一个新类 k_8，这时分类如下：

$$K_4 = \{k_8, a_5\}$$

最后，取平台高度为 4，将 k_8, a_5 重新组成一个新类 k_9，这时分类如下：

$$K_5 = \{k_9\}$$

Mandist 函数可以求矩阵列向量之间的两两绝对值距离，正好可以用于简单连接法；计算的 Matlab 程序如下：

```
clc,clear
a=[2,1;2,2;4,3;5,4;3,6];
[j,k]=size(a);
b=zeros(j);
b=mandist(a');                % mandis 用来求矩阵列向量族之间两两绝对值距离
b=tril(b);                    %截取下三角的元素
nb=nonzeros(b);               %将 b 中的零元素去掉，非零元素按列排列
nb=unique(nb)                 %将重复的非零元素去掉
fori=1:j-1
nb_min=min(nb);
    [row,col]=find(b==nb_min);tu=union(row,col);      % col 和 row 归为一类
tu=reshape(tu,1,length(tu));                          % 将数组 tu 变成行向量
fprintf('第%d 次合成，平台高度为%d 时的分类结果为：%s\n',...
i,nb_min,int2str(tu));
nb(nb==nb_min)=[];                                    % 把已经归类的元素删除
if length(nb)==0;
break
end
end
```

其运行结果为：

第 1 次合成，平台高度为 1 时的分类结果为：1　2；
第 2 次合成，平台高度为 2 时的分类结果为：3　4；
第 3 次合成，平台高度为 3 时的分类结果为：2　3；
第 4 次合成，平台高度为 4 时的分类结果为：1　3　4　5。

3.4.2. *k*-均值聚类

一旦某个个体被确定归属某一族群，其后续将固定在该族群内，不再具备加入其他族群的资格。这便决定了层次聚类它仅仅能达到局部最优，一般不可能达到全局最优。所以便有了非层次聚类的分割法，分割法中最常用的方法为 *k*-Means 算法。

k-均值法试图寻找 *k* 个族群 $(G_1, \cdots, G_k)$ 的划分方式，使得划分后的族群内方差(WGSS)最小。

$$\mathrm{WGSS}=\sum_{l=1}^{k}\sum_{i\in G_l}\sum_{j=1}^{p}\left(y_{ij}-\bar{y}_j^{(l)}\right)^2=\sum_{l=1}^{k}\sum_{i\in G_i}\left(y_i-\bar{y}^{(l)}\right)'\left(y_i-\bar{y}^{(l)}\right)$$

其中族群 G_l 中第 j 个变量的均值 $\bar{y}_j^{(l)}=\sum_{i\in G_l}y_{ij}/n_i$，$y_i=\left(y_{i1},\cdots,y_{ip}\right)'$，$\bar{y}^{(l)}=\left(\bar{y}_1^{(l)},\cdots,\bar{y}_p^{(l)}\right)'$

在给定 k 的情况下，我们可以通过以下步骤将 WGSS 最小化：

首先，我们选取 k 个初始点，将其视作族群的代表或称为“种子”。

每个个体都被划分到距离它最近的种子所在的族群中。

当族群划分完成后，我们计算每个新族群的中心点，并将其作为新的“种子”。

接着，我们重复进行步骤 2 和 3，直到族群不再发生显著的变化或满足终止条件。

k 的选取：

我们所需要的是选择一个使得 WGSS 足够小(但不是最小)的 k 值。可以用 WGSS 的碎石图来寻找最优的 k，通常选取拐点为最优的 k。

有了 k 的选取，我们可以通过下列方法去选择初始种子：

在相互间隔超过指定最小距离的前提下，随机选择 k 个个体。

选择数据集前 k 个相互间隔超过某指定最小距离的个体。

选择 k 个相互激励最远的个体。

选择 k 个等距网格点，可能不是数据集的点。

这里我们可以将 k 均值法作为层次聚类的改进。我们首先用层次聚类法进行聚类，再以各个族群的质心作为 k 均值法的初始种子。这样就可以载层次聚类的基础上，使得个体有被重新分配到其他族群的机会。

Matlab 提供了 2 个基本的生成伪随机数的函数 rand 和 randn 其中 rand 函数用来生成[0, 1]上均匀分布随机数，randn 函数用来生成标准正态分布随机数。由[0, 1]上均匀分布随机数可以生成其他分布的随机数，由标准正态分布随机数可以生成一般正态分布随机数。在不指明分布的情况下，通常所说的随机数是指[0, 1]上均匀分布的随机数。在不同版本的 Matlab 中，rand 函数有几种通用的调用格式[4]，如下表 2 所示：

Table 2. Call format for the rand function
表 2. Rand 函数的调用格式[4]

调用格式	说明
Y = rand	生成一个服从[0, 1]上均匀分布的随机数
Y = rand(n)	生成 n × n 的随机数矩阵
Y = rand(m,n)	生成 m × n 的随机数矩阵
Y = rand([m,n])	生成 m × n 的随机数矩阵
Y = rand(m,n,p,…)	生成 m × n × p × …的随机数矩阵或数组
Y = rand([m,n,p,…])	生成 m × n × p × …的随机数矩阵或数组
Y = rand(size(A))	生成与矩阵或者数组 A 具有相同大小的随机数矩阵或者数组

下面调用 rand 函数生成随机数，用来作为 k-均值法的数据：

在 Matlab 中，k-均值法可以直接调用 kmeans 函数来计算，比如我们要将 rand 函数随机生成的 100 个二维数据点分为三组，我们就可以调用为 kmeans (data, 3)。

```
clc
num = rand(100,2);
[it,c] = kmeans(num, 3);
figure;
hold on;
scatter(num(it==1,1), num(it==1,2),'r');
scatter(num(it==2,1), num(it==2,2),'g');
scatter(num(it==3,1), num(it==3,2),'b');
scatter(c(:,1), c(:,2),'k','filled');
legend('Cluster 1','Cluster 2','Cluster 3','Centroids');
```

从图中可以看出随机生成的 100 个点已经被分为了红绿蓝三个部分，图中黑点为各个部分的质心。

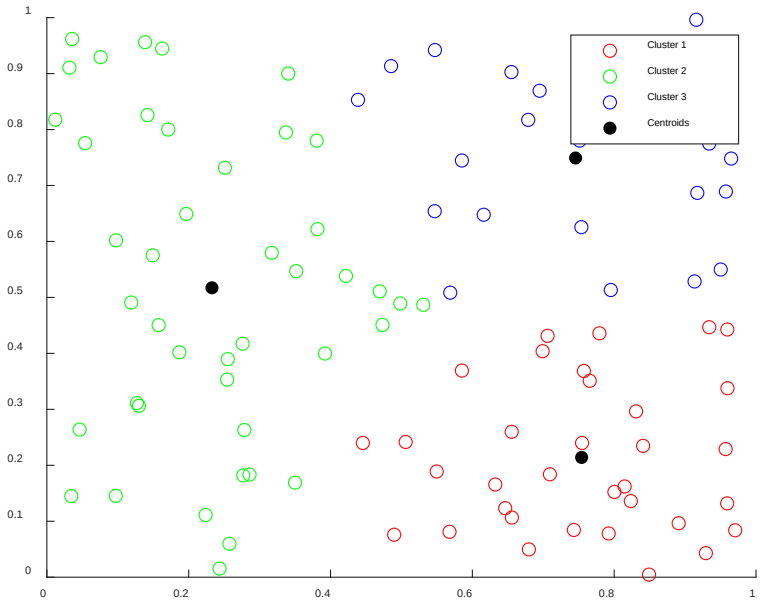

Figure 4. K-means clustering scatter distribution
图 4. k-均值聚类散点分布图

从图 4 中可以看出随机生成的 100 个点已经被分为了红绿蓝三个部分，图中黑点为各个部分的质心。

4. 结语

基于 Matlab 的基础上，我们对概率论与数理统计的具体实例进行了处理和分析，模拟随机现象，并对结果进行统计分析，直观便捷地展现了问题的结果，提升了解决实际问题的效率，让该课程的“有用论”在学生心中油然而生。未来，我们期待进一步利用 Matlab 的优势，深入研究和探索概率论与数理统计中的各种问题，使教学质量得到进一步提升。

基金项目

2024 年广西高校中青年教师科研基础能力提升项目：分数阶微分方程边值问题解的研究(编号：2024KY1727)；2021 年校级重点应用型课程建设项目：概率论与数理统计；2023 年大学生创新训练项目：基于 MATLAB 软件的概率论与数理统计问题的研究与算法(编号：S202313644049)。

参考文献

[1] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012.

[2] 何晓飞, 张洋. Matlab 在统计学与数据分析中的应用[M]. 北京: 清华大学出版社, 2014.

[3] 吴冲锋, 徐乃忠. Matlab 在数学建模中的应用[M]. 北京: 电子工业出版社, 2016.

[4] 王鹏, 董平轩, 冉玉梅. 《概率论与数理统计》教学中随机数的 Matlab 实现[J]. 电脑知识与技术, 2023, 19(2): 138-140+161. https://doi.org/10.14004/j.cnki.ckt.2023.0106