第八次作业 _ 参考答案

郑盼盼

2024-11-22

目录

2.43 设 $\xi \sim B(1, 0.3)$，对于 $i = 1, 2, \ldots, 10000$，分别使用 R 语言模拟 ξ 的重复观测结果 $10i$ 次，并计算相应的重复观测结果的算术平均值 $\bar{x}_i$；绘制 $(i, \bar{x}_i)$ 的折线图，分析该折线随着 i 增大变化的趋势及原因

- **答**：这道题目要求我们计算不同样本量下的样本均值，并将结果使用折线图绘制出来。考察：(1) 对于 `for` 循环的使用；(2) 对于向量的基础操作，即如何往向量中添加元素；(3) 基础绘图
  - **方法 1**：直接生成 10×10000 个随机数，在 `for` 循环中每次计算前 $10i$ 个元素的算术平均值，并存成向量。

```
# 2.43 程序
i <- 1:10000                         # 根据题意定义向量 i=1,2,...,10000
xi <- rbinom(10*max(i),1,0.3)    # 生成 10*10000 个服从 B(1,0.3) 的随机数
x_bar <- c()                         # 定义一个空向量，用于存储不同样本量的均值
# 通过 for 循环，每次计算 xi 前 10*j 个元素的均值，并作为最后一个元素添加给
↪ 向量 x_bar
for (j in i){
  x_bar <- c(x_bar, mean(xi[1:10*j]))
}
plot(i, x_bar, type="l") # 使用 plot 绘图，第一个是横坐标向量，第二个是纵坐
↪ 标向量，type="l" 即设置其为折线图
abline(h=0.3, col="red", lty=2, lwd=2) # 绘制参考线
```

  - **方法二**：通过 `for` 循环，每次生成 $10i$ 个样本，并计算这 $10i$ 个样本的均值，然后存成向量。

```
# 2.43 程序
i <- 1:10000     # 根据题意定义向量 i=1,2,....,10000
x_bar <- c()     # 定义一个空向量，用于存储不同样本量的均值
# 通过 for 循环，每次生成样本量为 10*j 的随机数，并计算其均值，作为最后一个
↪ 元素添加给向量 x_bar
```

```r
for (j in i){
  xi <- rbinom(10*j,1,0.3)
  x_bar <- c(x_bar, mean(xi))
}
plot(i, x_bar, type="l")    # plot 绘制折线图
abline(h=0.3, col="red", lty=2, lwd=2) # 添加总体期望的参考线
```

2.44 设 ξ 服从以 $1000, 10$ 和 50 为参数的超几何分布，对于 $i = 1, 2, \ldots, 10000$，分别使用 R 语言模拟 ξ 的重复观测结果 $10i$ 次，并计算相应的重复观测结果的算术平均值 $\bar{x}_i$；绘制 $(i, \bar{x}_i)$ 的折线图，分析该折线随着 i 增大变化的趋势及原因

- **答**：这道题目要求我们计算不同样本量下的样本均值，并将结果使用折线图绘制出来。考察：(1) 对于 `for` 循环的使用；(2) 对于向量的基础操作，即如何往向量中添加元素；(3) 基础绘图
  - **方法 1**：直接生成 10×10000 个随机数，在 `for` 循环中每次计算前 $10i$ 个元素的算术平均值，并存成向量。

```r
# 2.44 程序
i = 1:10000                          # 根据题意定义向量 i=1,2,....,10000
xi <- rhyper(max(i) * 10, 10, 990, 50)# 生成 10000*10 个服从参数为 1000,10 和
↪  50 的超几何分布的随机数
x_bar <- c()          # 定义一个空向量用于存储不同样本量下的算术平均值
# 利用 for 循环计算前 10*j 个元素的算术平均值，并作为最后一个元素添加到向量
↪  x_bar 中
for (j in i){
  x_bar <- c(x_bar, mean(xi[1:j*10]))
}
plot(i, x_bar, type="l") # 通过 plot 函数，将设置 type="l" 绘制折线图
abline(h=0.5, lty=2, lwd=2, col="red")
```

  - **方法二**：通过 `for` 循环，每次生成 $10i$ 个样本，并计算这 $10i$ 个样本的均值，然后存成向量。

```r
# 2.44 程序
i = 1:10000        # 根据题意定义向量 i=1,2,...,10000
x_bar <- c()       # 定义一个空向量用于存储不同样本量下的均值
# 通过 for 循环，每次生成不同样本量的样本，并计算均值，将其作为最后一个元素
↪  添加到向量 x_bar 中
for (j in i){
  xi <- rhyper(j * 10, 10, 990, 50)
  x_bar <- c(x_bar, mean(xi))
```

```
}
plot(i, x_bar, type="l")    # 利用 plot 绘制折线图
abline(h=0.5, lty=2, lwd=2, col="red") # 绘制参考线 (总体数学期望)
```

2.49 设 $\xi \sim B(10000, 0.3)$，写出蒙特卡洛方法近似计算 $F_\xi(5555)$ 的 R 语言程序代码，并将计算结果与 pbinom(5555,10000,0.3) 的计算结果相比较，分析近似计算的精度和重复观测次数之间的关系。(这道题出的不是很好。

- **答**：这道题让我们用蒙特卡洛模拟的方法来估计已知分布的分布函数值，我们可以借用经验分布函数的方法来进行估计，设我们有 n 个对于该分布的观测值 $X_1, \dots, X_n$，则其经验分布函数 $F_n(x)$ 为

$$F_n(x) = \frac{n(\{i : 1 \le i \le n,\, X_i \le x\})}{n}$$

其中分子 $n(\{i : 1 \le i \le n,\, X_i \le x\})$ 表示 n 个样本 X_i 中小于 x 个数，在 r 语言中我们可以十分简单的计算得到。根据大数定律，有：

$$\lim_{n\to\infty} F_n(x) = F(x) = \mathbb{P}(X \le x)$$

我们可以根据如上概念进行蒙特卡洛模拟。

```
# 2.49 程序
sim_freq <- c() # 定义空向量用于存储不同样本量下的蒙特卡洛模拟结果
true_prob <- pbinom(5555, 10000,0.3) # 计算实际的分布函数值
# 在不同样本量 (10,100,1000) 下进行蒙特卡洛模拟
for (n in c(10,100,1000)) {
  sim_B <- rbinom(n, 10000,0.3)  # 生成 n 个服从 B(10000,0.3) 的随机数 (即对总体
↪  B(10000,0.3) 进行 n 次重复观测)
  tmp_freq <- mean(sim_B <= 5555) # 计算经验分布, 即 n 次重复观测中小于等于 5555
↪  的次数占总观测次数的比例; 需要了解: 向量的关系运算, 逻辑型和数值型之间转换
↪  的规则。
  sim_freq <- c(sim_freq,tmp_freq) # 将模拟的结果存储到向量 sim_freq 中
  abs_diff <- abs(true_prob - tmp_freq) # 计算模拟结果和实际结果之间的误差
  cat(" 重复观测数为 ",n," 的估计结果为 ", tmp_freq, " 其和真实值之间的差距 ",
   ↪ abs_diff,"\n") # 打印出结果
}
```

```
## 重复观测数为  10  的估计结果为  1  其和真实值之间的差距  0
## 重复观测数为  100  的估计结果为  1  其和真实值之间的差距  0
## 重复观测数为  1000  的估计结果为  1  其和真实值之间的差距  0
```

2.52 设 $\xi \sim U(-10,10)$，写出蒙特卡洛方法近似计算 $F_\xi(0.5)$ 的 R 语言程序代码，并将计算结果与 `punif(0.5,-10,10)` 的计算结果相比较，分析近似计算的精度和重复观测次数之间的关系。

- **答**：和上道题类似，这道题让我们用蒙特卡洛模拟的方法来估计已知分布的分布函数值，我们可以借用经验分布函数的方法来进行估计，设我们有 n 个对于该分布的观测值 $X_1,\dots,X_n$，则其经验分布函数 $F_n(x)$ 为

$$F_n(x) = \frac{n(\{i: 1 \le i \le n,\, X_i \le x\})}{n}$$

其中分子 $n(\{i: 1 \le i \le n,\, X_i \le x\})$ 表示 n 个样本 X_i 中小于 x 个数，在 r 语言中我们可以十分简单的计算得到。根据大数定律，有：

$$\lim_{n\to\infty} F_n(x) = F(x) = \mathbb{P}(X \le x)$$

我们可以根据如上概念进行蒙特卡洛模拟。

```r
# 2.52 程序
sim_freq <- c()       # 定义空向量用于存储不同样本量下的蒙特卡洛模拟结果
true_prob <- punif(0.5, -10, 10) # 计算实际的分布函数值
# 在不同样本量 (10,100,1000) 下进行蒙特卡洛模拟
for (n in c(10,100,1000,10000)) {
  sim_U <- runif(n, -10,10) # 生成 n 个服从 U(-10,10) 的随机数 (即对总体
  ↪  U(-10,10) 进行 n 次重复观测)
  tmp_freq <- mean(sim_U <= 0.5) # 计算经验分布 0.5; 需要了解: 向量的关系运算，逻
  ↪  辑型和数值型之间转换的规则。
  sim_freq <- c(sim_freq,tmp_freq) # 将模拟的结果存储到向量 sim_freq 中
  abs_diff <- abs(true_prob - tmp_freq) # 计算模拟结果和实际结果之间的误差
  cat(" 重复观测数为",n," 的估计结果为", tmp_freq, " 其和真实值之间的差距",
    ↪  abs_diff,"\n") # 打印结果
}
```

```
## 重复观测数为 10 的估计结果为 0.7 其和真实值之间的差距 0.175
## 重复观测数为 100 的估计结果为 0.45 其和真实值之间的差距 0.075
## 重复观测数为 1000 的估计结果为 0.525 其和真实值之间的差距 0
## 重复观测数为 10000 的估计结果为 0.5249 其和真实值之间的差距 1e-04
```

2.54 试用蒙特卡洛方法估算定积分 $\int_0^1 x^2 e^{x^2}\, dx$

- **答** 参考 P88 页的方法：此处我们的积分区间是 $[0,1]$ 被积函数 $f(x) = x^2 e^{x^2}$。对于最一般的积分形式：

$$\int_a^b f(x)\, dx$$

蒙特卡洛模拟的步骤如下：

1. 生成积分区间内服从均匀分布的 n 个随机数，即 $Y_i \sim U(a,b)\,(i=1,2,\ldots,n)$
2. 根据被积函数 $f(x)$ 对第一步所得的随机数进行计算，得到 n 个随机变量 $f(Y_i)\,(i=1,2,\ldots,n)$
3. 根据如下公式进行计算:

$$\int_a^b f(x)\,dx \approx (b-a)\frac{1}{N}\sum_{i=1}^{N} f(Y_i)$$

其实就是计算上一步得到的 n 个随机变量 $f(Y_i)$ 的算术平均值后乘以积分上下限的差 $b-a$

```
# 2.54 程序
Y <- runif(10000, 0,1) # 生成 10000 个服从 U(0,1) 的随机数
f_Y = Y^2 * exp(Y^2)  # 计算 f(Y)
estim_int <- mean(f_Y) * (1-0) # 计算 f(Y) 的算术平均值，并乘以积分上下限的差
↪  (1-0)
true_int <- integrate(function(x) x^2 * exp(x^2), 0,1) # 使用 integrate 函数精确计
↪  算
cat(" 蒙特卡洛模拟的结果为", estim_int, "，实际的计算结果为", true_int$value) #
↪  比较二者差异
```

```
## 蒙特卡洛模拟的结果为 0.6319227 ，实际的计算结果为 0.627815
```

2.60 若大学生中男生的身高均值为 173.61 cm，标准差为 4.96 cm。随机选取 16 名男大学生。在 RStudio 中，写出应用中心极限定理近似计算这 16 名大学生的身高均值落在区间 $(168.65, 178.57)$（单位：cm）内的概率的程序代码，并给出近似计算的结果。

- **答** 此题目主要考察了中心极限定理的应用：根据中心极限定理，对于 16 名男大学生的身高均值 $\bar{X}$，其近似于正态分布

$$\bar{X} \sim N(\mathbb{E}(X), D(X)/16)$$

其中

- $\mathbb{E}(X) = 173.61$
- $D(X) = 4.96^2$

题目要求我们计算 $\mathbb{P}(168.65 \le \bar{X} \le 178.57)$，根据正态分布的性质，我们可以将 $\bar{X}$ 进行标准化为标准正态分布：

$$\mathbb{P}\left(\frac{168.65-\mathbb{E}(X)}{\sqrt{D(X)/16}} \le \frac{\bar{X}-\mathbb{E}(X)}{\sqrt{D(X)/16}} \le \frac{178.57-\mathbb{E}(X)}{\sqrt{D(X)/16}}\right) = \Phi\left(\frac{178.57-\mathbb{E}(X)}{\sqrt{D(X)/16}}\right) - \Phi\left(\frac{168.65-\mathbb{E}(X)}{\sqrt{D(X)/16}}\right)$$

此处 $\Phi(x)$ 为标准正态分布的分布函数在 x 点的函数值。

```
# 2.54 程序
mu = 173.61      # 总体均值
sigma2 = 4.96^2  # 总体方差
```

```
mean_sigma2 = sigma2 / 16 # 样本均值的方差
# 可以不标准化，直接根据样本均值服从分布的参数计算
up = pnorm(178.57, mu, sqrt(mean_sigma2))
low = pnorm(168.65, mu, sqrt(mean_sigma2))
up - low

# 标准化后进行计算
z_up = (178.57 - mu)/(sqrt(mean_sigma2))
z_down = (168.65 - mu)/(sqrt(mean_sigma2))
pnorm(z_up) - pnorm(z_down) # pnorm 不写其他参数就默认为标准正态分布的参数：均值
↪ 为 0，标准差为 1
```