第四次上机作业

斯蓬 220810332

2024-11-19

目录

1 第六章课后习题编程部分

1.1 Task6.1

```
# 参数设定
set.seed(123)
```

```r
n1 <- 100
n2 <- 200
mu1 <- 1
mu2 <- 1
sigma1 <- 1
sigma2 <- 2

# 生成独立样本
X <- rnorm(n1, mean = mu1, sd = sigma1)
Y <- rnorm(n2, mean = mu2, sd = sigma2)

# 方差估计
var_X <- var(X)
var_Y <- var(Y)
var_ratio <- var_X / var_Y

# Bootstrap 方法
B <- 1000
bootstrap_var_ratios <- numeric(B)

for (b in 1:B) {
  X_star <- sample(X, size = n1, replace = TRUE)
  Y_star <- sample(Y, size = n2, replace = TRUE)
  bootstrap_var_ratios[b] <- var(X_star) / var(Y_star)
}

# 计算偏差和方差
bias <- mean(bootstrap_var_ratios) - var_ratio
variance <- var(bootstrap_var_ratios)

cat("6.1 结果:\n")
```

```
## 6.1 结果:
```

```
cat(" 真实方差比值:", var_ratio, "\n")
```

```
## 真实方差比值: 0.2246849
```

```
cat("Bootstrap 估计偏差:", bias, "\n")
```

```
## Bootstrap估计偏差: 0.001239208
```

```
cat("Bootstrap 估计方差:", variance, "\n")
```

```
## Bootstrap估计方差: 0.001547056
```

1.2 Task6.5

```
# 设置随机种子
set.seed(123)

# 参数设定
n <- 200             # 样本大小
sigma <- 2           # 正态分布标准差
B <- 1000            # Bootstrap 重采样次数

# 生成数据
X <- rnorm(n, mean = 0, sd = sigma)

# 定义方差估计量公式
var_estimator <- function(x) {
  n <- length(x)
  return(sum((x - mean(x))^2) / n)  # 偏差估计量公式
}

# Bootstrap 方法计算偏差
bootstrap_var <- numeric(B)
for (b in 1:B) {
```

```r
  resample <- sample(X, size = n, replace = TRUE)  # Bootstrap 重采样
  bootstrap_var[b] <- var_estimator(resample)
}
bootstrap_bias <- mean(bootstrap_var) - var_estimator(X)  # Bootstrap 偏差

# Jackknife 方法估计 Bootstrap 偏差的方差
jackknife_bias <- numeric(n)
for (i in 1:n) {
  # 去掉第 i 个样本
  jackknife_sample <- X[-i]

  # 对 Jackknife 样本进行 Bootstrap 偏差计算
  bootstrap_var_j <- numeric(B)
  for (b in 1:B) {
    resample_j <- sample(jackknife_sample, size = n - 1, replace = TRUE)
    bootstrap_var_j[b] <- var_estimator(resample_j)
  }
  # 计算去掉第 i 个样本后的 Bootstrap 偏差
  jackknife_bias[i] <- mean(bootstrap_var_j) - var_estimator(jackknife_sample)
}

# 估计 Bootstrap 偏差的方差
jackknife_bias_variance <- (n - 1) / n * sum((jackknife_bias - mean(jackknife_bias))^2)

# 输出结果
cat("Bootstrap 偏差估计量:\n")
```

```
## Bootstrap 偏差估计量:
```

```r
cat("  偏差 =", bootstrap_bias, "\n")
```

```
##   偏差 = -0.0002434071
```

```
cat("\nJackknife 方法估计 Bootstrap 偏差的方差:\n")
```

```
##
## Jackknife 方法估计 Bootstrap 偏差的方差:
```

```
cat("  方差 =", jackknife_bias_variance, "\n")
```

```
##   方差 = 0.03283107
```

2 第七章例题复现部分

2.1 7.2.3

```
set.seed(220810332)
n=1000
m=600
x=rnorm(n,mean=2,sd=2)
sx=sum(x[1:m])
sx2=sum(x[1:m]^2)
max.iter=100
hmu=rep(0,max.iter)
hsigma2=rep(0,max.iter)
hmu[1]=0
hsigma2[1]=1

for (i in 1:max.iter) {
  s1=sx+(n-m)*hmu[i]
  s2=sx2+(n-m)*(hmu[i]^2+hsigma2[i])
  hmu[i+1]=s1/n
  hsigma2[i+1]=s2/n-hmu[i+1]^2
  if (abs(hmu[i+1]-hmu[i])<1e-8
  & abs(hsigma2[i+1]-hsigma2[i])<1e-8) break
}
```

```r
mu_est <- hmu[i + 1]
sigma2_est <- hsigma2[i + 1]

cat("Estimated Mean (mu):", mu_est, "\n")
```

```
## Estimated Mean (mu): 1.980367
```

```r
cat("Estimated Variance (sigma^2):", sigma2_est, "\n")
```

```
## Estimated Variance (sigma^2): 3.546603
```

2.2 7.3.1

```r
# 设置随机数种子以保证结果可复现
set.seed(1)

# 生成观测数据
N <- 1000  # 样本总数
A <- rbinom(N, 1, 0.7)  # 生成服从伯努利分布的 0-1 变量，概率为 0.7
X <- rnorm(N, 1, sqrt(3)) * (A == 0) + rnorm(N, 10, 1) * (A == 1)  # 生成两类混合正态数据

# 最大迭代次数
max.iter <- 100

# 初始化参数存储向量
a1 <- c(); a2 <- c()      # 混合系数
mu1 <- c(); mu2 <- c()    # 均值
sig1 <- c(); sig2 <- c()  # 方差

# 初始化参数
a1[1] <- 0.4; a2[1] <- 0.6       # 初始混合系数
mu1[1] <- 2; mu2[1] <- 5         # 初始均值
sig1[1] <- 4; sig2[1] <- 2       # 初始方差
```

```r
# EM 算法迭代
for (i in 1:max.iter) {
  # E 步：计算每个数据点属于两个分布的后验概率
  p1 <- dnorm(X, mu1[i], sqrt(sig1[i]))  # 第一个正态分布的概率密度
  p2 <- dnorm(X, mu2[i], sqrt(sig2[i]))  # 第二个正态分布的概率密度

  d1 <- a1[i] * p1 / (a1[i] * p1 + a2[i] * p2)  # 数据点属于第一个分布的责任概率
  d2 <- a2[i] * p2 / (a1[i] * p1 + a2[i] * p2)  # 数据点属于第二个分布的责任概率

  # M 步：根据责任概率更新参数
  a1[i + 1] <- mean(d1)  # 更新混合系数 1
  a2[i + 1] <- mean(d2)  # 更新混合系数 2

  mu1[i + 1] <- sum(X * d1) / sum(d1)  # 更新第一个正态分布的均值
  mu2[i + 1] <- sum(X * d2) / sum(d2)  # 更新第二个正态分布的均值

  sig1[i + 1] <- sum((X - mu1[i + 1])^2 * d1) / sum(d1)  # 更新第一个正态分布的方差
  sig2[i + 1] <- sum((X - mu2[i + 1])^2 * d2) / sum(d2)  # 更新第二个正态分布的方差

  # 检查收敛条件
  if (abs(mu1[i + 1] - mu1[i]) < 1e-8 &&
      abs(mu2[i + 1] - mu2[i]) < 1e-8 &&
      abs(sig1[i + 1] - sig1[i]) < 1e-8 &&
      abs(sig2[i + 1] - sig2[i]) < 1e-8) break
}

# 输出最终估计的参数
cat(" 最终参数估计结果：\n")
```

```
## 最终参数估计结果：
```

```r
cat(" 混合系数：a1 =", a1[i + 1], ", a2 =", a2[i + 1], "\n")
```

```
## 混合系数：a1 = 0.3046471 , a2 = 0.6953529
```

```
cat(" 均值: mu1 =", mu1[i + 1], ", mu2 =", mu2[i + 1], "\n")
```

```
## 均值: mu1 = 0.9885232 , mu2 = 9.992192
```

```
cat(" 方差: sigma1^2 =", sig1[i + 1], ", sigma2^2 =", sig2[i + 1], "\n")
```

```
## 方差: sigma1^2 = 3.197053 , sigma2^2 = 1.084535
```

3 第五章例题复现部分

3.1 5.7

```
set.seed(220810332)
n = 100   # 样本大小
N = 1000  # 重复实验次数
I11 = NULL
I21 = NULL
I22 = NULL

# 模拟实验
for (i in 1:N) {
  X = rnorm(n, mean = 0, sd = 1)  # 生成样本 X
  Y = rnorm(n, mean = 0, sd = 1)  # 生成样本 Y

  I11[i] = var(X)  # 样本方差（同样本）
  I21[i] = n * var(X) / (n - 1)  # 修正后的样本方差（同样本）
  I22[i] = n * var(Y) / (n - 1)  # 修正后的样本方差（不同样本）
}

# 比较同样本不同估计方法的精度
VDif1 = var(I11 - I21)

# 比较不同样本不同估计方法的精度
```

```
VDif2 = var(I11 - I22)

# 输出结果
cat("E(I11) =", mean(I11), " Var(I11) =", var(I11), "\n")
```

```
## E(I11) = 1.005493  Var(I11) = 0.01904802
```

```
cat("E(I21) =", mean(I21), " Var(I21) =", var(I21), "\n")
```

```
## E(I21) = 1.01565  Var(I21) = 0.01943478
```

```
cat("E(I22) =", mean(I22), " Var(I22) =", var(I22), "\n")
```

```
## E(I22) = 1.017355  Var(I22) = 0.02057543
```

```
cat(" 相同样本，VDif1 =", VDif1, "\n")
```

```
## 相同样本，VDif1 = 1.943478e-06
```

```
cat(" 不同样本，VDif2 =", VDif2, "\n")
```

```
## 不同样本，VDif2 = 0.0388018
```

4 第五章课后习题编程部分

4.1 Task5.3

```
# 设置随机数种子，保证结果可重复
set.seed(123)

# 样本大小和实验重复次数
n <- 500        # 样本大小
N <- 1000       # 重复实验次数

# 初始化向量存储均值估计和控制变量法估计的结果
mean_estimates <- numeric(N)
```

```
control_estimates <- numeric(N)

# 理论期望值
E_X <- 1  # Exp(1) 分布的均值为 1

# 模拟实验
for (i in 1:N) {
  # 生成指数分布样本
  X <- rexp(n, rate = 1)  # 参数 rate = 1, 对应 Exp(1) 分布

  # 生成控制变量 Y, 其分布为 U(0, 1)
  Y <- runif(n, min = 0, max = 1)  # 控制变量 Y

  # 均值估计
  mean_estimates[i] <- mean(X)

  # 控制变量法估计
  control_estimates[i] <- mean(X) - cov(X, Y) / var(Y) * (mean(Y) - 0.5)
}

# 计算均值估计和控制变量法估计的均值和方差
mean_est_mean <- mean(mean_estimates)
mean_est_var <- var(mean_estimates)

control_est_mean <- mean(control_estimates)
control_est_var <- var(control_estimates)

# 输出结果
cat(" 均值估计法: \n")
```

```
## 均值估计法:
```

```
cat("  E(X) 的估计均值 =", mean_est_mean, "\n")
```

```
##   E(X) 的估计均值 = 0.9996929
```

```r
cat("  E(X) 的估计方差 =", mean_est_var, "\n\n")
```

```
##   E(X) 的估计方差 = 0.001991974
```

```r
cat(" 控制变量法：\n")
```

```
## 控制变量法：
```

```r
cat("  E(X) 的估计均值 =", control_est_mean, "\n")
```

```
##   E(X) 的估计均值 = 0.9997225
```

```r
cat("  E(X) 的估计方差 =", control_est_var, "\n")
```

```
##   E(X) 的估计方差 = 0.001990947
```