R related homework

尹超

中国科学院大学，北京 100049

Carter Yin

University of Chinese Academy of Sciences, Beijing 100049, China

2024.8 – 2025.1

序言

R related homework

目录

序言 I

目录 II

1 2024.11.6 1

第 1 次作业　2024.11.6

Let's analyze and explain the code line by line, then compute the result for the provided sample size and for a larger sample size (115).

Here's the original code with explanations:

```
mysamplesize=15              # 设置样本量为 15
alpha=0.05                   # 设置显著性水平为 0.05（对应 95% 置信区间）
mytheta=1                    # 设置指数分布的偏移参数为 1

mysample=rexp(mysamplesize, rate=1) + mytheta  # 生成一个大小为 15 的样本，来源于偏移参数为 my
mymle=sort(mysample)[1]      # 取样本的最小值作为偏移参数的最大似然估计（MLE）

bootstrapsize=1000           # 设置自助法样本数量为 1000
bootstrapestimates=rep(0, bootstrapsize) # 初始化一个数组，用于存储自助法的估计值

# 生成自助样本，并计算每个样本的最小值
for(ii in 1:bootstrapsize){
    bootstrapsample=rexp(mysamplesize, rate=1) + mymle # 生成新的自助法样本，并使用 mymle 作为
    bootstrapestimates[ii]=sort(bootstrapsample)[1]    # 存储样本的最小值（自助法估计）
}

bootstrapquantiles=sort(bootstrapestimates - mymle) # 计算自助法估计与 MLE 之间差值的分布

# 计算置信区间的下分位数和上分位数
lowerquantile=bootstrapquantiles[round(bootstrapsize * alpha * 0.5)]
upperquantile=bootstrapquantiles[round(bootstrapsize * (1 - alpha * 0.5))]

# 计算置信区间的下界和上界
lowerbound=mymle - upperquantile
upperbound=mymle - lowerquantile
```

结果

样本量为 15 时，0.95 置信区间为 $(1.010786, 1.242433)$

样本量为 115 时，0.95 置信区间为 $(0.9734284, 1.001642)$

可以看出，随着样本量的增大，置信区间变得更窄，说明对 θ 的估计更加精确。

```
> mysamplesize=15;
> alpha=0.05
> mytheta=1;
> mysample=rexp(mysamplesize,rate=1)+mytheta;
> mymle=sort(mysample)[1];
>
> bootstrapsize=1000;
> bootstrapestimates=rep(0,bootstrapsize);
> for(ii in 1:bootstrapsize){
+ bootstrapsample=rexp(mysamplesize,rate=1)+mymle;
+ bootstrapestimates[ii]=sort(bootstrapsample)[1];
+ }
> bootstrapquantiles=sort(bootstrapestimates-mymle)
>
> lowerquantile=bootstrapquantiles[round(bootstrapsize*alpha*0.5)]
> upperquantile=bootstrapquantiles[round(bootstrapsize*(1-alpha*0.5))]
> lowerbound=mymle-upperquantile
> upperbound=mymle-lowerquantile
>
> lowerbound
[1] 1.010786
> upperbound
[1] 1.242433
> lowerquantile
[1] 0.001889443
> upperquantile
[1] 0.2335364
>
```

```
> mysamplesize=115;
> alpha=0.05
> mytheta=1;
> mysample=rexp(mysamplesize,rate=1)+mytheta;
> mymle=sort(mysample)[1];
> bootstrapsize=1000;bootstrapestimates=rep(0,bootstrapsize);
> for(ii in 1:bootstrapsize){
+ bootstrapsample=rexp(mysamplesize,rate=1)+mymle;
+ bootstrapestimates[ii]=sort(bootstrapsample)[1];
+ }
> bootstrapquantiles=sort(bootstrapestimates-mymle)
> lowerquantile=bootstrapquantiles[round(bootstrapsize*alpha*0.5)]
> upperquantile=bootstrapquantiles[round(bootstrapsize*(1-alpha*0.5))]
> lowerbound=mymle-upperquantile
> upperbound=mymle-lowerquantile
> lowerbound
[1] 0.9734284
> upperbound
[1] 1.001642
> lowerquantile
[1] 0.0001751082
> upperquantile
[1] 0.02838861
>
```

```
> # 参数设定
> mysamplesize = 15                # 样本量
> alpha = 0.05                     # 显著性水平 0.05（对应 95% 置信区间）
> mytheta = 1                      # 偏移参数设为 1
>
> # 生成原始样本
> set.seed(0)  # 设置随机种子，确保结果可重复
> mysample = rexp(mysamplesize, rate=1) + mytheta  # 生成偏移后的指数分布样本
> T2_hat = min(mysample)           # 计算矩估计 T2，即原始样本的最小值
>
> # 自助法参数设定
> bootstrapsize = 1000             # 自助法样本数
> bootstrapestimates = rep(0, bootstrapsize) # 初始化数组用于存储每个自助法样本的 T2 估计
>
> # 自助法过程
> for (ii in 1:bootstrapsize) {
+     bootstrapsample = sample(mysample, size = mysamplesize, replace = TRUE) # 有放回地抽取自助法样本
+     bootstrapestimates[ii] = min(bootstrapsample)  # 计算自助法样本的 T2 矩估计
+ }
>
> # 计算置信区间
> bootstrapquantiles = sort(bootstrapestimates - T2_hat) # 计算自助法估计与原估计的差值
> lowerquantile = bootstrapquantiles[round(bootstrapsize * alpha * 0.5)]
> upperquantile = bootstrapquantiles[round(bootstrapsize * (1 - alpha * 0.5))]
>
> # 计算置信区间的上下界
> lowerbound = T2_hat - upperquantile
> upperbound = T2_hat - lowerquantile
>
> # 输出结果
> list(lowerbound = lowerbound, upperbound = upperbound)
$lowerbound
[1] 1.095554

$upperbound
[1] 1.139795

>
> lowerquantile
[1] 0
> upperquantile
[1] 0.04424132
>
```