**云南大学数学与统计学院上机实验报告 （ 3 ）**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **实验课程名** | | | **抽样技术与应用实验** | | | | | | | | | **成绩** | |  | |
| **学号** |  | | | | **姓名** | | **枫叶** | **专业** | | **统计学** | **年级** | | **2021级** | | |
| **实验项目名称** | | | | **不等概率抽样** | | | | **日期** | | **2023.4.18** | **实验时间** | | | |  |
| **指导教师** | | **潘东东老师** | | | | **实验地址（室）** | | | **格物楼3507** | | | | | | |
| **教师评语** | |  | | | | | | | | | | | | | |
|  | |  | | | | | | | | | | | | | |

1. **实验目的：**

1、学会编写函数，实现pps抽样。

2、使用sample（），rmultionm（）函数实现pps抽样。

1. **实验内容:**

1、编写函数generate.pps（），参变量可以输入不等概率向量，也可以输入度量总体单元规模或大小的向量值，另一个参变量是产生的pps样本量。

2、使用sample（），rmultnorm（）函数从表6.1所示的多项分布（概率为Zi）中抽取30个样本值。（pps抽样）学号作为随机数种子。

3、自编R程序实现表6.4的布鲁尔抽样。

1. **实验环境:**

**R软件**

1. **实验步骤及实验结果:**

1、程序实验：R代码

1PPS抽样程序

generate.pps <- function(p,n){

  library(sampling)

  num <- length(p)

  index <- 1:num

  is.wholenumber <- #判别是否存在整数

    function(x, tol = .Machine$double.eps^0.5)  abs(x - round(x)) < tol

  result <- c()

  for (j in 1:n){

    if(min(p)>1){ #规模度量 #什么都不做，如果加了next会导致整个循环结束，别问我怎么知道的

  }else{ #概率向量 #其实这个判别没用啊，但是懒得改了喵

    while(T){

      p <- p\*10

      if(all(is.wholenumber(p)) == T) break #存在小数就乘10

    }

  }

  M <- sum(p)

  c <- sample(M,1)

  options(digits=1)

  c <- runif(1,1,M) #生成随机整数

breaks <- c(0)

  for (i in 1:num) {

    breaks[i+1] <- sum(p[1:i])

  }

  group <- cut(0:M,breaks = breaks, right=T,include.lowest = T) #代码法的分组

  levels(group) <- 1:num

  result[j] <- group[c+1] #返回一个因子水平，表示抽到的是第几个单位，存入result

  }

result

}

2

（1）sample

library(sampling)

set.seed(172)

p <- 1:10

prob <- c(0.08,0.1,0.17,0.06,0.24,0.09,0.05,0.07,0.04,0.10)

sample(p,30,replace = T,prob = prob)

（2）rmultinorm

library(sampling)

set.seed(172)

p <- 1:10

prob <- c(0.08,0.1,0.17,0.06,0.24,0.09,0.05,0.07,0.04,0.10)

rmultinom(1,30,prob = prob)[,1]

3

#假定输入的是度量单元规模的向量

Brewer.sub <- function(m,n){

    result <- cumulation <- c()

    M <- sum(m)

    Z <- m/M

    p <- Z\*(1-Z)/(1-n\*Z)

      for (j in 1:length(m)){

        cumulation[j] <- sum(p[1:j])

      }

    r <- runif(1)

    a <- sum(p)\*r #这里参考了149页例6.3

    judge <- a > cumulation

    k <- 1

      while (T){

      if (judge[k] == F) break #换了新的找落点方式喵

      k <- k+1

      }

      k #k表示落到的是第几个区间，也即抽中的是第几个样本

}

Brewer <- function(m,n){

  result <- c()

  for (i in 1:n){

    a <- result[i] <- Brewer.sub(m,n)

    m[a] <- 0 #去掉抽中的单位

    n <- n-i

  }

  result

}

1. **程序结果：截屏**

1

输入概率向量

p <- c(0.1,0.2,0.2,0.4,0.1)

generate.pps(p,n=10)



输入规模度量

p <- c(1,2,2,2,3,4)

generate.pps(p,n=10)



2

（1）sample



每次抽抽到第几个

（2）rmultinorm



每个各抽到了几次

3

m <- c(17,12,15,24,10) #Brewer方法要求Z\_i<1/n

n <- 3

Brewer(m,n)



**五、结果分析**