

L07-离散型随机变量的模拟

郑盼盼

2024-11-10

目录

7.1 模拟 X 的重复观测值	1
7.2 计算 X 的分位数	2
7.3 大数定律与算术平均值的估计	3
Questions	3

7.1 模拟 X 的重复观测值

若随机变量 X 的密度矩阵如下所示

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_m \\ p_1 & p_2 & \cdots & p_m \end{pmatrix}$$

则可以利用如下代码模拟这个随机变量的 n 次重复观测:

```
sample(c(x1,...,xm), size=n, replace=T, prob=c(p1,...,pm))
```

例如 当 X 的矩阵密度为

$$\begin{pmatrix} -0.5 & -1 & 1.1 & 1.2 \\ 0.2 & 0.3 & 0.25 & 0.25 \end{pmatrix} \quad (7.1)$$

时, 模拟它的 10000 次重复观测的程序为

```
x <- c(-0.5, -1, 1.1, 1.2)
p <- c(0.2, 0.3, 0.25, 0.25)
rX <- sample(x, size=10000, replace=T, prob=p)
mean(rX == -0.5)
```

```
## [1] 0.1977
```

```
mean(rX == -1)
```

```
## [1] 0.3001
```

```
mean(rX == 1.1)
```

```
## [1] 0.2524
```

```
mean(rX == 1.2)
```

```
## [1] 0.2498
```

7.2 计算 X 的分位数值

如何使用 R 计算如 (7.1) 所示的分布的分位数值? 1. 将第一行的随机变量取值改为递增的密度矩阵:

$$\begin{pmatrix} -1 & -0.5 & 1.1 & 1.2 \\ 0.3 & 0.2 & 0.25 & 0.25 \end{pmatrix} \quad (7.2)$$

2. 利用如上密度矩阵计算其 0.4 分位数

```
x <- c(-1, -0.5, 1.1, 1.2); p <- c(0.3, 0.2, 0.25, 0.25)
```

```
tmpAlpha <- 0.4
```

```
for (i in 1:length(x)) {
```

```
  tmpF <- sum(p[1:i])      # 计算前 i 个元素的密度之和
```

```
  # 若密度 (概率) 之和大于等于 tmpAlpha, 返回其对应的随机变量取值, 并终止循环
```

```
  if (tmpF >= tmpAlpha) {
```

```
    xAlpha <- x[i]
```

```
    break
```

```
  }
```

```
}
```

```
xAlpha
```

```
## [1] -0.5
```

7.3 大数定律与算术平均值的估计

若某高中一年级有 N 名学生，他们的身高分别为 x_1, \dots, x_N ，如何估计这些学生的平均身高？

$$a = \frac{1}{N} \sum_{i=1}^N x_i = \sum_{i=1}^N x_i \times \frac{1}{N}$$

若将 $\frac{1}{N}$ 看成是概率 $\mathbb{P}(X = x_i)$ ，则 X 的密度矩阵为：

$$\begin{pmatrix} x_1 & \cdots & x_N \\ \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix}$$

数学期望

$$\mathbb{E}(X) = \sum_{i=1}^N x_i \mathbb{P}(X = x_i) = \sum_{i=1}^N x_i \times \frac{1}{N}$$

根据大数定律，有 $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = a$ 的概率为 1。可以用

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n X_i$$

估计学生的平均身高 a 。

R 语言实现 可以利用随机模拟的方法考察上述方法估计算术平均值的效果：

```
x <- c(rnorm(98, mean=1.60, sd=0.16), rnorm(86, mean=1.70, sd=0.17))
tmpId <- sample(1:length(x), size=100000, replace="T")
xSample <- x[tmpId]
hatMean <- mean(xSample)
trueMean <- mean(x)
hatMean
```

```
## [1] 1.647603
```

```
abs(trueMean - hatMean)
```

```
## [1] 4.532178e-05
```

Questions

1. 若 X 的密度矩阵为

$$\begin{pmatrix} 0 & -1 & 1 \\ 0.5 & 0.1 & 0.4 \end{pmatrix}$$

写出模拟 X 的 100 次重复观测值的程序代码，并用这些重复观测值估计 X 的数学期望，给出估计的误差（估计误差等于估计值减真值）