

E03 上机实验 03

郑盼盼

2024-10-24

目录

1. 从 1:10 中用取后放回的方法依次抽取 8 个数，计算事件 $A = \{1, 2, 3, 4, 5\}$ 的频率。

```
A <- c(1,2,3,4,5) # 根据事件 A 的定义，定义一个向量 A，其包含五个元素：1, 2, 3,
↪ 4, 5 (也可以使用 1:5)
samps <- sample(1:10, 8, T) # 从 1:10 中使用取后放回的方法抽取 8 个数组成一个样
↪ 本
freq = sum(samps %in% A) / 8 # %in% 用于判断样本中每个元素是否属于 A, sum() 用于
↪ 求和 (用于计算出样本中属于 A 的元素的个数)
print(freq)
```

```
## [1] 0.75
```

2. 从 1:10 中用取后不放回的方法依次抽取 8 个数，计算事件 A 的频率。

```
A <- c(1,2,3,4,5)
samps <- sample(1:10, 8, F) # 除了此处使用不放回方法进行抽样外，其余均和 1. 相同
freq = sum(samps %in% A) / 8
print(freq)
```

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

3. 将 1. 重复 100 次，得到 100 个频率值 x_1, x_2, \dots, x_{100} ；将 2. 重复 100 次得到 100 个频率值 u_1, u_2, \dots, u_{100} ；用红色将 100 个点：

$$(x_1, 1), (x_2, 1), \dots, (x_{100}, 1)$$

绘制在直角坐标系中，用蓝色将 100 个点

$$(u_1, 2), (u_2, 2), \dots, (u_{100}, 2)$$

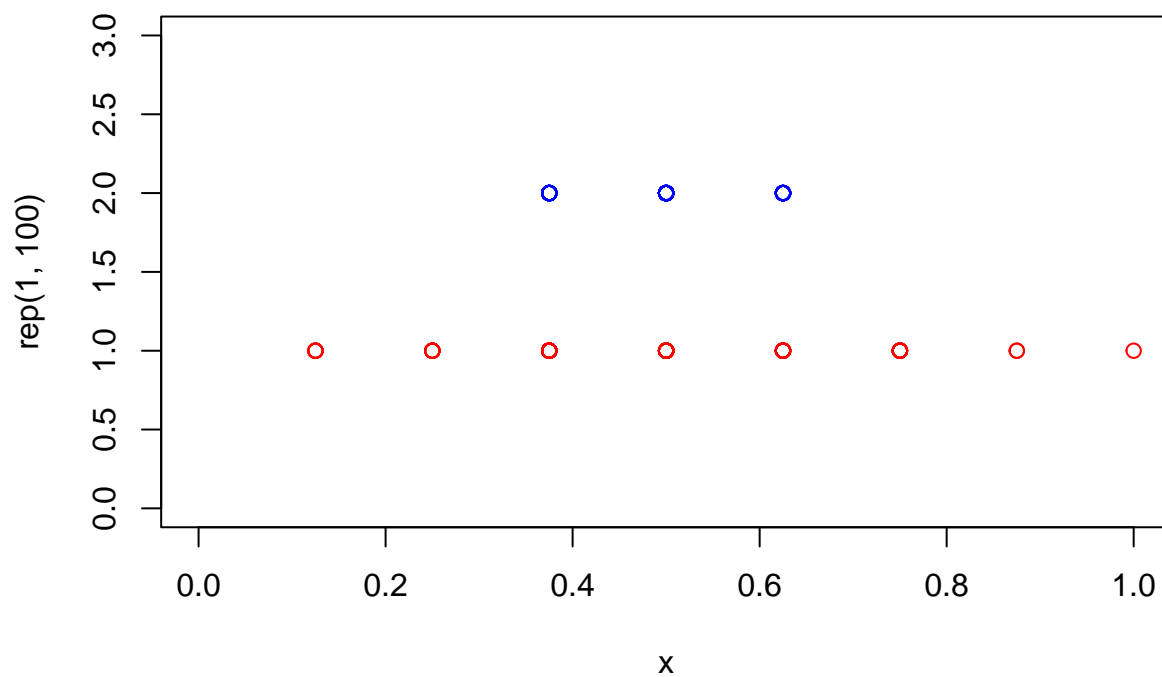
绘制在同一坐标系中（用函数 `points()` 添加新的点）

```
x <- c() # 定义空向量 x, 用于存储频率 x1,x2,...,x100
u <- c() # 定义空向量 u, 用于存储频率 u1,u2,...,u100

# 根据题意重复 100 次, 因此采用从 1 到 100 进行遍历的 for 循环
for (i in 1:100) {
  samp1 <- sample(1:10, 8, T) # 采用放回的方式得到第一个样本
  samp2 <- sample(1:10, 8, F) # 采用不放回的方式得到第二个样本
  x <- c(x, sum(samp1 %in% A) / 8) # 计算样本 1 的频率, 并作为最后一个元素添加到
  ↪ 向量 x 上
  u <- c(u, sum(samp2 %in% A) / 8) # 计算样本 2 的频率, 并作为最后一个元素添加到
  ↪ 向量 u 上
}

# 以 x 为横坐标, 1 为纵坐标, 绘制散点图
plot(x, rep(1,100), # 由于绘图需要保证横坐标向量和纵坐标向量的长度一致, 我们使
  ↪ 用 rep(1,100), 构建一个长度为 100 元素全为 1 的向量
  xlim = c(0,1), # 限制 x 轴的范围为 0 到 1; xlim: x's limitations
  ylim=c(0,3),   # 限制 y 轴的范围为从 0 到 3; ylim: y's limitations
  col="red"      # 设置颜色为红色; col: color
)

# points() 用于在已经绘制的图上面添加新的散点 (若使用 plot 会新建一张图, 无法将
  ↪ 两种散点绘制在一张图上)
points(u, rep(2,100),
  col="blue")
```



4. 在 3. 中所绘制图上两种颜色点的中心位置坐标都是什么? 请解释人们更喜欢用 2. 中的频率值来估计 $\mathbb{P}(A)$ 的原因

答: 中心位置的坐标为 0.5 即事件 A 发生的概率; 但是 2. 的分布相较而言更为集中。因此, 其更适合用于估计 $\mathbb{P}(A)$