E06

郑盼盼

2024-10-24

示例代码

```r
m <- 10000 # 观测次数
n <- 1:m
barX <- n # 用于存放均值
myCol <- c("blue", "green", "yellow", "grey", "pink", "black") # 线的颜色
for (i in 1:6) {
  x <- runif(m) # 模拟 m 个 U(0,1) 随机数
  for (j in 1:m) {
    barX[j] <- mean(x[1:j]) # 计算前 j 次模拟的随机数的均值
  }
  if (i == 1) {
    # 创建绘图画面，在其上绘制第 1 次模拟的 myCol[1] 色的均值变化折线
    plot(n, barX, col = myCol[1], type = "l", lwd = 1.5, ylim = c(0, 1), ylab =
     ↪  "arithmetic mean")
  } else{
    lines(n, barX, col = myCol[i], lwd = 1.5)
  }
  if (i == 6) {
    abline(h = 0.5, col = "red", lwd = 2)
  }
}
```

1. 请说明示例程序代码中的内外两个循环语句的功能，解释为什么要用条件语句。

   答：

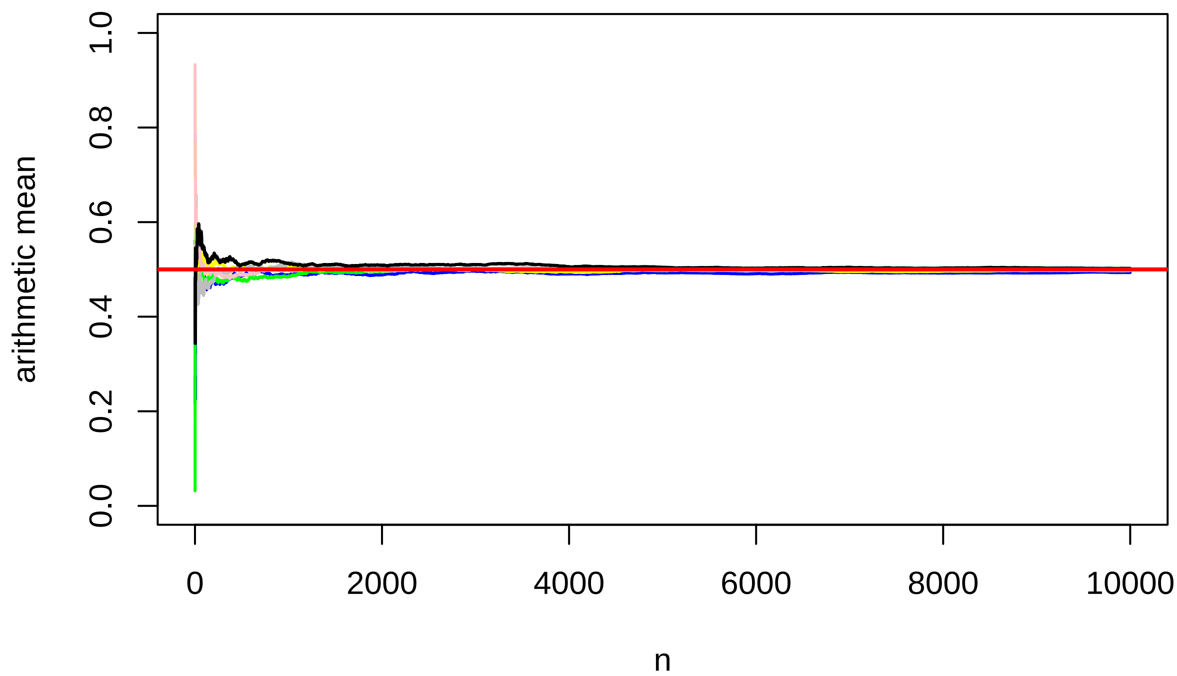
   - 两个循环语句的功能：外层循环 `for (i in 1:6)` 迭代 6 次，每次通过 runif(m) 生成 m 个在区间 [0, 1] 上均匀分布的随机数（模拟六次均匀分布随机数的生成过程），并存储在向量 `x` 中。绘

制不同颜色的均值变化线。每次迭代时，通过不同的颜色绘制均值变化折线，颜色是从预先定义好的颜色向量 `myCol` 中依次取出；内层循环 `for (j in 1:m)` 遍历每次模拟的随机数，计算从第 1 次到第 j 次生成的随机数的均值，并将结果存储在 `barX[j]` 中。

- 条件语句的功能：
  1. **区分第一次绘图和后续绘制**：在 R 的基础绘图系统中，plot() 函数不仅绘制数据点或线，还会创建整个绘图区域和坐标轴。因此，在第一次（i == 1）绘图时，必须使用 plot() 函数来初始化图像，并设置坐标轴、刻度、图例等内容，在第 2 次及以后的绘图中，图像的坐标系已经由 plot() 函数生成，因此只需要使用 lines() 函数在已有的坐标系上绘制新的曲线。lines() 不会重新创建坐标轴，它只是在现有的绘图区域内添加新的线条。
  2. **添加红色基准线（水平线）**：这里的条件语句确保红色的基准线只在最后一次绘图完成后添加。这样做是因为，你希望在所有模拟均值曲线绘制完成后，最后再添加这条期望值为 0.5 的水平线。如果在每次循环中都调用 abline()，则会多次重复绘制这条线，影响图像的美观。

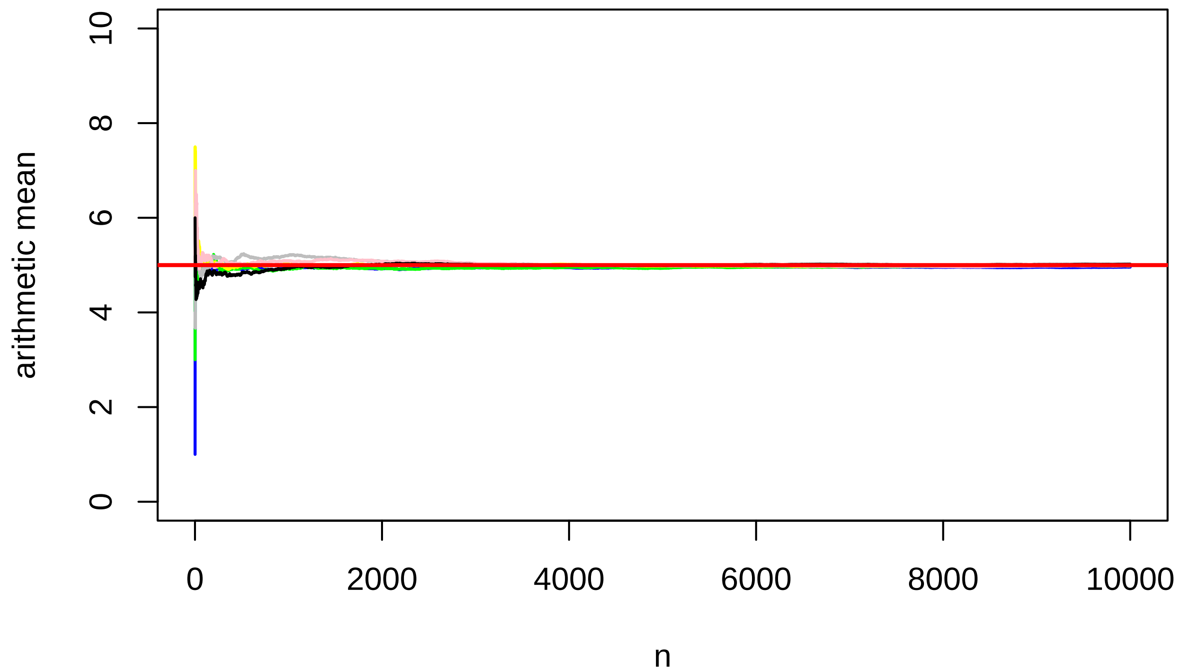
2. 如何避免示例程序代码中的条件语句，并给出相应代码。

```r
m <- 10000  # 观测次数
n <- 1:m    # 用来表示每次的观测次数
barX <- n   # 用于存放每次模拟后的均值
myCol <- c("blue", "green", "yellow", "grey", "pink", "black")  # 线的颜色

# 先创建空图，设置好坐标轴、y 轴标签和 ylim 范围
plot(n, barX, type = "n", ylim = c(0, 1), ylab = "arithmetic mean")

# 使用 for 循环绘制 6 条均值变化线
for (i in 1:6) {
  x <- runif(m)  # 模拟 m 个 U(0,1) 随机数
  for (j in 1:m) {
    barX[j] <- mean(x[1:j])  # 计算前 j 次模拟的随机数的均值
  }
  # 使用 lines() 绘制每条均值变化线
  lines(n, barX, col = myCol[i], lwd = 1.5)
}

# 绘制红色水平线 y = 0.5
abline(h = 0.5, col = "red", lwd = 2)
```



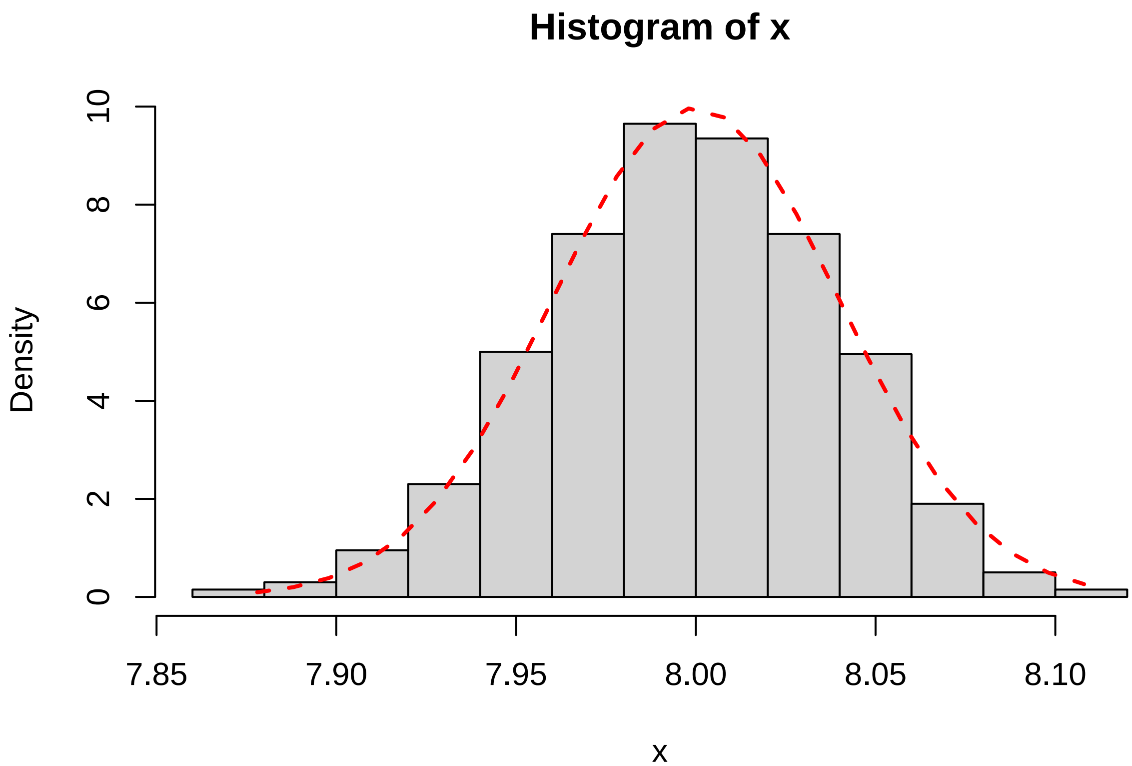
3. 用 R 语言程序代码模拟研究 $X \sim P(5)$ 的重复观测数据的算术平均值与观测次数之间的关系，总结规律。

```
m <- 10000  # 观测次数
n <- 1:m    # 用来表示每次的观测次数
barX <- n   # 用于存放每次模拟后的均值
myCol <- c("blue", "green", "yellow", "grey", "pink", "black")  # 线的颜色

# 先创建空图，设置好坐标轴、y 轴标签和 ylim 范围
plot(n, barX, type = "n", ylim = c(0, 10), ylab = "arithmetic mean")

# 使用 for 循环绘制 6 条均值变化线
for (i in 1:6) {
  x <- rpois(m, 5)  # 模拟 m 个 P(5) 随机数
  for (j in 1:m) {
    barX[j] <- mean(x[1:j])  # 计算前 j 次模拟的随机数的均值
  }
  # 使用 lines() 绘制每条均值变化线
  lines(n, barX, col = myCol[i], lwd = 1.5)
}

# 绘制红色水平线 y = 0.5
abline(h = 5, col = "red", lwd = 2)
```



4. 试通过 1000 次模拟观测数据估计 $X \sim B(10, 0.8)$ 的数学期望 $\mathbb{E}(X)$，讨论估计结果是否为随机变量，并判断估计误差的取值范围。

```
m = 1000
n = 1000
x <- replicate(n, mean(rbinom(m, 10, 0.8)))
hist(x, freq=F)
xx <- seq(min(x), max(x), by=0.01)
y <- dnorm(xx, 8, sqrt(1.6/m))
lines(xx, y,
      lty=2,
      col="red",
      lwd=2)
```



5. 已知数学考试的平均成绩（5 分制）为 4.10，标准差为 0.3，估算 100 名成绩之和小于 400 的概率。
答：假设数学考试的平均成绩 X 服从正态分布 $X \sim N(4.10, 0.3^2)$，令随机变量 X' 为 100 次观测值 $X_1, X_2, \cdots, X_{100}$ 之和，根据中心极限定理，100 个成绩的算术平均值 $\bar{X}_{100}$：

$$\bar{X}_{100} \sim N(4.10, 0.3^2/100)$$

而 $X' = 100\bar{X}_{100}$，因此：

$$X' \sim N(410, 100 \times 0.3^2)$$

因此，我们可以将分数总和进行标准化：

$$\mathbb{P}(X' < 400) = \mathbb{P}\left(\frac{X' - 410}{3} < -10/3\right) \approx \Phi(-3.33)$$

可以利用 R 语言计算得：

```r
pnorm(-3.33)
```

```
## [1] 0.0004342299
```