

## E04 上机实验 04

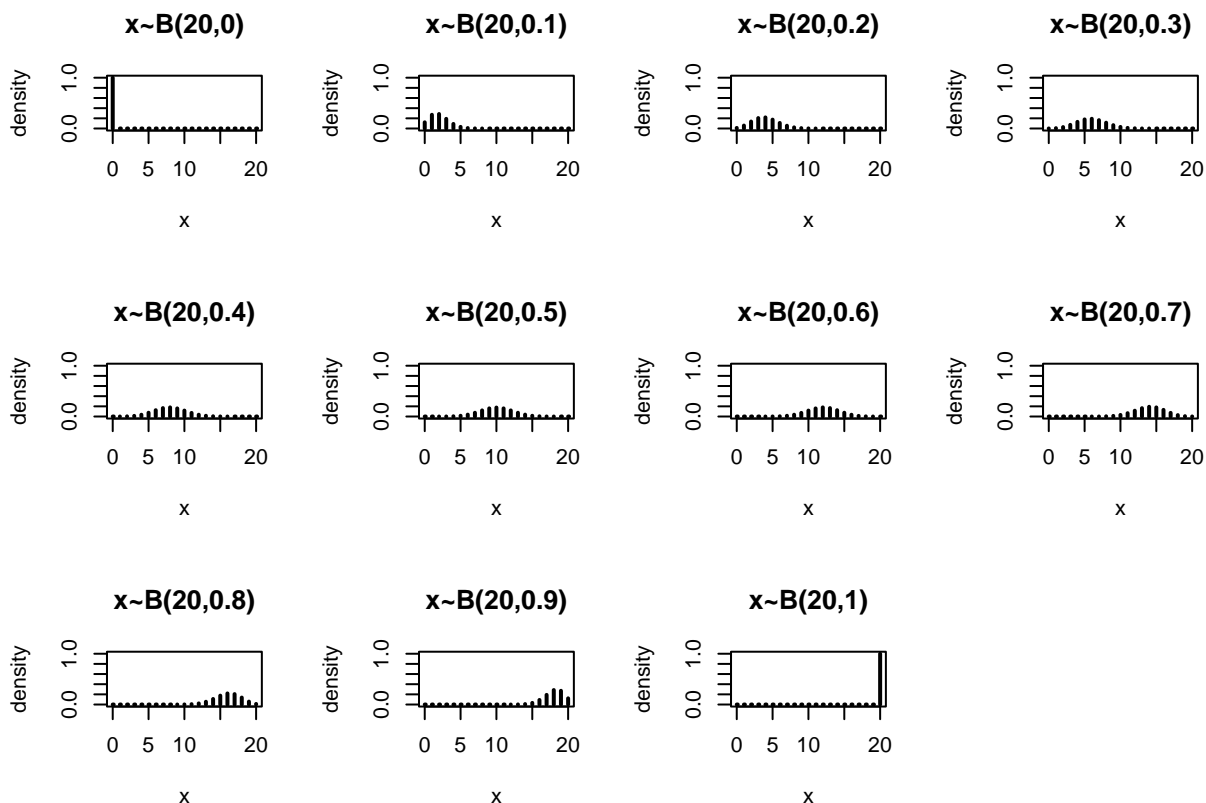
郑盼盼

2024-10-30

1. 对于  $p = 0, 0.1, \dots, 0.9, 1$  绘制  $B(20, p)$  的密度函数图像，考察随着成功概率由小到大的变化，密度图像的变化特征。

```
x <- 0:20          # B(20,p) 所有可能取值: 0,1,...,20
par(mfrow = c(3, 4)) # 定义多张图片的排列方式: 3 行 4 列 (因为 p=0,0.1,...,1 一
  ↪ 共有 11 种取值

# 使用 for 循环绘制 11 张图像
for (p in seq(0,1, by=0.1)){
  y <- dbinom(x,20,p) # 计算 B(20,p) 所有的密度
  plot(x, y,
        type="h", # 选择绘图的类型
        lwd=2,    # 设置线的粗细; lwd: line's width
        xlab = "x", # 设置 x 轴的标签; xlab: x's label
        ylab = "density", # 设置 y 轴的标签; ylab: y's label
        xlim = c(0,20), # 设置 x 轴的范围; xlim: x's limitations
        ylim = c(0,1),  # 设置 y 轴的范围; ylim: y's limitations
        main=paste("x~B(20,"p,")",sep="")) # 设置图像的题目
}
```



2. 已知  $X \sim B(4, \frac{1}{6})$ , 模拟  $X$  的  $m = 10$  次重复观测值, 用  $f_i$  表示  $m$  次观测中  $\{X = i\}$  的频率, 绘制密度矩阵

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ f_0 & f_1 & f_2 & f_3 & f_4 \end{pmatrix}$$

的密度图像, 当  $m = 100, 1000, 10000$  时, 上述密度图像的变化规律是什么。

```
rm(list=ls()) # 删除所有变量
```

```
all_df = data.frame("x" = 0:4) # 定义一个数据框, 用于存储每次模拟的结果
```

```
par(mfrow=c(2,2)) # 对多张图片进行排列 (排成 2 行 2 列), 此处
```

```
# 遍历不同的模拟次数 m, 由于模拟的次数分别为 10^1, 10^2, 10^3, 10^4, 我们可以对
# 幂进行遍历
```

```
for (i in 1:4){
```

```
  m <- 10^i # 模拟次数 m 为 10 的 i 次方
```

```
  x <- rbinom(m, 4, 1/6) # 生成 m 个 B(4, 1/6) 的随机数
```

```
  tmp_freq = c() # 定义一个向量 tmp_freq 用于存储模拟所得的频率 f_i
```

```
  k <- 0:4 # k 为所有可能的结果 B(4, 1/6) 的样本空间只有 {0, 1, 2, 3, 4}
```

```

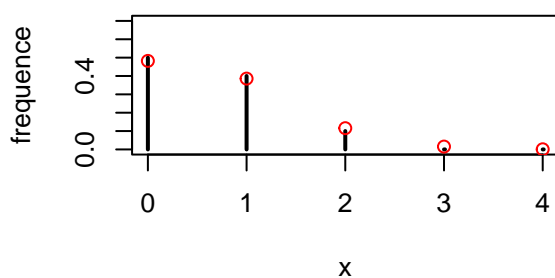
# 遍历计算  $f_i$ 
for (j in k){
  tmp_freq <- c(tmp_freq, sum(x == j)/m) # 将  $f_i$  作为向量 tmp_freq 的最后一个
  ↪ 元素
}

# 绘制频率图
plot(k, tmp_freq,
     type="h",
     lwd=2,
     xlab="x",
     ylab="frequency",
     xlim=c(0,4),
     ylim=c(0,0.7),
     main=paste("m=",m,sep=""))
)

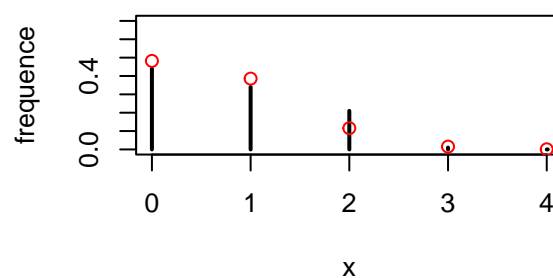
# 绘制理论上的密度值
points(k, dbinom(k,4,1/6), col="red")
}

```

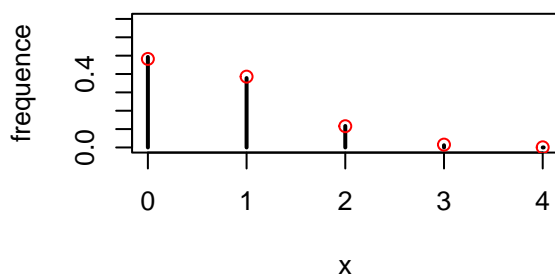
m=10



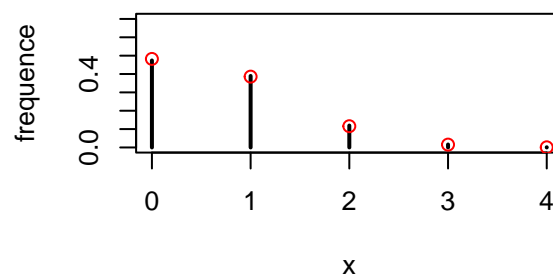
m=100



m=1000



m=10000



3. 袋中有 10 个红球和 10 个黑球。从袋中取后放回的方法依次任取 9 个球，用  $X$  表示取出红球的个数；从袋中用取后不放回的方法依次任取 9 个球，用  $Y$  表示取出红球的个数，将  $X$  和  $Y$  的分布函数曲线用不同颜色绘制在同一图中（绘制坐标位于区间  $[0, 9.5]$  内），解释  $X$  和  $Y$  的分布函数为什么不同。由题意可知， $X \sim B(9, 0.5)$ ， $Y \sim H(20, 10, 9)$

```
rm(list=ls())
par(mfrow=c(1,1))
k <- 0:9
X_probs <- pbinom(k, 9, 0.5)
Y_probs <- phyper(k, 10, 10, 9)

plot(c(0,1), c(X_probs[1], X_probs[1]),
     type="l",
     lwd=.5,
     col="blue",
     xlim=c(0,9.5),
     ylim=c(0,1),
     xlab="x",
     ylab="p")
lines(c(0,1), c(Y_probs[1], Y_probs[1]),
     type="l",
     lwd=.5,
     col="red",
     )

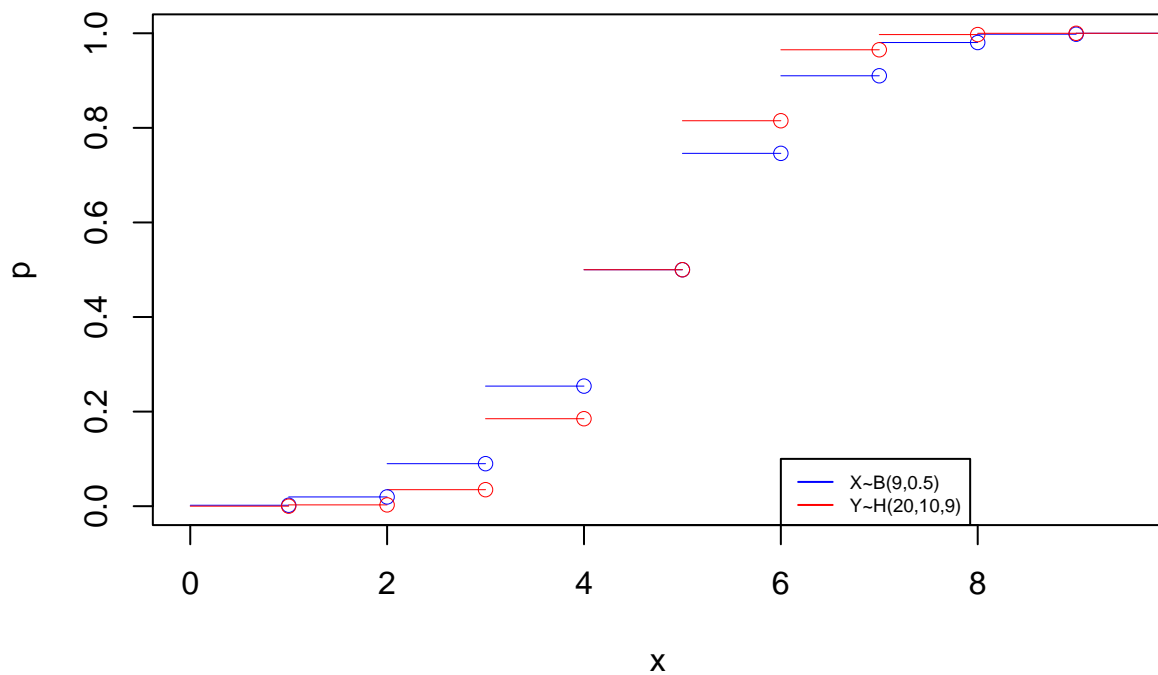
for (i in 2:10){
  lines(c(i-1,i), c(X_probs[i], X_probs[i]),
       type = "l",
       lwd = .5,
       col = "blue")
  lines(c(i-1,i), c(Y_probs[i], Y_probs[i]),
       type = "l",
       lwd = .5,
       col = "red")
  points(i-1, X_probs[i-1],
        col="blue",
        lwd=.5)
```

```

points(i-1, Y_probs[i-1],
      col="red",
      lwd=.5)
}

legend(6, 0.1, legend=c("X~B(9,0.5)", "Y~H(20,10,9)"),
      col=c("blue", "red"), lty=c(1,1), cex=0.6)

```



- 对于 X（有放回抽样），红球每次被取出的概率固定，因此分布函数较为平滑，概率集中在  $X = 4$  附近。
- 对于 Y（无放回抽样），取出的红球数量影响后续抽样的概率，因此分布更加集中，分布曲线变化较为陡峭。