2020 秋季本科时间序列

第 1 次作业答案

10 月 14 日

1. (a) 随机变量 X, Y 的联合分布列为：

Y \ X	1	2	$p_{\cdot j}$
1	0.2	0.3	0.5
2	0.3	0.2	0.5
$p_{i\cdot}$	0.5	0.5	1

代入验证可知，对于任意 $i, j \in \{1, 2\}$，均有：

$$\mathbb{P}(X = i, Y = j) = p_{ij} \neq p_i \cdot p_{\cdot j} = \mathbb{P}(X = i)\mathbb{P}(Y = j)$$

成立，这不满足题目中所给的独立条件，故 X, Y 不独立。

(b) XY 的取值范围为 $\{1, 2, 4\}$，由 (a) 问中的 X, Y 分布易得 XY 的分布列：

XY	1	2	4
p	0.2	0.6	0.2

故 $\mathbb{E}XY = 1 \times 0.2 + 2 \times 0.6 + 4 \times 0.2 = 2.2$，而 $\mathbb{E}XY = \mathbb{E}Y = 1 \times 0.5 + 2 \times 0.5 = 1.5$，进而可得 $\text{cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = 2.2 - 1.5 \times 1.5 = -0.05$。

2. (a) 方法一：

当 X, Y 是离散型随机变量时：

$$\begin{aligned}\mathbb{E}(aX+bY) &= \sum_x\sum_y (ax+by)\mathbb{P}(X=x,Y=y)\\ &= \sum_x\sum_y ax\mathbb{P}(X=x,Y=y)+\sum_x\sum_y by\mathbb{P}(X=x,Y=y)\\ &= \sum_x ax\sum_y \mathbb{P}(X=x,Y=y)+\sum_y by\sum_x \mathbb{P}(X=x,Y=y)\\ &= \sum_x ax\mathbb{P}(X=x)+\sum_y by\mathbb{P}(Y=y)=a\mathbb{E}(X)+b\mathbb{E}(Y)\end{aligned}$$

当 X,Y 是连续型随机变量时：

$$\begin{aligned}\mathbb{E}[aX+bY] &= \iint_{-\infty}^{+\infty}(ax+by)dF(x,y)\\ &= \iint_{-\infty}^{+\infty}(ax+by)f(x,y)dxdy\\ &= \int_{-\infty}^{+\infty}ax\left(\int_{-\infty}^{+\infty}f(x,y)dy\right)dx+\int_{-\infty}^{+\infty}by\left(\int_{-\infty}^{+\infty}f(x,y)dx\right)dy\\ &= \int_{-\infty}^{+\infty}axf_X(x)dx+\int_{-\infty}^{+\infty}byf_Y(y)dy\\ &= a\mathbb{E}X+b\mathbb{E}Y\end{aligned}$$

方法二：

利用 Riemann-Stieltje 积分性质可得：

$$\begin{aligned}\mathbb{E}[aX+bY] &= \iint_{-\infty}^{+\infty}(ax+by)dF(x,y)\\ &= \iint_{-\infty}^{+\infty}axdF(x,y)+\iint_{-\infty}^{+\infty}bydF(x,y)\\ &= a\mathbb{E}X+b\mathbb{E}Y\end{aligned}$$

(b) 根据方差的性质可得：

$$\begin{aligned}\mathrm{var}(\alpha X+Y) &= \alpha^2\mathrm{var}(X)+\mathrm{var}(Y)+2\alpha\mathrm{cov}(X,Y)\\ &= \alpha^2\sigma^2(X)+\sigma^2(Y)+2\alpha\mathrm{cov}(X,Y)\end{aligned}$$

将该方程看作关于 α 的二次方程，由于 $\mathrm{var}(\alpha X+Y)$ 始终大于等于零，所以可得判别式：

$$\Delta=b^2-4ac=4\mathrm{cov}(X,Y)^2-4\sigma^2(X)\sigma^2(Y)\le 0$$

因此 $\mathrm{cov}(X,Y)^2\le\sigma^2(X)\sigma^2(Y)\Rightarrow|\mathrm{cov}(X,Y)|^2\le\sigma^2(X)\sigma^2(Y)$。

又因为:

$$\rho(X,Y)=\frac{\mathrm{cov}(X,Y)}{\sigma(X)\sigma(Y)}$$

所以 $|\mathrm{cov}(X,Y)|^2\le\sigma^2(X)\sigma^2(Y)$，不等式两端同时除以 $\sigma(X)\sigma(Y)$ 可证 $\left|\rho(X,Y)\right|\le 1$。

3. 证明：

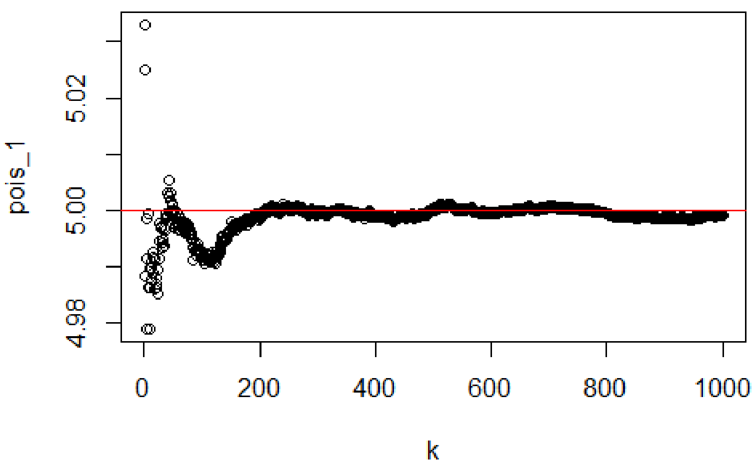
$$
\begin{aligned}
\mathbb{E}\mathbf{A}\mathbf{X} &= \mathbb{E}\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{km} \end{bmatrix}\begin{bmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & \ddots & \vdots \\ X_{m1} & \cdots & X_{mn} \end{bmatrix} \\
&= \mathbb{E}\begin{bmatrix} \sum_{i=1}^m a_{1i}X_{i1} & \cdots & \sum_{i=1}^m a_{1i}X_{in} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^m a_{ki}X_{i1} & \cdots & \sum_{i=1}^m a_{ki}X_{in} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^m a_{1i}\mathbb{E}X_{i1} & \cdots & \sum_{i=1}^m a_{1i}\mathbb{E}X_{in} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^m a_{ki}\mathbb{E}X_{i1} & \cdots & \sum_{i=1}^m a_{ki}\mathbb{E}X_{in} \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{km} \end{bmatrix}\begin{bmatrix} \mathbb{E}X_{11} & \cdots & \mathbb{E}X_{1n} \\ \vdots & \ddots & \vdots \\ \mathbb{E}X_{m1} & \cdots & \mathbb{E}X_{mn} \end{bmatrix} = \mathbf{A}\mathbb{E}\mathbf{X}
\end{aligned}
$$

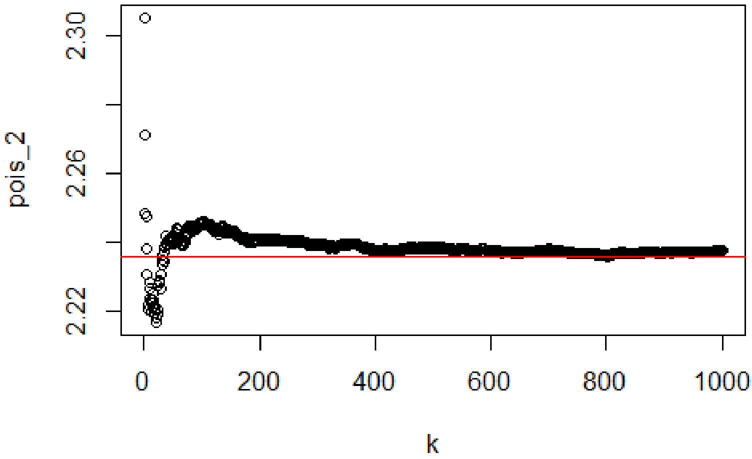
4. (a)

```
library(ggplot2)
#生成抽样
F1<-rpois(1000000,5)          #泊松分布，均值为5，标准差为√5
F2<-rnorm(1000000,1,2)        #正态分布，均值为1，标准差为2
F3<-rgeom(1000000,0.5)        #几何分布，均值为1，标准差为√2
#构造累计均值
pois_1<-c(0)
norm_1<-c(0)
geom_1<-c(0)
for (i in seq(1000,1000000,1000)){
 pois_1[i/1000]<-mean(F1[1:i])
 norm_1[i/1000]<-mean(F2[1:i])
 geom_1[i/1000]<-mean(F3[1:i])
}
#构造累计标准差
pois_2<-c(0)
norm_2<-c(0)
geom_2<-c(0)
for (i in seq(1000,1000000,1000)){
 pois_2[i/1000]<-sd(F1[1:i])
 norm_2[i/1000]<-sd(F2[1:i])
```

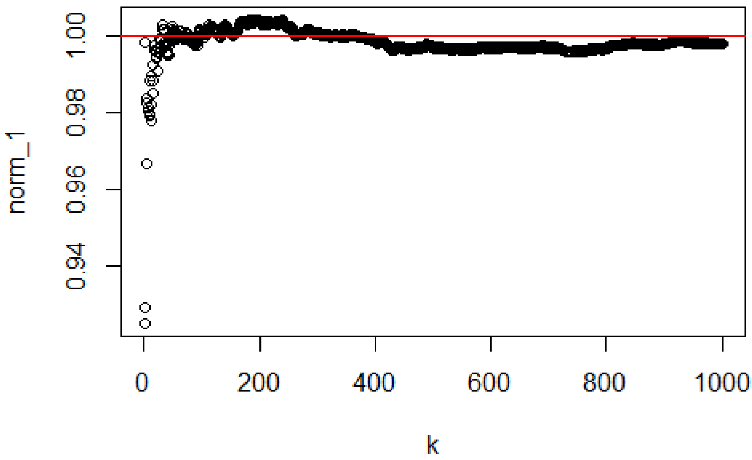
```
 geom_2[i/1000]<-sd(F3[1:i])
}
#绘图
#泊松分布：
k<-c(1:1000)
plot(k,pois_1)
abline(h=5, col="red")    #均值收敛于5
plot(k,pois_2)
abline(h=sqrt(5), col="red")    #标准差收敛√5
#正态分布：
plot(k,norm_1)
abline(h=1, col="red")    #均值收敛于1
plot(k,norm_2)
abline(h=2, col="red")    #标准差收敛于2
#几何分布：
plot(k,geom_1)
abline(h=1, col="red")    #均值收敛于1
plot(k,geom_2)
abline(h=sqrt(2), col="red")    #标准差收敛于√2
```

下边分别是泊松、正态和几何分布的均值和标准差收敛图：

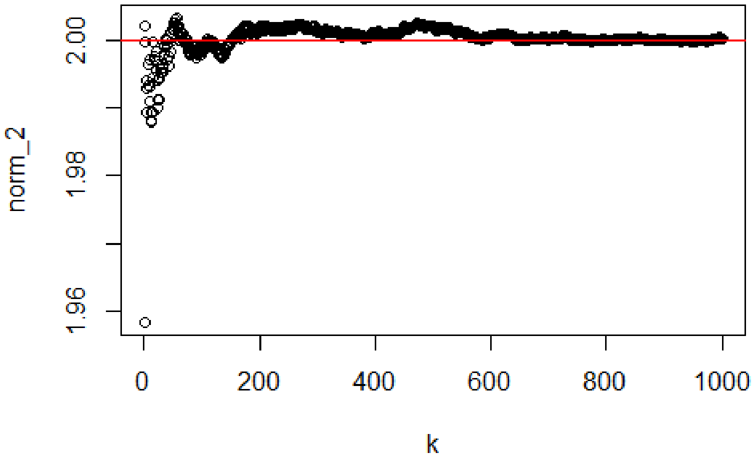




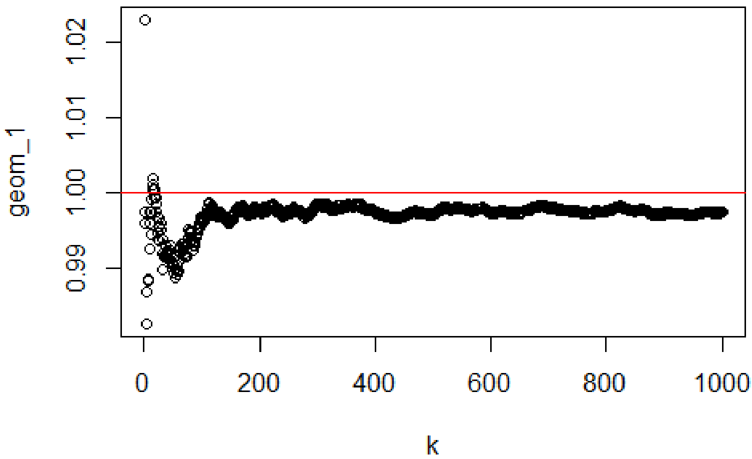
pois_2
2.22
2.26
2.30
0
200
400
600
800
1000
k



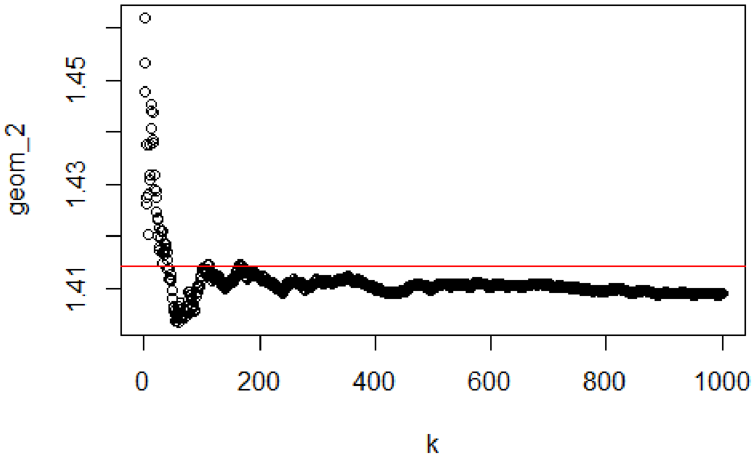
norm_1
0.94
0.96
0.98
1.00
0
200
400
600
800
1000
k



norm_2
1.96
1.98
2.00
0
200
400
600
800
1000
k



geom_1
0.99
1.00
1.01
1.02
0
200
400
600
800
1000
k



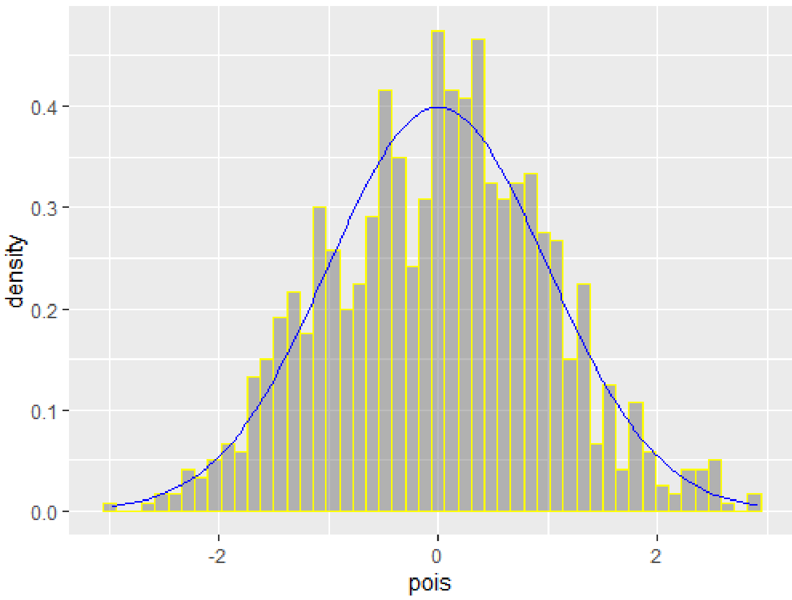
由此可以看出随着 k 的增加，$\mu_k^{(i)}$ 与 $\sigma_k^{(i)}$ 逐渐逼近 $\mu^{(i)}$ 与 $\sigma^{(i)}$。

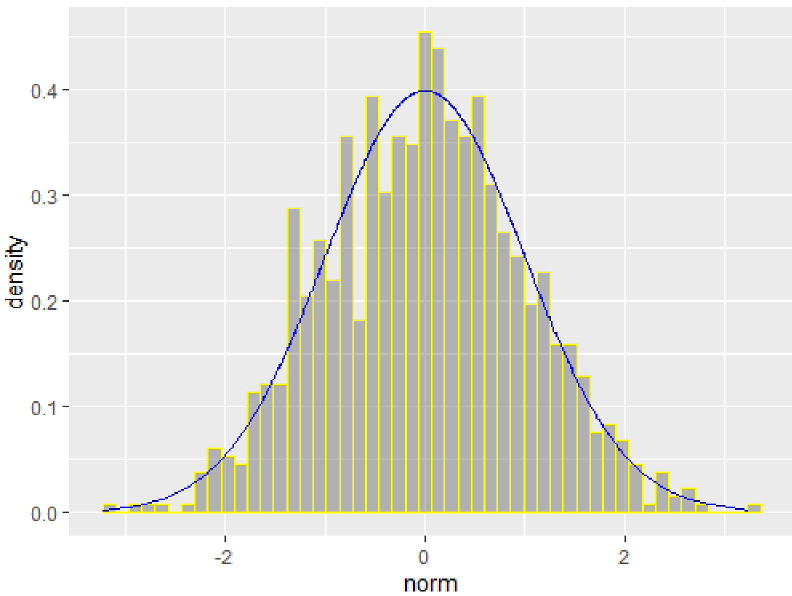
(b)

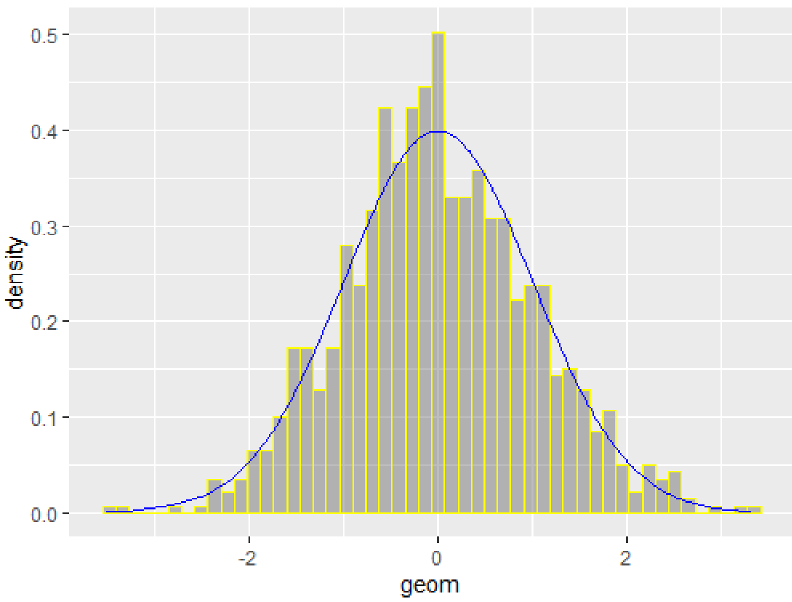
```
#构造SK
pois<-c(0)
norm<-c(0)
geom<-c(0)
for (i in seq(1000,1000000,1000)){
 pois[i/1000]<-mean(F1[(i-999):i])
 norm[i/1000]<-mean(F2[(i-999):i])
 geom[i/1000]<-mean(F3[(i-999):i])
}
#标准化SK
p<-scale(pois)
n<-scale(norm)
g<-scale(geom)
#绘制比较图
#图一：泊松vs标准正态
ggplot()+
geom_histogram(mapping = aes(x = sort(p),y = ..density..),
     color = "yellow",alpha = 0.4,bins = 50)+
geom_line(mapping = aes(x = sort(p),y = dnorm(sort(p))),
```

```
color = "blue")+labs(x = "pois")
#图二： 正态vs标准正态
ggplot()+
geom_histogram(mapping = aes(x = sort(n),y = ..density..),
    color = "yellow",alpha = 0.4,bins = 50)+
geom_line(mapping = aes(x = sort(n),y = dnorm(sort(n))),
color = "blue")+labs(x = "norm")
#图三： 几何vs标准正态
ggplot()+
geom_histogram(mapping = aes(x = sort(g),y = ..density..),
    color = "yellow",alpha = 0.4,bins = 50)+
geom_line(mapping = aes(x = sort(g),y = dnorm(sort(g))),
color = "blue")+labs(x = "geom")
```

下边分别是泊松和标准正态分布、正态和标准正态分布以及几何和标准正态分布的对比图：







由图可知 $\xi_k^{(i)}$ 的分布接近于标准正态的分布。