概率论与数理统计 第5次作业

Name: 宋昊原　　Student ID: 2022010755

October 19, 2023

1 取球

1.1

$$P(X=x,Y=y)=\frac{3!}{x!y!(3-x-y)!}(\frac{3}{12})^x(\frac{4}{12})^y(\frac{5}{12})^{3-x-y}$$

于是

(X,Y)	(0,Y)	(1,Y)	(2,Y)	(3,Y)
(X,0)	$\frac{125}{1728}$	$\frac{225}{1728}$	$\frac{135}{1728}$	$\frac{27}{1728}$
(X,1)	$\frac{300}{1728}$	$\frac{360}{1728}$	$\frac{108}{1728}$	0
(X,2)	$\frac{240}{1728}$	$\frac{144}{1728}$	0	0
(X,3)	$\frac{64}{1728}$	0	0	0

1.2

$$P(X=1)=\sum_{j=0}^{3}P(X=1,Y=j)=\frac{729}{1728}=\frac{27}{64}$$

2 联合分布的公式

$$\begin{aligned}F(b,d)-F(a,d)-F(b,c)+F(a,c)&=(F(b,d)-F(a,d))-(F(b,c)-F(a,c))\\&=(P(X\le b,Y\le d)-P(X\le a,Y\le d))-(P(X\le b,Y\le c)-P(X\le a,Y\le c))\\&=P(a<X\le b,Y\le d)-P(a<X\le b,Y\le c)\\&=P(a<X\le b,c<Y\le d)\end{aligned}$$

3 圆盘均匀分布

3.1

$$f(x,y)=\begin{cases}\frac{1}{\pi} & x^2+y^2\le 1\\ 0 & else\end{cases} \tag{1}$$

3.2

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}dy=2\frac{\sqrt{1-x^2}}{\pi}$$

同理

$$f_Y(y)=2\frac{\sqrt{1-y^2}}{\pi}$$

3.3

$$P(R\le r)=\iint_{x^2+y^2\le r^2}f(x,y)dxdy=\frac{\pi r^2}{\pi}=r^2$$

3.4

将R视为一个新的随机变量，则其概率密度函数

$$f_R(r) = \frac{d}{dr}P(R \leq r) = 2r, \forall r \in [0,1]$$

则

$$E(R) = \int_0^1 r f_R(r) dr = \int_0^1 2r^2 dr = \frac{2}{3}$$

4 二元正态分布的边际密度

首先给出二元正态分布的PDF

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right)$$

其对X的边际密度为

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy$$

令$\xi = \frac{x-\mu_1}{\sigma_1}$，$\eta = \frac{y-\mu_2}{\sigma_2}$，则

$$f_X(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2(1-\rho^2)}(\xi^2 + \eta^2 - 2\rho\xi\eta)\right) d(\sigma_2\eta + \mu_2)$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2(1-\rho^2)}((\eta-\rho\xi)^2 + (1-\rho^2)\xi^2)\right) d\eta$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \exp(-\frac{1}{2}\xi^2) \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2(1-\rho^2)}(\eta-\rho\xi)^2\right) d\eta$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \exp(-\frac{1}{2}\xi^2) \int_{-\infty}^{+\infty} \exp\left(-\left(\frac{\eta-\rho\xi}{\sqrt{2(1-\rho^2)}}\right)^2\right) d\eta$$

$$= \frac{1}{\sqrt{2}\pi\sigma_1} \exp(-\frac{1}{2}\xi^2) \int_{-\infty}^{+\infty} \exp\left(-\left(\frac{\eta-\rho\xi}{\sqrt{2(1-\rho^2)}}\right)^2\right) d\frac{\eta-\rho\xi}{\sqrt{2(1-\rho^2)}}$$

$$= \frac{1}{\sqrt{2}\pi\sigma_1} \exp(-\frac{1}{2}\xi^2) \int_{-\infty}^{+\infty} \exp\left(-\left(\frac{\eta-\rho\xi}{\sqrt{2(1-\rho^2)}}\right)^2\right) d\frac{\eta-\rho\xi}{\sqrt{2(1-\rho^2)}}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{\xi^2}{2}}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

相当于一元正态分布.
同理

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$$

5 二元正态分布的条件密度

$$
\begin{aligned}
f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\
&= \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{y-\mu_2}{\sigma_2}\right)^2-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right)}{\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}} \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_1}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\rho\frac{y-\mu_2}{\sigma_2}\right)^2-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right) \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_1}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}-\rho\frac{y-\mu_2}{\sigma_2}\right)^2\right)
\end{aligned}
$$

6 三角形域内的均匀分布

6.1 联合分布

设三角形域为D，则

$$
f(x,y)=\begin{cases} 2 & (x,y)\in D \\ 0 & else \end{cases} \tag{2}
$$

6.2 边际分布

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{+\infty} f(x,y)dx \\
&= \int_0^{1-y} 2dx \\
&= 1-y, \forall y\in[0,1]
\end{aligned}
$$

6.3 条件分布

$$
\begin{aligned}
f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\
&= \frac{2}{1-y}, \forall (x,y)\in D
\end{aligned}
$$

7 独立随机变量

7.1 条件分布

条件分布的PMF：

$$
\begin{aligned}
f_{X_1|X_1+X_2}(m|n) &= \frac{f_{X_1,X_2}(m,n-m)}{f_{X_1+X_2}(n)} \\
&= \frac{f_{X_1}(m)f_{X_2}(n-m)}{f_{X_1+X_2}(n)} \\
&= \frac{\frac{\lambda_1^m}{m!}e^{-\lambda_1}\frac{\lambda_2^{n-m}}{(n-m)!}e^{-\lambda_2}}{\frac{(\lambda_1+\lambda_2)^n}{n!}e^{-\lambda_1-\lambda_2}} \\
&= \binom{n}{m}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^m\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-m}
\end{aligned}
$$

7.2 解释

上述条件分布质量函数在形式上接近二项分布. 事实上，这种情形在语义上也接近二项分布.
这里X_1和X_2服从泊松分布的含义可以认为是：对一个系统观测一段时间，其中事件1发生的次数为X_1，事件2发生的次数为X_2.
当前语境下，已知事件1和2共发生了n次，求此条件下事件1发生m次的概率，这非常类似于二项分布.
同时，由于λ_1和λ_2分别为这段时间内发生事件1和2的次数的期望值，因此可以认为从发生的每个事件中随机抽一个事件，是事件1、事件2的概率的比值为$\frac{\lambda_1}{\lambda_2}$.
也就是，这个条件分布其实是$B(n, \frac{\lambda_1}{\lambda_1+\lambda_2})$.

8 约定见面

规定两人到达时间以2点为1，1点为0，均匀分划刻度.

8.1

$$f(x,y) = \begin{cases} 1 & (x,y) \in D = [0,1] \times [0,1] \\ 0 & else \end{cases} \tag{3}$$

8.2

这里要求的事件相当于以下条件：

$$X - Y > \frac{1}{6} \vee Y - X > \frac{1}{6}$$

D是单位正方形，而D中的上述区域（设为A）相当于2个直角边长为$\frac{5}{6}$的等腰直角三角形. 则

$$P((x,y) \in A) = \int_A f(x,y)dxdy = \int_A dxdy = S(A) = \frac{25}{36}$$

即所求概率为$\frac{25}{36}$.

9 Farlie-Morgenstein族

9.1

$$H_X(x) = \lim_{y \to +\infty} H(x,y)$$

$$= \lim_{y \to +\infty} F(x)G(y)\{1 + \alpha[1 - F(x)][1 - G(y)]\}$$

由于

$$\lim_{y \to +\infty} G(y) = 1$$

有

$$H_X(x) = F(x)$$

同理

$$H_Y(y) = G(y)$$

9.2

这里只需要讨论$(x,y) \in D = [0,1] \times [0,1]$的情况.
此时，$F(x) = x, G(y) = y$.
对于$\alpha = 1$，有

$$H_1(x,y) = xy(1 - (1-x)(1-y)) = x^2y + xy^2 - x^2y^2$$

求出D上的概率密度函数（D外恒取0）

$$h_1(x,y) = \frac{\partial^2}{\partial x \partial y} H_1(x,y) = 2x + 2y - 4xy$$

同理，对于$\alpha = -1$

$$H_2(x,y) = 2xy - x^2y - xy^2 + x^2y^2$$
$$h_2(x,y) = 2 - 2x - 2y + 4xy$$

10 Copula函数

令

$$H(x,y) = C(F(x), G(y))$$

由于$C(u,v)$是连续分布的联合累积分布，它在$\mathbb{R}^2$内任何一点都应当连续. 于是

$$H_X(x) = \lim_{y\to+\infty} H(x,y) = C(F(x), G(+\infty)) = C(F(x), 1)$$

由于$C(u,v)$对X和Y的边际分布都是$[0,1]$上均匀分布

$$C(F(x), 1) = C_X(F(x)) = F(x)$$

于是

$$H_X(x) = F(x)$$

同理

$$H_Y(y) = G(y)$$

11 全概率公式和Bayes公式

在接下来的讨论中，我们都研究$f_X(x)$的全概率公式和$f_{Y|X}(y|x)$的Bayes公式，也就是在$X = x$发生的情况下推断$Y = y$发生的后验概率.
将指出，这一公式的形式与X是连续还是离散无关，只与Y有关，但这只是形式上的. 实际上相应的函数f要根据X是离散还是连续选择理解成PMF还是PDF.

11.1 Y离散的情形

全概率公式

$$f_X(x) = \sum_y f_{X|Y}(x|y) f_Y(y)$$

Bayes公式

$$f_{Y|X}(y_0|x) = \frac{f_{X|Y}(x|y_0) f_Y(y_0)}{\sum\limits_y f_{X|Y}(x|y) f_Y(y)}$$

11.2 Y连续的情形

全概率公式

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X|Y}(x|y) f_Y(y) dy$$

Bayes公式

$$f_{Y|X} y_0|x = \frac{f_{X|Y}(x|y_0) f_Y(y_0)}{\int_{-\infty}^{+\infty} f_{X|Y}(x|y) f_Y(y) dy}$$

12

12.1

由$f(x,y)$是PDF，应有

$$\iint\limits_{x^2+y^2\le 1} \frac{c}{1+x^2+y^2}dxdy = 1$$

作极坐标变换，有

$$\int_0^{2\pi}(\int_0^1 \frac{c\rho}{1+\rho^2}d\rho)d\phi$$

$$= \pi\int_0^1 \frac{c}{1+\rho^2}d(1+\rho^2)$$

$$= \pi c\ln(1+\rho^2)|_0^1$$

$$= \pi c\ln 2 = 1$$

于是

$$c = \frac{1}{\pi\ln 2}$$

12.2

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy$$

$$= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi\ln 2(1+x^2+y^2)}dy$$

$$= \frac{1}{\pi\ln 2\sqrt{1+x^2}}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{(\frac{y}{\sqrt{1+x^2}})^2+1}d\frac{y}{\sqrt{1+x^2}}$$

$$= \frac{1}{\pi\ln 2\sqrt{1+x^2}}\arctan\frac{y}{\sqrt{1+x^2}}|_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}$$

$$= \frac{2}{\pi\ln 2\sqrt{1+x^2}}\arctan\sqrt{\frac{1-x^2}{1+x^2}}$$

同理

$$f_Y(y) = \frac{2}{\pi\ln 2\sqrt{1+y^2}}\arctan\sqrt{\frac{1-y^2}{1+y^2}}$$

考虑$(0,0)$点

$$f_X(0) = f_Y(0) = \frac{1}{2\ln 2}$$

$$f(0,0) = c = \frac{1}{\pi\ln 2}$$

故

$$f(0,0) \ne f_X(0)f_Y(0)$$

于是X和Y不独立.

13 边际正态分布但不是二元正态分布的反例

13.1

只需证明$g(x,y)\geq 0$且$\iint\limits_{\mathbb{R}^2} g(x,y)dxdy=1$.
由于$X,Y\sim N(0,1)$且X和Y独立

$$f(x,y)=\frac{1}{2\pi}\exp{(-\frac{x^2+y^2}{2})}$$

$\forall (x,y)\in D=\{(x,y)|x^2+y^2\leq 1\}$，有

$$\frac{e^{-\frac{1}{2}}}{2\pi}\leq f(x,y)\leq\frac{1}{2\pi}$$

而

$$|\frac{xy}{100}|\leq\frac{1}{200}<\frac{e^{-\frac{1}{2}}}{2\pi}$$

故在D上

$$g(x,y)=f(x,y)+\frac{xy}{100}\geq 0$$

而在D外$g(x,y)=f(x,y)$非负，这证明了非负性.
要证明归一性，只需要证明在D上g和f积分相同，即

$$\iint\limits_{D}\frac{xy}{100}dxdy=0$$

由于$\frac{xy}{100}$关于x轴对称，这个积分确实为0，这证明了归一性.
于是$g(x,y)$是一个二维概率密度函数.

13.2

考虑U的边际分布

$$g_U(x)=\int_{-\infty}^{+\infty}g(x,y)dy$$

$$=\int_{-\infty}^{+\infty}f(x,y)dy+\phi(x)$$

上述第一项就是标准正态分布，考虑余项$\phi(x)$，当$x\notin(-1,1)$时，$\phi(x)=0$，只需考虑$x\in(-1,1)$的情形，此时

$$\phi(x)=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{xy}{100}dy$$

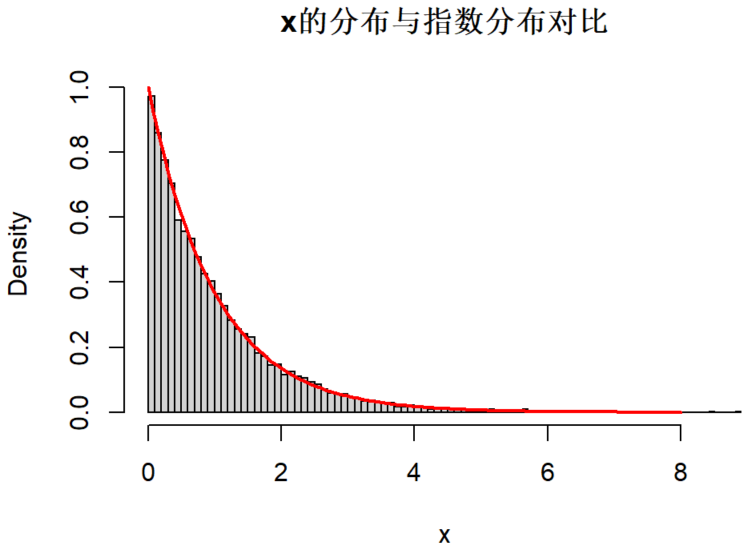
这是奇函数在对称区间上的积分，为0. 于是$U\sim N(0,1)$，同理，$V\sim N(0,1)$.
然而，二元正态分布的概率密度函数是连续函数，而$g(x,y)$在任何满足

$$x^2+y^2=1\wedge xy\neq 0$$

的点不连续，因此(U,V)不服从二元正态分布.

14 计算机实验：随机变量的函数

以下是x_i的直方图与指数分布概率密度函数的比较，其中红色曲线为指数分布概率密度函数.



以下是我使用的代码.

```
> y<-runif(10000)
> x<- -log(1-y)
> hist(x,main="x的分布与指数分布对比",breaks = 100,probability = TRUE)
> x1<-seq(0,8,length.out=100)
> y1<-exp(-x1)
> lines(x1, y1, col = "red", lwd = 2)
```