# Homework 5

姓名: 方嘉聪 学号: 2200017849

**Problem 1.** 定义二维随机变量 $X, Y$. 证明: $\operatorname{Corr}(X,Y) = \pm 1$ 当且仅当存在实数 $a \neq 0, b$, 使得 $\mathbb{P}(Y = aX + b) = 1$.

提示: 利用结论 (无需证明), 若随机变量 $Z$ 满足 $\operatorname{Var}(Z) = 0$, 则 $\mathbb{P}(Z = \mathbb{E}(Z)) = 1$. ◀

**Solution.** 分别证明两个方向. 记 $X, Y$ 的均值分别为 $\mu_X, \mu_Y$, 标准差分别为 $\sigma_X, \sigma_Y$.

(1) 若 $\operatorname{Corr}(X,Y) = \pm 1$. 这里我们证明 $\operatorname{Corr}(X,Y) = 1$ 的情况, $\operatorname{Corr}(X,Y) = -1$ 的情况同理. 由定义知

$$\operatorname{Corr}(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sigma_X \sigma_Y} = 1. \implies \operatorname{Cov}(X,Y) = \sigma_X \sigma_Y.$$

考虑 $a = \sigma_Y / \sigma_X \neq 0$, 我们来计算 $\operatorname{Var}(Y - aX)$.

$$\operatorname{Var}(Y - aX) = \sigma_Y^2 + a^2 \sigma_X^2 - 2a \operatorname{Cov}(X,Y) = a^2 \sigma_X^2 - 2a \sigma_X \sigma_Y + \sigma_Y^2 = 0.$$

由结论知 $\mathbb{P}(Y - aX = \mathbb{E}(Y - aX)) = 1$. 令 $b = \mathbb{E}(Y - aX)$, 则 $\mathbb{P}(Y = aX + b) = 1$. 证毕.
注: 对于 $\operatorname{Corr}(X,Y) = -1$ 的情况, 可以取 $a = -\sigma_Y / \sigma_X, b = \mathbb{E}(Y - aX)$.

(2) 若存在实数 $a \neq 0, b$, 使得 $\mathbb{P}(Y = aX + b) = 1$. 我们先来证明, 对于任意随机变量 $Z$ 有

$$\mathbb{P}(Z = \mathbb{E}(Z)) = 1 \implies \operatorname{Var}(Z) = 0.$$

证明: 注意到

$$\mathbb{P}(Z = \mathbb{E}(Z)) = 1 \implies \mathbb{P}((Z - \mathbb{E}(Z))^2 = 0) = 1 \implies \operatorname{Var}(Z) = \mathbb{E}((Z - \mathbb{E}(Z))^2) = 0.$$

回到本题, 由于 $\mathbb{P}(Y = aX + b) = 1$, 那么 $\operatorname{Var}(Y - aX) = 0$. 故

$$0 = \operatorname{Var}(Y - aX) = \sigma_Y^2 + a^2 \sigma_X^2 - 2a \operatorname{Cov}(X,Y)$$

那么 (利用均值不等式)

$$\operatorname{Corr}(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{1}{2a} \frac{\sigma_Y}{\sigma_X} + \frac{a}{2} \frac{\sigma_X}{\sigma_Y} \implies |\operatorname{Corr}(X,Y)| \geq 1.$$

由于 $|\operatorname{Corr}(X,Y)| \leq 1$, 那么 $\operatorname{Corr}(X,Y) = \pm 1$. 证毕.

综上所述, $\operatorname{Corr}(X,Y) = \pm 1$ 当且仅当存在实数 $a \neq 0, b$, 使得 $\mathbb{P}(Y = aX + b) = 1$. ◁

**Problem 2.** 对于 $\sigma_1 > 0, \sigma_2 > 0, -1 < \rho < 1$, 二维随机变量 $(U,V) \sim \mathcal{N}(0, 0, \sigma_1^2, \sigma_2^2, \rho)$. 本题中我们将计算 $\mathbb{E}(\operatorname{ReLU}(U) \cdot \operatorname{ReLU}(V))$. 其中 $\operatorname{ReLU}(x) = \max\{0, x\}$.

设二维随机变量 $(X,Y) \sim \mathcal{N}(0,0,1,1,\rho)$, 令二维随机变量 $(R, \Theta)$ 满足 $R \geq 0, \Theta \in [0, 2\pi]$, 且

$$\begin{cases} X = R \cdot \left( \sqrt{1 - \rho^2} \cdot \cos\Theta + \rho \cdot \sin\Theta \right) = R \cdot \sin(\arccos\rho + \Theta) \\ Y = R \sin\Theta \end{cases} \tag{1}$$

(1) 令 $x = r\cdot\left(\sqrt{1-\rho^2}\cdot\cos\theta+\rho\cdot\sin\theta\right)$, $y=r\sin\theta$. 验证 $x^2+y^2-2\rho xy = r^2(1-\rho^2)$.

(2) 计算 $R,\Theta$ 的联合密度函数, $R$ 和 $\Theta$ 的各自边际密度函数, 并判断 $R$ 和 $\Theta$ 的独立性.

(3) 计算 $\mathbb{E}(\mathrm{ReLU}(X)\cdot\mathrm{ReLU}(Y))$. 提示: 利用结论 (无需证明)

$$\int_0^{+\infty} x^3 e^{-x^2/2}\mathrm{d}x = 2,$$

$$\int_0^{\pi-\arccos\rho}\left(\rho\cdot\sin^2\theta+\sqrt{1-\rho^2}\sin\theta\cos\theta\right)\mathrm{d}\theta = \frac{1}{2}\left(\rho(\pi-\arccos\rho)+\sqrt{1-\rho^2}\right).$$

(4) 验证 $(\sigma_1 X,\sigma_2 Y)$ 与 $(U,V)$ 具有相同的分布.

(5) 计算 $\mathbb{E}(\mathrm{ReLU}(U)\cdot\mathrm{ReLU}(V))$.

◀

**Solution.** (1) 直接带入验证即可.

$$\begin{aligned}
x^2+y^2-2\rho xy &= r^2\left(\sqrt{1-\rho^2}\cos\theta+\rho\sin\theta\right)^2 + r^2\sin^2\theta - 2\rho r^2\left(\sqrt{1-\rho^2}\cos\theta+\rho\sin\theta\right)\sin\theta \\
&= r^2(1-\rho^2)\cos^2\theta + r^2\rho^2\sin^2\theta + r^2\sin^2\theta - 2\rho r^2\sin\theta\cos\theta \\
&= r^2(1-\rho^2).
\end{aligned}$$

(2) 我们先来计算

$$\left|\frac{\partial(X,Y)}{\partial(R,\Theta)}\right| = \left|\begin{pmatrix}\sqrt{1-\rho^2}\cos\Theta+\rho\sin\Theta & R(-\sqrt{1-\rho^2}\sin\Theta+\rho\cos\Theta) \\ \sin\Theta & R\cos\Theta\end{pmatrix}\right| = R\sqrt{1-\rho^2}.$$

那么 $R,\Theta$ 的联合密度函数为

$$\begin{aligned}
f_{R,\Theta}(r,\theta) &= f_{X,Y}(x(r,\theta),y(r,\theta))\cdot\left|\frac{\partial(X,Y)}{\partial(R,\Theta)}\right| \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right)\cdot r\sqrt{1-\rho^2} \\
&= \frac{1}{2\pi}r\exp\left(-\frac{r^2}{2}\right). \quad \text{利用 (1) 中结论}
\end{aligned}$$

那么 $R$ 的边际密度函数为

$$f_R(r) = \int_0^{2\pi} f_{R,\Theta}(r,\theta)\mathrm{d}\theta = re^{-r^2/2}\int_0^{2\pi}\frac{1}{2\pi}\mathrm{d}\theta = re^{-r^2/2}.$$

而 $\Theta$ 的边际密度函数为

$$f_\Theta(\theta) = \int_0^{+\infty} f_{R,\Theta}(r,\theta)\mathrm{d}r = \frac{1}{2\pi}\int_0^{+\infty} re^{-r^2/2}\mathrm{d}r = \frac{1}{2\pi}.$$

由于 $f_{R,\Theta}(r,\theta) = f_R(r)f_\Theta(\theta)$, 那么 $R$ 和 $\Theta$ 是独立的.

(3) 注意到

$$\begin{aligned}
\mathbb{E}(\mathrm{ReLU}(X)\cdot\mathrm{ReLU}(Y)) &= \iint_{x\geq 0\wedge y\geq 0} xy\cdot f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
\text{(使用(1)换元)} \quad &= \iint_{\mathbf{D}} xy\cdot f_{X,Y}(x(r,\theta),y(r,\theta))\cdot\left|\frac{\partial(X,Y)}{\partial(R,\Theta)}\right|\mathrm{d}r\,\mathrm{d}\theta \\
&= \iint_{\mathbf{D}} f_{R,\Theta}(r,\theta)\cdot r^2\left(\sqrt{1-\rho^2}\sin\theta\cos\theta+\rho\sin^2\theta\right)\mathrm{d}r\,\mathrm{d}\theta
\end{aligned}$$

其中区域

$$\begin{aligned}\mathbf{D} &= \left\{(r,\theta): r\geq 0, \sin\theta\geq 0, \sqrt{1-\rho^2}\cos\theta+\rho\sin\theta\geq 0\right\}\\ &= \{(r,\theta): r\geq 0, \theta\in[0,\pi-\arccos\rho]\}.\end{aligned}$$

那么我们有

$$\begin{aligned}\mathbb{E}(\mathrm{ReLU}(X)\cdot\mathrm{ReLU}(Y)) &= \frac{1}{2\pi}\int_0^{\pi-\arccos\rho}\int_0^{+\infty} r^3 e^{-r^2/2}\cdot\left(\sqrt{1-\rho^2}\sin\theta\cos\theta+\rho\sin^2\theta\right)\mathrm{d}r\,\mathrm{d}\theta\\ &= \frac{1}{2\pi}\int_0^{+\infty} r^3 e^{-r^2/2}\int_0^{\pi-\arccos\rho}\left(\rho\cdot\sin^2\theta+\sqrt{1-\rho^2}\sin\theta\cos\theta\right)\mathrm{d}\theta\,\mathrm{d}r\\ &= \frac{1}{2\pi}\left(\rho(\pi-\arccos\rho)+\sqrt{1-\rho^2}\right).\end{aligned}$$

最后一个等式利用了提示中的结论.

(4) 记 $(W,T)=(\sigma_1X,\sigma_2Y)$, 那么

$$\begin{aligned}f_{W,T}(w,t) &= f_{X,Y}(x,y)\cdot\left|\frac{\partial(X,Y)}{\partial(W,T)}\right|\\ &= \frac{1}{\sigma_1\sigma_2}\cdot\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left(x^2+y^2-2\rho xy\right)\right)\\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_1^2}+\frac{y^2}{\sigma_2^2}-\frac{2\rho xy}{\sigma_1\sigma_2}\right)\right\}.\end{aligned}$$

故 $(\sigma_1X,\sigma_2Y)=(W,T)\sim\mathcal{N}(0,0,\sigma_1^2,\sigma_2^2,\rho)$, 与 $(U,V)$ 具有相同的分布.

(5) 由于 $(\sigma_1X,\sigma_2Y)$ 与 $(U,V)$ 具有相同的分布, 那么

$$\begin{aligned}\mathbb{E}(\mathrm{ReLU}(U)\cdot\mathrm{ReLU}(V)) &= \mathbb{E}(\mathrm{ReLU}(\sigma_1X)\cdot\mathrm{ReLU}(\sigma_2Y)) = \sigma_1\sigma_2\mathbb{E}(\mathrm{ReLU}(X)\cdot\mathrm{ReLU}(Y))\\ &= \frac{\sigma_1\sigma_2}{2\pi}\left(\rho(\pi-\arccos\rho)+\sqrt{1-\rho^2}\right).\end{aligned}$$

◁

**Problem 3.** 在课上我们考虑了如下矩阵 $A\in\mathbb{R}^{n\times n}$: 对于任意 $1\leq i,j\leq n, A_{i,j}\sim\mathcal{N}(0,1)$, 且不同元素相互独立. 计算 $\mathbb{E}(\mathrm{tr}(A^3))$ 和 $\mathbb{E}(\mathrm{tr}(A^4))$. 提示: 考虑 $n=1$ 的情况, 并参考作业三第六题. ◀

**Solution.** 我们先来计算 $\mathbb{E}(A_{i,j}^3)$ 与 $\mathbb{E}(A_{i,j}^4)$. 有对称性知 $\mathbb{E}(A_{i,j}^3)=0$. 而由作业三第六题知 $X\sim\mathcal{N}(0,1)\implies X^2\sim\chi^2(1)$. 那么 $\mathbb{E}(A_{i,i}^2)=1, \mathrm{Var}(A_{i,i}^2)=2$. 由此知

$$\mathbb{E}(A_{i,j}^4)=\mathbb{E}(A_{i,j}^2)^2+\mathrm{Var}(A_{i,j}^2)=3.$$

记 $A_{(i,j)}^{(t)}$ 表示 $A^t$ 的第 $(i,j)$ 个元素. 那么我们有

$$A_{i,j}^{(2)}=\sum_{k=1}^n A_{i,k}A_{k,j}\implies A_{i,i}^{(3)}=\sum_{j=1}^n A_{i,j}^{(2)}A_{j,i}=\sum_{j=1}^n\sum_{k=1}^n A_{i,k}A_{k,j}A_{j,i}.$$

那么

$$\mathbb{E}(A_{i,i}^{(3)})=\mathbb{E}\left(\sum_{j=1}^n\sum_{k=1}^n A_{i,k}A_{k,j}A_{j,i}\right)=\sum_{j=1}^n\sum_{k=1}^n\mathbb{E}(A_{i,k}A_{k,j}A_{j,i})=\mathbb{E}(A_{i,i}^3)=0.$$

$$\mathbb{E}(\mathrm{tr}(A^3))=\mathbb{E}\left(\sum_{i=1}^n A_{i,i}^{(3)}\right)=\sum_{i=1}^n\mathbb{E}\left(A_{i,i}^{(3)}\right)=0.$$

类似地, 我们有

$$A_{i,j}^{(4)} = \sum_{k=1}^{n} A_{i,k} A_{k,j}^{(3)} \implies A_{i,i}^{(4)} = \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{t=1}^{n} A_{i,k}A_{k,j}A_{j,t}A_{t,i}.$$

那么

$$\begin{aligned}\mathbb{E}(A_{i,i}^{(4)}) &= \mathbb{E}\left(\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{t=1}^{n} A_{i,k}A_{k,j}A_{j,t}A_{t,i}\right) = \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{t=1}^{n}\mathbb{E}(A_{i,k}A_{k,j}A_{j,t}A_{t,i}) \\ &= \sum_{j=1}^{n}\mathbb{E}(A_{i,j}A_{j,j}A_{j,j}A_{j,i}) = \mathbb{E}(A_{i,i}^4) + \sum_{j\neq i}\mathbb{E}(A_{i,j}^2)\mathbb{E}(A_{j,j}^2) = n+2\end{aligned}$$

故

$$\mathbb{E}(\mathrm{tr}(A^4)) = \mathbb{E}\left(\sum_{i=1}^{n} A_{i,i}^{(4)}\right) = \sum_{i=1}^{n}\mathbb{E}\left(A_{i,i}^{(4)}\right) = n^2+2n.$$

综上, $\mathbb{E}(\mathrm{tr}(A^3)) = 0, \mathbb{E}(\mathrm{tr}(A^4)) = 3n.$ ◁

**Problem 4.** 回答下列问题:

(1) 令 $X_1, X_2, \cdots, X_n$ 为独立同分布随机变量, 且 $X_i \sim \mathcal{N}(0,1)$. 令 $Y = \sum_{i=1}^n X_i^2$. 对于任意实数 $t \in [0, 1/4)$, 证明

$$\mathbb{E}\left(e^{t(Y-n)}\right) \le e^{2t^2 n}.$$

提示: 首先考虑 $n=1$ 的情况, 并参考作业三第六题, 以及作业一第三题的提示.

(2) 对于任意 $0 \le \Delta < 1$, 证明

$$\mathbb{P}(Y \ge (1+\Delta)n) \le e^{-n\Delta^2/8}.$$

提示: 根据 $0 \le \Delta < 1$, 选择合适的 $t$ 使得 $t \in [0, 1/4)$, 并利用马尔可夫不等式.

(3) 对于任意 $0 \le \Delta < 1$, 证明

$$\mathbb{P}(Y \le (1-\Delta)n) \le e^{-n\Delta^2/8}.$$

◀

**Solution.** (1) 在作业三第六题中我们已经计算了, 对于 $X \sim \mathcal{N}(0,1)$, 有

$$\mathbb{E}\left(e^{tX^2}\right) = \frac{1}{\sqrt{1-2t}}, \quad \text{for } t \in (-\infty, 1/2).$$

注意到 $\{X_i\}_{i=1}^n \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$, 那么

$$\mathbb{E}\left(e^{t(Y-n)}\right) = e^{-nt}\mathbb{E}\left(e^{t\sum_{i=1}^n X_i^2}\right) = e^{-nt}\prod_{i=1}^{n}\mathbb{E}\left(e^{tX_i^2}\right) = e^{-nt}\frac{1}{(1-2t)^{n/2}}$$

对于 $t \in (-1/4, 1/4)$, 等价于需证明

$$\begin{aligned}\frac{e^{-nt}}{(1-2t)^{n/2}} \le e^{2t^2 n} &\iff (1-2t)^{n/2} \ge e^{-2t^2 n - tn} \\ &\iff \ln(1-2t) \ge -(2t)^2 - 2t\end{aligned}$$

令 $h(t) = \ln(1-2t) + 4t^2 + 2t$, 而

$$h'(t) = \frac{4t(1-4t)}{1-2t} \implies h(t) \text{ 在 } (-1/4, 0) \text{ 上单调递减, 在 } (0, 1/4) \text{ 上单调递增.}$$

那么 $h(t) \ge h(0) = 0$, 证毕.

(2) 取 $t = \Delta/4 \in [0, 1/4)$, 那么

$$\begin{aligned}\mathbb{P}(Y \ge (1+\Delta)n) &= \mathbb{P}\left(e^{tY} \ge e^{t(1+\Delta)n}\right) = \mathbb{P}\left(e^{t(Y-n)} \ge e^{t\Delta n}\right) \\ (\text{Markov's inequality}) \quad &\le \frac{\mathbb{E}\left[e^{t(Y-n)}\right]}{e^{tn\Delta}} \le \frac{e^{2t^2 n}}{e^{tn\Delta}} \\ (t = \Delta/4) \quad &= e^{-n\Delta^2/8}.\end{aligned}$$

证毕.

(3) 类似 (2) 中的证明, 取 $t = \Delta/4 \in [0, 1/4)$, 那么:

$$\begin{aligned}\mathbb{P}(Y \le (1-\Delta)n) &= \mathbb{P}\left(e^{tY} \le e^{t(1-\Delta)n}\right) = \mathbb{P}\left(e^{-t(Y-n)} \ge e^{t\Delta n}\right) \\ (\text{Markov's inequality}) \quad &\le \frac{\mathbb{E}\left[e^{-t(Y-n)}\right]}{e^{tn\Delta}} \\ (-t = -\frac{\Delta}{4} \in (-\frac{1}{4}, 0],\ \text{Use (1)}) \quad &\le \frac{e^{2t^2 n}}{e^{tn\Delta}} = e^{-n\Delta^2/8}.\end{aligned}$$

证毕.

$\triangleleft$