第六次作业参考解答

1. 直接法：

$$\begin{aligned}
P(Y=k) &= P(X_1+X_2=k) \\
&= \sum_{k_1=0}^{k} P(X_1=k_1, X_2=k-k_1) \\
&= \sum_{k_1=0}^{k} P(X_1=k_1)P(X_2=k-k_1) \\
&= \sum_{k_1=0}^{k} \frac{\lambda_1^{k_1}}{k_1!}e^{-\lambda_1}\frac{\lambda_2^{k-k_1}}{(k-k_1)!}e^{-\lambda_2} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{k_1=0}^{k}\frac{k!}{k_1!(k-k_1)!}\lambda_1^{k_1}\lambda_2^{k-k_1} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}(\lambda_1+\lambda_2)^k
\end{aligned}$$

因此 Y 服从参数为 $\lambda_1+\lambda_2$ 的泊松分布

矩母函数法：作业 7-5

一种直观解释：泊松分布多用于表示一定时间或一定空间内出现的小概率事件次数。当考虑 $Y=X_1+X_2$ 时，实际上考虑了两个过程中发生的事件总次数。而事件的总数仍然保持随机和独立，从而形成一个新的泊松分布。

2. X, Y 的联合概率密度函数为

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

(1) 引入 $X'=X$, 有

$$\begin{cases} X'=X \\ Z=\dfrac{Y}{X} \end{cases} \Rightarrow \begin{cases} X=X' \\ Y=X'Z \end{cases}$$

那么 (X',Z) 的联合概率密度函数为 $g(x',z)=f(x',x'z)|J|$, 其中

$$J=\det\begin{bmatrix} \frac{\partial x}{\partial x'} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial x'} & \frac{\partial y}{\partial z} \end{bmatrix}=\det\begin{bmatrix} 1 & 0 \\ z & x' \end{bmatrix}=x'$$

因此 Z 的概率密度函数为

$$\begin{aligned}
g_Z(z) &= \int_{-\infty}^{\infty} f(x',x'z)|J|\mathrm{d}x' = \int_{-\infty}^{\infty}\frac{1}{2\pi}e^{-\frac{x'^2(1+z^2)}{2}}|x'|\mathrm{d}x' \\
&= \frac{1}{\pi}\int_0^{\infty} x'e^{-\frac{x'^2(1+z^2)}{2}}\mathrm{d}x' \\
&= \frac{1}{2\pi}\int_0^{\infty} e^{-\frac{t^2(1+z^2)}{2}}\mathrm{d}t \\
&= \frac{1}{2\pi}\left(-\frac{2}{1+z^2}e^{-\frac{t^2(1+z^2)}{2}}\bigg|_0^{+\infty}\right) \\
&= \frac{1}{\pi(1+z^2)}
\end{aligned}$$

Tips: 注意本题和课堂例题有区别，没有限制 $X > 0$.

(2)

$$J = \det \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix} = \det \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix} = r(\cos^2\theta + \sin^2\theta) = r$$

(R, Θ) 的联合概率密度函数为

$$g(r,\theta) = f(r\cos\theta, r\sin\theta)|J| = \frac{r}{2\pi}e^{-\frac{r^2}{2}}, \quad r \geq 0, 0 \leq \theta < 2\pi$$

同时 $g(r,\theta)$ 可以分解为以下两个函数的乘积:

$$g_R(r) = \begin{cases} re^{-\frac{r^2}{2}} & ,r \geq 0 \\ 0 & ,r < 0 \end{cases}, \quad g_\Theta(\theta) = \begin{cases} \frac{1}{2\pi} & ,0 \leq \theta < 2\pi \\ 0 & ,\text{其他情况} \end{cases}$$

所以 R, Θ 独立。

(3) 逆映射为

$$\begin{cases} X = \frac{1}{2}(U+V) \\ Y = \frac{1}{2}(U-V) \end{cases}$$

$$J = \det \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial v} & \frac{\partial y}{\partial v} \end{bmatrix} = \det \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} = -\frac{1}{2}$$

(U, V) 的联合概率密度函数为

$$g(u,v) = f(\frac{u+v}{2}, \frac{u-v}{2})|J| = \frac{1}{4\pi}e^{-\frac{u^2+v^2}{4}}$$

同时 $g(u,v)$ 可以分解:

$$g(u,v) = \left(\frac{1}{4\pi}e^{-\frac{u^2}{4}}\right)\left(e^{-\frac{v^2}{4}}\right)$$

所以 U, V 独立.

Tips: (3) 写成这个形式只是为了说明分解以后的两个函数可以不是概率密度函数。

3. 根据累积分布函数的定义直接求解即可

$$F_Y(y) = P(Y \leq y) = \prod_{i=1}^{n} P(X_i \leq y) = F^n(y)$$

$$F_Z(z) = P(Z \leq z) = 1 - P(Z > z) = 1 - \prod_{i=1}^{n}[1 - P(X_i \leq z)] = 1 - (1 - F(z))^n$$

Tips: 注意 $P(Z > z)$ 不是分布函数。

4. (1) 卡方分布：若 $X_1, X_2, \cdots, X_n$ 相互独立，都服从正态分布 $N(0,1)$, 则 $Y = X_1^2 + \cdots + X_n^2$ 服从自由度

为 n 的卡方分布 χ_n^2。概率密度函数

$$k_n(x)=\begin{cases}\dfrac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}e^{-\frac{x}{2}}x^{\frac{n-2}{2}}, & x>0\\ 0, & x\leq 0\end{cases}$$

(2) t 分布：设 X_1,X_2 独立，$X_1\sim\chi_n^2, X_2\sim N(0,1)$, 而 $Y=\frac{X_2}{\sqrt{\frac{X_1}{n}}}$, 则 Y 服从自由度 n 的 t 分布。概率密度函数

$$t_n(y)=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{y^2}{n})^{-\frac{n+1}{2}}$$

(3) F 分布：设 X_1,X_2 独立，$X_1\sim\chi_n^2, X_2\sim\chi_m^2$, 则 $Y=\frac{m^{-1}X_2}{n^{-1}X_1}$ 服从自由度为 (m,n) 的 F 分布。概率密度函数

$$f_{mn}(y)=\begin{cases}m^{\frac{m}{2}}n^{\frac{n}{2}}\dfrac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}y^{\frac{m}{2}-1}(my+n)^{-\frac{m+n}{2}}, & y>0\\ 0 & ,y\leq 0\end{cases}$$

5. (1) 设三匹马的赔率分别为 r_A,r_B,r_C. 同时每匹马上的下注金额为 x_A,x_B,x_C. 投注站设定的抽成比例为 β. 那么假设 A 马最终获胜，则有

$$x_A(1+r_A)=(1-\beta)(x_A+x_B+x_C)$$

B,C 马获胜的情况也同理。代入数据后有

$$\begin{cases}r_A=\dfrac{(1-\beta)(x_A+x_B+x_C)}{x_A}-1=\dfrac{9}{10}\\ r_B=\dfrac{(1-\beta)(x_A+x_B+x_C)}{x_B}-1=\dfrac{13}{6}\\ r_C=\dfrac{(1-\beta)(x_A+x_B+x_C)}{x_C}-1=\dfrac{15}{4}\end{cases}$$

Tips: 本题给出了赔率的定义，建议按照题目的要求来求解。

(2) 假设某匹马的获胜概率为 p，失败概率 $1-p$. 如果投资这匹马的资金为 x, 获胜后的额外收益即为 rx. 那么在市场的视角下，应有 $prx-(1-p)x=0$, 解得 $p=\frac{1}{1+r}$. 代入数据后可得每匹马隐含的获胜概率为

$$\begin{cases}p_A=\dfrac{10}{19}\\ p_B=\dfrac{6}{19}\\ p_C=\dfrac{4}{19}\end{cases}$$

(3) $p_A+p_B+p_C=\frac{20}{19}>1$.

一种看法：对比作业 2-14(2)，二者的原理都是一致的。参与投注的各方在总体上高估了自己获胜的概率 (在之前的题目中表现为预计对手获胜的概率之和小于 1, 在本题中表现为预计自己获胜的概率之和大于 1)，因此都会额外投注更多与获胜概率不匹配的资金或者接受事实上偏低的赔率。这部分额外投注就会被规则设计者拿走。

6. (1) 错误。

$$
\begin{aligned}
Var(XY) &= E(X^2Y^2) - E^2(XY) \\
&= E(X^2)E(Y^2) - E^2(X)E^2(Y) \\
Var(X)Var(Y) &= (E(X^2) - E^2(X))(E(Y^2) - E^2(Y)) \\
&= E(X^2)E(Y^2) - E(X^2)E^2(Y) - E^2(X)E(Y^2) + E^2(X)E^2(Y) \\
Var(XY) - Var(X)Var(Y) &= E(X^2)E^2(Y) + E^2(X)E(Y^2) - 2E^2(X)E^2(Y) \\
&= E^2(X)Var(Y) + E^2(Y)Var(X)
\end{aligned}
$$

上式不一定为 0.

事实上可以看出 $Var(XY) \geq Var(X)Var(Y)$.

(2) 错误，X 的中位数不一定唯一，而 $E(X)$ 为确定的值。或者也可以举出具体的例子说明，比如 X 的分布有一定的偏向时，很容易构造出反例。

7. 证明：由期望的线性性质和方差的定义可知

$$
\begin{aligned}
E((X-c)^2) - Var(X) &= E(X^2 - 2cX + c^2) - E(X^2) + E^2(X) \\
&= -2cE(X) + c^2 + E^2(X) \\
&= (E(X) - c)^2 \\
&\geq 0
\end{aligned}
$$

当且仅当 $c = E(X)$ 时等号成立。

8. 题目中说明了 X 存在概率密度函数，因此不考虑 X 离散的情况

证明：

$$
\begin{aligned}
E(|X-c|) - E(|X-m|) &= \int_{-\infty}^{\infty} (|x-c| - |x-m|) f(x) \mathrm{d}x \\
&= \int_{-\infty}^{m} (|x-c| - |x-m|) f(x) \mathrm{d}x + \int_{m}^{\infty} (|x-c| - |x-m|) f(x) \mathrm{d}x
\end{aligned}
$$

对 c 的取值分类讨论，当 $c \leq m$ 时，有

$$
|x-c| - |x-m| = \begin{cases} c - m, & x \leq c \\ 2x - c - m, & c < x \leq m \\ m - c, & x > m \end{cases}
$$

此时

$$
\begin{aligned}
\int_{-\infty}^{m} (|x-c| - |x-m|) f(x) \mathrm{d}x &\geq (c-m) \int_{-\infty}^{m} f(x) \mathrm{d}x = -\frac{1}{2}(m-c) \\
\int_{m}^{\infty} (|x-c| - |x-m|) f(x) \mathrm{d}x &= (m-c) \int_{m}^{\infty} f(x) \mathrm{d}x = \frac{1}{2}(m-c) \\
E(|X-c|) - E(|X-m|) &\geq 0
\end{aligned}
$$

当 $c > m$ 时，有

$$|x-c|-|x-m| = \begin{cases} c-m, & x \le m \\ m+c-2x, & m < x \le c \\ m-c, & x > c \end{cases}$$

此时

$$\begin{aligned} &\int_{-\infty}^{m} (|x-c|-|x-m|)f(x)\mathrm{d}x = (c-m)\int_{-\infty}^{m} f(x)\mathrm{d}x = -\frac{1}{2}(m-c) \\ &\int_{m}^{\infty} (|x-c|-|x-m|)f(x)\mathrm{d}x \ge (m-c)\int_{m}^{\infty} f(x)\mathrm{d}x = \frac{1}{2}(m-c) \\ &E(|X-c|)-E(|X-m|) \ge 0 \end{aligned}$$

综上所述，对于任何常数 c 都成立不等式 $E(|X-c|) \ge E(|X-m|)$.

9. 对数正态分布的概率密度函数 $f(x) = \frac{1}{x\sqrt{2\pi}\sigma}\exp\left\{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right\}, \quad x > 0.$

期望：

$$\begin{aligned} E(X) &= \int_0^{\infty} xf(x)\mathrm{d}x \\ &= \int_0^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right\}\mathrm{d}x \\ &\overset{令 t=\ln x}{=} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\}e^t\mathrm{d}t \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{t^2-2t\mu+\mu^2-2\sigma^2 t}{2\sigma^2}\right\}\mathrm{d}t \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(t-(\mu+\sigma^2))^2}{2\sigma^2}\right\}\exp\left\{\mu+\frac{\sigma^2}{2}\right\}\mathrm{d}t \\ &= \exp\left\{\mu+\frac{\sigma^2}{2}\right\} \end{aligned}$$

最后一步是因为正态分布 $N(\mu+\sigma^2, \sigma^2)$ 的密度函数在 $\mathbb{R}^2$ 上的积分为 *1*.

方差：

$$\begin{aligned} E(X^2) &= \int_0^{\infty} x^2 f(x)\mathrm{d}x \\ &= \int_0^{\infty} \frac{x}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right\}\mathrm{d}x \\ &\overset{令 t=\ln x}{=} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\}e^{2t}\mathrm{d}t \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(t-(\mu+2\sigma^2))^2}{2\sigma^2}\right\}\exp\left\{2\mu+2\sigma^2\right\}\mathrm{d}t \\ &= \exp\left\{2\mu+2\sigma^2\right\} \\ Var(X) &= E(X^2)-E^2(X) \\ &= \exp\left\{2\mu+2\sigma^2\right\} - \exp\left\{2\mu+\sigma^2\right\} = e^{2\mu+\sigma^2}(e^{\sigma^2}-1) \end{aligned}$$

10.

$$Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^n X_i}{n}\right) = Var\left(\sum_{i=1}^n \frac{X_i}{n}\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$
\begin{aligned}
E(\sum_{i=1}^{n}(X_i-\bar{X})^2) &= E(\sum_{i=1}^{n}(X_i^2-2X_i\bar{X}+\bar{X}^2)) \\
&= E(\sum_{i=1}^{n}X_i^2)-2E(\bar{X}\sum_{i=1}^{n}X_i)+nE(\bar{X}^2) \\
&= E(\sum_{i=1}^{n}X_i^2)-nE(\bar{X}^2) \\
E(S^2) &= E(\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{n-1}) \\
&= \frac{1}{n-1}(\sum_{i=1}^{n}E(X^2)-nE(\bar{X}^2)) \\
&= \frac{1}{n-1}(\sum_{i=1}^{n}(\sigma^2+\mu^2)-n(\frac{\sigma^2}{n}+\mu^2)) \\
&= \sigma^2
\end{aligned}
$$

11. (1)(2)(3)(4) 互相等价。首先 $\mathrm{Cov}(X,Y)=0$ 即是 X 与 Y 不相关的定义。(2) 与 (1) 等价。

$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)$,$\mathrm{Cov}(X,Y)=0$ 等价于 $E(XY)=E(X)E(Y)$. (1) 与 (3) 等价。

$Var(X+Y)=E((X+Y)^2)-E^2(X+Y)=E(X^2+2XY+Y^2)-E^2(X)-E^2(Y)-2E(X)E(Y)=Var(X)+Var(Y)+2\mathrm{Cov}(X,Y)$, 则 $Var(X+Y)=Var(X)+Var(Y)$ 等价于 $\mathrm{Cov}(X,Y)=0$. (1) 与 (4) 等价。

备注： 如果把每个命题抽象成一个节点，把已经证明的命题之间的关系看做有向边，我们的证明过程会在这个图中不断添加连边。当这个图成为一个强连通的图时，即完成了所有命题之间相互等价的证明。通常来说，最常用的策略是以一个命题为前提，推导出下一个命题作为结论，构造一个单向的环状图。以上的证明策略略有不同，但也满足要求。注意等价关系是双向边。

Tips: 注意独立和相关在定义上区别。我们很多时候用两个随机变量独立来说明它们满足上述性质，但并不意味着这是随机变量独立的定义。

12. 证明：

$$
\mathrm{Corr}(X,Y)=E(\frac{X-\mu_1}{\sigma_1}\frac{Y-\mu_2}{\sigma_2})=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2})f(x,y)\mathrm{d}x\mathrm{d}y
$$

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

$$
\begin{aligned}
\mathrm{Corr}(X,Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}uv\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)\right\}\sigma_1\sigma_2\mathrm{d}u\mathrm{d}v \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}uv\exp\left\{-\frac{1}{2(1-\rho^2)}[(u-\rho v)^2+(1-\rho^2)v^2]\right\}\mathrm{d}u\mathrm{d}v \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}ve^{-\frac{v^2}{2}}\left(\int_{-\infty}^{\infty}u\frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left\{-\frac{(u-\rho v)^2}{2(1-\rho^2)}\right\}\mathrm{d}u\right)\mathrm{d}v \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\rho v^2e^{-\frac{v^2}{2}}\mathrm{d}v \\
&= \frac{\rho}{\sqrt{2\pi}}\sqrt{2\pi} \\
&= \rho
\end{aligned}
$$

Tips: 本题需要验证这一结论，不建议直接从 $\mathrm{Cov}(X,Y)=\rho\sigma_1\sigma_2$ 出发，本题需要延续课堂上的过程进行计算。

13. 设随机变量 X_i 表示第 i 个人是否拿到了自己的帽子, 如果拿到则 $X_i=1$, 否则 $X_i=0$, 那么 $X=\sum_{i=1}^n X_i$. 且已知 $P(X_i=1)=\frac{1}{n}$.

$$
\begin{aligned}
E(X) &= E(\sum_{i=1}^n X_i) = n\cdot\frac{1}{n} = 1 \\
E(X^2) &= E((\sum_{i=1}^n X_i)^2) \\
&= E(\sum_{i=1}^n X_i^2 + 2\sum_{1\le i<j\le n} X_iX_j) \\
&= 1 + 2\sum_{1\le i<j\le n} E(X_iX_j)
\end{aligned}
$$

我们考察 X_iX_j, 当 $X_iX_j=1$ 时说明对应的两个人都拿到了自己的帽子

$$
P(X_i=1,X_j=1) = P(X_i=1|X_j=1)P(X_j=1) = \frac{1}{(n-1)n}
$$

因此

$$
\begin{aligned}
E(X^2) &= 1 + 2\sum_{1\le i<j\le n} E(X_iX_j) \\
&= 1 + 2\binom{n}{2}\frac{1}{(n-1)n} \\
&= 2 \\
Var(X) &= E(X^2) - E^2(X) = 2 - 1 = 1
\end{aligned}
$$

Tips: 用 $n\to\infty$ 时的近似进行计算，虽然能够求出正确的结果，但逻辑上来说不是很严谨。

14. 证明:

(1) 构造二次函数 $f(t)=t^2E(U^2)+2tE(UV)+E(V^2), t\in\mathbb{R}$, 则我们有

$$
f(t) = E(t^2U^2+2tUV+V^2) = E(tU+V)^2 \ge 0
$$

这意味着此二次函数在实数范围内至多有一个零点，判别式

$$
\begin{aligned}
\Delta &= (2E(UV))^2 - 4E(U^2)E(V^2) \\
&= 4E^2(UV) - 4E(U^2)E(V^2) \\
&\le 0 \\
E^2(UV) &\le E(U^2)E(V^2)
\end{aligned}
$$

等号成立条件为 $tU+V=0$ 成立，也即存在常数 c 使得 $V=cU$ 恒成立，$P(V=cU)=1$

备注: 这实际上是 *Cauchy-Schwartz* 不等式的期望版本，在之前的微积分和线性代数课程中还讲过积分版本和向量内积的版本，其数学本质是类似的。但也因为本质上是在证明 *Cauchy-Schwartz* 不等式，

所以不建议直接用 Cauchy-Schwartz 不等式来证明。

(2) 设 $U = \frac{X-\mu_1}{\sigma_1}, V = \frac{Y-\mu_2}{\sigma_2}$, 则 U,V 都是均值为 0，方差为 1 的随机变量。由 (1) 得 $|\mathrm{Corr}(X,Y)| = |E(\frac{X-\mu_1}{\sigma_1}\frac{Y-\mu_2}{\sigma_2})| = |E(UV)| \le \sqrt{E^2(U)E^2(V)} = 1$, 等号成立的条件为存在常数 c 使得 $V = cU$，也即存在常数 a,b 使得 $Y = aX+b, P(Y = aX+b) = 1$. 这里的 $a \neq 0$ 需要从 U,V 都不是常量，即 σ_1,σ_2 均严格大于 0 说明。

15. (1) 证明:

$$\begin{aligned}
\mathrm{Cov}(X_i-\bar{X},\bar{X}) &= \mathrm{Cov}(X_i,\bar{X}) - \mathrm{Cov}(\bar{X},\bar{X}) \\
&= \mathrm{Cov}(X_i,\sum_{k=1}^{n}\frac{X_k}{n}) - Var(\bar{X}) \\
&= \sum_{k=1}^{n}\mathrm{Cov}(X_i,\frac{X_k}{n}) - \frac{\sigma^2}{n} \\
\text{当}i \neq k\text{时，两随机变量独立，}&\mathrm{Cov}(X_i,\frac{X_k}{n}) = 0 \\
&= \frac{1}{n}\mathrm{Cov}(X_i,X_i) - \frac{\sigma^2}{n} \\
&= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} \\
&= 0
\end{aligned}$$

(2) 不一定。以 Bernoulli 分布为例说明，假设 X_1,X_2 独立，均服从以 p 为参数的 Bernoulli 分布，则 $P(X_1=1)=p, P(X_1=0)=1-p$. 同时 $P(\bar{X}=0) = P(X_1=0)P(X_2=0) = (1-p)^2$,$P(\bar{X}=1) = P(X_1=1)P(X_2=1) = p^2$. 则 $P(X_1-\bar{X}=0) = P(X_1=X_2) = p^2+(1-p)^2$. 然而

$$P(\bar{X}=0, X_1-\bar{X}=0) = P(X_1=X_2=0) = (1-p)^2$$
$$P(\bar{X}=0)P(X_1-\bar{X}=0) = (1-p)^2\cdot[p^2+(1-p)^2]$$

当 $0<p<1$ 时，有 $P(\bar{X}=0, X_1-\bar{X}=0) \neq P(\bar{X}=0)P(X_1-\bar{X}=0)$. 说明 $X_i-\bar{X}$ 与 $\bar{X}$ 不独立.

16. *注意：本题中题目所述的绘制方法和当下 Q-Q 图主流的绘制方式有出入。在 Python/R/Matlab 相关包的实现中，Q-Q 图的横纵坐标与题目中所述相反。如果想严格按照题目中的要求绘制，需要手动进行代码实现或者调整调用函数得到的图形对象。*

(1) 绘图略。理论上此时的正态 Q-Q 图近似为一条直线，且斜率近似为 1.

(2) 绘图略。理论上此时的正态 Q-Q 图不能近似为一条直线。

(3) 可以。只需要将计算纵坐标时所使用的函数改为所假设分布对应累积分布函数的反函数即可。

(4) 可以。取 x 轴和 y 轴分别表示两个数据集中数据的观测分位数，观察 Q-Q 图是否近似为直线。

一部分同学指出仅从 Q-Q 图的形态是否是直线不足以判断同分布，如无法看出具体的参数是否匹配，或者认为不足以说明来自同一未知总体，均有一定道理。在具体的推广上有更多的细节需要考虑，但 Q-Q 图确实是一种用于比较分布的非参数方法，如有兴趣可自行查阅相关资料。

作业总结:

- 本次作业的内容主要涉及随机向量的函数，密度函数变换法，期望，方差，协方差，相关与独立等。
- 主要集中出现的问题包括：第 2 题中的基本方法和概念出错，如对连续型随机变量讨论 $P(Z=z)$，用协方差论证独立性。第 8 题的证明完成度 (除了参考解答中的方法，当然还有其他的方法或者写法)。第 10 题求 $E(S^2)$ 时会混淆相关记号的含义。第 11 题的基本概念不清。第 12 题和第 14 题属于验证性质的证明，或者从课堂内容出发进行补充，部分同学直接使用了几乎完全等价的定理进行证明。除此之外，部分“说明，举例，直观理解”类的题目论证和表达不够清晰和完善。