Homework 6

抹茶奶绿

March 29, 2025

Exercise 1.

设 $X_1 \sim \text{Poisson}(\lambda_1), X_2 \sim \text{Poisson}(\lambda_2)$ 且相互独立，求随机变量 $Y = X_1 + X_2$ 的分布，并给出直观解释.

Solution.

由于

$$M_{X_1+X_2}(t) = M_{X_1}(t)M_{X_2}(t) = e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}$$

因此 $Y = X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

若两个独立的泊松过程分别描述两种独立的随机事件的发生，则合并后的过程的发生率为两者之和.

Exercise 2.

假设随机变量 X 与 Y 相互独立，且都服从标准正态分布.

I 求

$$Z = \frac{Y}{X}(X \neq 0)$$

的概率密度函数.

II 令

$$X = R\cos\Theta, Y = R\sin\Theta$$

计算 (R, Θ) 的概率密度函数，并确定 Θ 与 R 是否独立.

III 令

$$U = X + Y, V = X - Y$$

求 (U, V) 的联合概率密度函数，并确定 U 与 V 是否独立.

Solution.

I 由 X 和 Y 的联合分布可得

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

令 $Z = \frac{Y}{X}$, 则 $Y = ZX$, $X = X$, Jacobian 为 $|J| = |x|$.

$$f_{Z,X}(z, x) = f_{X,Y}(x, zx)|J| = \frac{1}{2\pi}e^{-\frac{x^2(1+z^2)}{2}}|x|$$

由此可得边际分布

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{Z,X}(z, x)dx = \frac{1}{\pi(1+z^2)}$$

因此 Z 服从标准 Cauchy 分布.

II 由极坐标变换可得

$$f_{R,\Theta}(r, \theta) = f_{X,Y}(r\cos\theta, r\sin\theta) \cdot r = \frac{1}{2\pi}e^{-\frac{r^2}{2}}$$

故变量是分离的，因此 Θ 与 R 独立.

III 由线性变换可得

$$f_{U,V}(u, v) = f_{X,Y}\left(\frac{u+v}{2}, \frac{u-v}{2}\right) \cdot \frac{1}{4} = \frac{1}{4\pi}e^{-\frac{u^2+v^2}{4}}$$

故变量是分离的，因此 U 与 V 独立.

Exercise 3.
设随机变量 $X_1, X_2, \ldots, X_n$ 独立同分布，其分布函数为 $F(x)$,
令
$$Y = \max\{X_1, X_2, \ldots, X_n\}, Z = \min\{X_1, X_2, \ldots, X_n\}$$
分别求 $F_Y(y)$ 和 $F_Z(z)$.

Solution.

$$F_Y(y) = P(Y \le y) = P(X_1 \le y, \ldots, X_n \le y) = [F(y)]^n$$

$$F_Z(z) = P(Z \le z) = 1 - P(Z > z) = 1 - P(X_1 > z, \ldots, X_n > z) = 1 - [1 - F(z)]^n$$

Exercise 4.
了解统计上（与正态分布相关）的三大分布：卡方分布、t 分布和 F 分布（参阅 John Rice 的《数理统计与数据分析》或其他资料），给出其定义.

Solution.

I 卡方分布：若 $Z_1, \ldots, Z_k$ 为相互独立的标准正态随机变量，则随机变量
$$\chi^2 = \sum_{i=1}^{k} Z_i^2$$
服从自由度为 k 的卡方分布，记作 $\chi^2(k)$.

II t 分布：设 $Z \sim N(0,1)$ 与 $\chi^2(k)$ 独立，则随机变量
$$T = \frac{Z}{\sqrt{\chi^2(k)/k}}$$
服从自由度为 k 的 t 分布，记作 $t(k)$.

III F 分布：设 $\chi^2(m)$ 与 $\chi^2(n)$ 独立，则随机变量
$$F = \frac{\chi^2(m)/m}{\chi^2(n)/n}$$
服从 F 分布，记作 $F(m,n)$.

Exercise 5.

假设有一场 3 匹马（分别记为 A, B, C）的比赛，在下注结束时假设有 500 元下注在 A 马，300 元下注在 B 马，200 元下注在 C 马. 如果投注站想确保每下注 100 元中可赚取 5 元，则

I 投注站设定的每匹马的赔率应该是多少？

II 此时赔率所隐含的每匹马获胜概率是多少？（对比作业题 1-5）

III 比较 3 匹马的隐含获胜概率之和与 1 的大小，你对此有什么看法？（对比作业题 2-14（2））

Solution.

I 在众筹投注中，常用的做法是将总赌资扣除掉抽水后作为奖金池。总赌资为

$$500 + 300 + 200 = 1000\text{元},$$

抽水 5% 后奖金池为

$$1000 \times (1 - 0.05) = 950\text{元}.$$

若某匹马获胜，则投注该马者按比例瓜分奖金池，赔率由奖金池与该马投注额的比值决定。对于马 i，若投注额为 M_i，则其赔率（指总返还额与投注额之比）为

$$\text{赔率}_i = \frac{950}{M_i}.$$

因此：

- A 马：投注 500 元，其赔率为 $950/500 = 1.9$
- B 马：投注 300 元，其赔率为 $950/300 \approx 3.1667$
- C 马：投注 200 元，其赔率为 $950/200 = 4.75$

II 赔率所隐含的每匹马获胜概率即为赔率的倒数，分别为：

- A 马：$500/950 \approx 0.5263$
- B 马：$300/950 \approx 0.3158$
- C 马：$200/950 \approx 0.2105$

III 3 匹马的隐含获胜概率之和为 $1000/950 \approx 1.0526$，大于 1.
这是由于投注站的“抽水”让下注的人都觉得赔率还可以，但其实是压低了他们的回报率.

Exercise 6.

判断下列结论对错并说明理由，假设所涉及的期望和方差皆存在.

I 若 X 与 Y 独立，则 $\mathrm{Var}(XY) = \mathrm{Var}(X)\,\mathrm{Var}(Y)$.

II X 的中位数若存在则一定等于 $E(X)$.

Solution.

I 错误，由独立性有

$$\mathrm{Var}(XY) = E[X^2]E[Y^2] - \big(E[X]E[Y]\big)^2$$

而

$$\mathrm{Var}(X)\,\mathrm{Var}(Y) = \big(E[X^2] - E[X]^2\big)\big(E[Y^2] - E[Y]^2\big)$$

当且仅当 $E[X]E[Y] = 0$ 时二者相等.

II 错误，中位数是使得 $P(X \le m) = 0.5$ 的值，而期望是加权平均值，二者不一定相等.

例如，考虑偏斜分布:

x	$P(X=x)$
0	0.7
10	0.3

此时中位数 $m = 0$ (因为 $P(X \le 0) = 0.7 > 0.5$)，而期望 $E(X) = 0 \times 0.7 + 10 \times 0.3 = 3$，两者不相等.

Exercise 7.

证明：对任何常数 c 有 $E((X-c)^2) \ge \mathrm{Var}(X)$，当且仅当 $c = E(X)$ 时取等.

Solution.

$$\begin{aligned} E\big[(X-c)^2\big] &= E\big[(X - E(X) + E(X) - c)^2\big] \\ &= E\big[(X - E(X))^2\big] + 2(E(X) - c)E[X - E(X)] + (E(X) - c)^2 \\ &= \mathrm{Var}(X) + (E(X) - c)^2 \end{aligned}$$

Exercise 8.

设 X 有概率密度函数 $f(x)$，其中位数为 m. 证明：对任何常数 c 都成立不等式

$$E\big(|X-c|\big) \ge E\big(|X-m|\big)$$

Solution.

记

$$\phi(c) = E(|X-c|) = \int_{\mathbb{R}} |x-c| f(x)dx = \int_{-\infty}^{c} (c-x)f(x)dx + \int_{c}^{+\infty} (x-c)f(x)dx$$

那么

$$\phi'(c) = \int_{-\infty}^{c} f(x)dx - \int_{c}^{+\infty} f(x)dx = 2F(c) - 1$$

依定义即知 $E(|X-c|)$ 最小时 $c = m$.

Exercise 9.

计算对数正态分布的期望和方差（对数正态分布定义可参见作业题 4-9）.

Solution.

设 X 服从对数正态分布，即 $\ln X \sim N(\mu, \sigma^2)$. 令 $Y = \ln X$，则 Y 的矩母函数为

$$M_Y(t) = E\left[e^{tY}\right] = \exp\!\left(t\mu + \frac{t^2\sigma^2}{2}\right)$$

那么

$$E[X] = E\left[e^{Y}\right] = M_Y(1) = \exp\!\left(\mu + \frac{\sigma^2}{2}\right)$$

$$E[X^2] = E\left[e^{2Y}\right] = M_Y(2) = \exp\!\left(2\mu + 2\sigma^2\right)$$

于是

$$\mathrm{Var}(X) = E[X^2] - \left(E[X]\right)^2 = \exp\!\left(2\mu + 2\sigma^2\right) - \left[\exp\!\left(\mu + \frac{\sigma^2}{2}\right)\right]^2 = \exp\!\left(2\mu + \sigma^2\right)\left[\exp\!\left(\sigma^2\right) - 1\right]$$

Exercise 10.

设随机变量 $X_1, X_2, \ldots, X_n$ 独立同分布（这样的序列也称为来自同一分布的随机样本），其公共期望为 μ，公共方差为 σ^2.

记

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

求 $\mathrm{Var}(\bar{X})$ 和 $E(S^2)$.

Solution.

I 由于 X_i 相互独立，故有

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

II 由 S^2 定义有

$$(n-1)S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2$$

取期望，有

$$E\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \sum_{i=1}^{n} E[(X_i - \mu)^2] - n\,E[(\bar{X} - \mu)^2] = n\sigma^2 - \sigma^2 = (n-1)\sigma^2$$

故

$$E(S^2) = \sigma^2$$

Exercise 11.

下列叙述是否等价? 请说明理由.

I $\mathrm{Cov}(X, Y) = 0$;

II X 与 Y 不相关;

III $E(XY) = E(X)E(Y)$;

IV $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Solution.

变量不相关的定义即为协方差为 0，因此 I 与 II 等价，而 III 也是协方差为零的等价表述，因为

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

另外，由

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$$

当 $\mathrm{Cov}(X, Y) = 0$ 时 IV 也成立. 因此，这四个叙述在期望与方差存在的条件下是等价的.

Exercise 12.

验证：若 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ ，则 $\rho = \mathrm{Corr}(X, Y)$.

Solution.

引入标准化变量：

$$U = \frac{X - \mu_1}{\sigma_1}, \quad V = \frac{Y - \mu_2}{\sigma_2}$$

则 (U, V) 服从标准二元正态分布 $N(0, 0, 1, 1, \rho)$，其协方差矩阵为：

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

对协方差矩阵 Σ 进行特征分解得：

- 特征值：$\lambda_1 = 1 + \rho$，$\lambda_2 = 1 - \rho$
- 正交特征向量：$\frac{1}{\sqrt{2}}[1, 1]^T$ 和 $\frac{1}{\sqrt{2}}[1, -1]^T$

即正交变换矩阵为：

$$A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

应用变换得到新变量：

$$Z_1 = \frac{U + V}{\sqrt{2}}, \quad Z_2 = \frac{U - V}{\sqrt{2}}$$

那么新变量的协方差矩阵为：

$$A \Sigma A^T = \begin{bmatrix} 1 + \rho & 0 \\ 0 & 1 - \rho \end{bmatrix}$$

其中 Z_1 和 Z_2 独立且方差分别为 $1 + \rho$ 和 $1 - \rho$.

那么

$$\begin{aligned}
\text{Corr}(X,Y) = \text{Cov}(U,V) &= \text{Cov}\left(\frac{Z_1+Z_2}{\sqrt{2}}, \frac{Z_1-Z_2}{\sqrt{2}}\right) \\
&= \frac{1}{2}\text{Cov}(Z_1+Z_2, Z_1-Z_2) \\
&= \frac{1}{2}\left[\text{Cov}(Z_1,Z_1) - \text{Cov}(Z_1,Z_2) + \text{Cov}(Z_2,Z_1) - \text{Cov}(Z_2,Z_2)\right] \\
&= \frac{1}{2}\left[\text{Var}(Z_1) - \text{Var}(Z_2)\right] \\
&= \frac{1}{2}\left((1+\rho) - (1-\rho)\right) \\
&= \rho
\end{aligned}$$

Exercise 13.

设 X 表示配对问题中“拿到自己帽子的人数”. 求 $E(X)$ 和 $\text{Var}(X)$.

Solution.

设有 N 个人，记 A_i 为第 i 个人拿到自己的帽子，$i = 1, \ldots, N$. 那么

$$P(A_iA_j) = P(A_i)P(A_j \mid A_i) = \frac{1}{N} \cdot \frac{1}{N-1}$$

故

$$E(X) = \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{n} \frac{1}{N} = 1$$

$$E\left[\binom{X}{2}\right] = \sum_{i<j} \frac{1}{N(N-1)} = \binom{N}{2} \cdot \frac{1}{N(N-1)}$$

于是

$$\text{Var}(X) = E[X^2] - (E[X])^2 = 1$$

Exercise 14.

I 证明：

$$E^2(UV) \le E(U^2)E(V^2)$$

且等号成立当且仅当存在常数 t 使得 $P(V = tU) = 1$.

II 利用 I 证明：

$$\left|\text{Corr}(X,Y)\right| \le 1,$$

且等号成立当且仅当存在常数 a, b（其中 $a \neq 0$）使得 $P(Y = aX + b) = 1$.

Solution.

I 对任意实数 $t \in \mathbb{R}$，定义非负随机变量：

$$E\left[(U+tV)^2\right] \ge 0$$

展开平方项：

$$E[U^2] + 2tE[UV] + t^2E[V^2] \ge 0$$

将其视为关于 t 的二次函数：

$$f(t) = E[V^2] \cdot t^2 + 2E[UV] \cdot t + E[U^2]$$

由于 $f(t) \geq 0$ 对所有 t 成立，判别式必须非正：

$$\Delta = (2E[UV])^2 - 4E[U^2] \cdot E[V^2] \leq 0$$

也即

$$E^2[UV] \leq E[U^2] \cdot E[V^2]$$

等号成立当且仅当存在某个 t 使得：

$$E^2(U + tV) = 0 \quad \Rightarrow \quad U + tV = 0 \quad \text{（a.s.）}$$

II 引入标准化变量

$$U = \frac{X - \mu_X}{\sigma_X}, \quad V = \frac{Y - \mu_Y}{\sigma_Y}$$

则

$$\rho(X, Y) = \mathrm{Cov}(U, V) = E(UV)$$

应用 Cauchy-Schwarz 不等式可得

$$|\rho(X, Y)| \leq \sqrt{E(U^2)E(V^2)} = 1$$

等号成立当且仅当存在常数 t 使 $V = tU$ （a.s.），整理可得所需结论，这里 $a = \dfrac{\sigma_Y}{\sigma_X}$.

Exercise 15.

设随机变量 $X_1, X_2, \ldots, X_n$ 独立同分布，其公共期望为 μ，公共方差为 σ^2

I 证明：

$$\mathrm{Cov}\left(X_i, \bar{X} - X_i\right) = 0.$$

II 判断 $X_i - \bar{X}$ 与 $\bar{X}$ 是否一定独立？尝试给出理由.

Solution.

由定义

$$\begin{aligned}
\mathrm{Cov}\left(X_i, \bar{X} - X_i\right) &= \mathrm{Cov}(X_i, \bar{X}) - \mathrm{Cov}(X_i, X_i) \\
&= \mathrm{Cov}\left(X_i, \frac{1}{n}\sum_{j=1}^{n} X_j\right) - \mathrm{Var}(X_i) \\
&= \frac{1}{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) - \sigma^2 \\
&= 0
\end{aligned}$$

最后一个等式成立是因为

$$\mathrm{Cov}(X_i, X_j) = \begin{cases} 0, & i \neq j \quad \text{由独立性} \\ \sigma^2, & i = j \end{cases}$$

通常来说，它们都是不相互独立的. 另外，当 X_i 为正态随机变量时，$\bar{X}$ 不仅与 $X_i - \bar{X}$ 独立，而且与整个序列 $X_j - \bar{X}, j = 1, 2, \ldots, n$ 相互独立. $\bar{X}$ 与样本方差 S^2 也相互独立，并且 $(n-1)S^2/\sigma^2$ 具有自由度为 $(n-1)$ 的 χ^2 分布.

Exercise 16.

利用 Q-Q 图验证正态性.*balabala*... 给出如下问题：

I 生成一组（100 个）标准正态随机数，画出其正态 Q-Q 图，检验是否近似为一条直线；

II 生成一组（100 个）服从参数为 $\lambda = 2$ 的指数分布随机数，画出其正态 Q-Q 图，检验是否近似为一条直线；

III 能否推广上述方法去验证给定数据集是否服从假设的分布？请说明；

IV 能否推广上述方法去验证给定的两个数据集是否来自同一未知总体？请说明.

Solution.

```python
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

normal_data = np.random.normal(0, 1, 100)

stats.probplot(normal_data, dist="norm", plot=plt)
plt.title("Q-Q Plot of Normal Distribution")
plt.show()

exponential_data = np.random.exponential(1/2, 100)

stats.probplot(exponential_data, dist="expon", sparams=(0, 1/2), plot=plt)
plt.title("Q-Q Plot of Exponential Distribution")
plt.show()
```

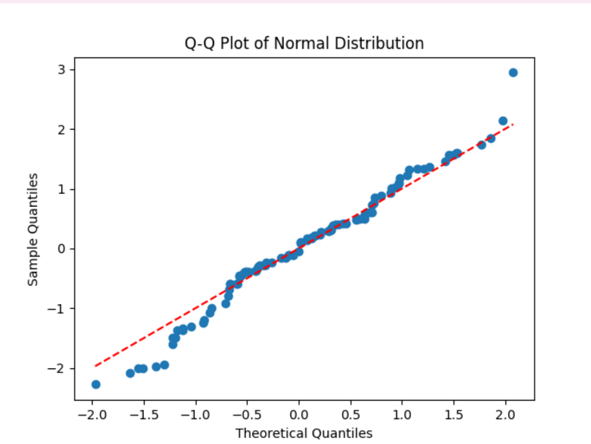

Figure 1: Q-Q Plot of Normal Distribution

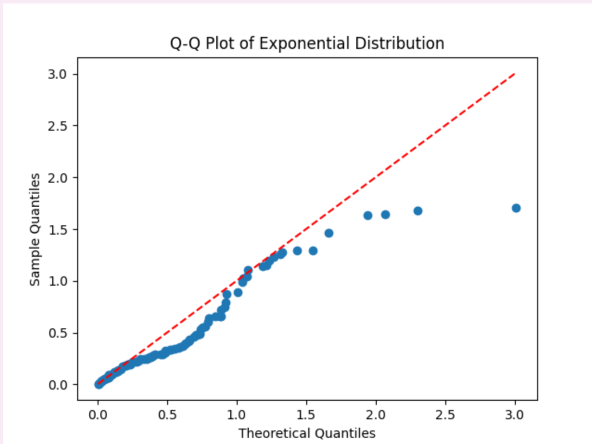

Figure 2: Q-Q Plot of Exponential Distribution

I 确为直线.

II 感觉不像喵.

III 可以，只需将 Φ^{-1} 替换为假设 CDF 的逆即可.

IV 可以，将两数据集分别排序，作出散点图，观察是否接近直线即可.