Homework 7

抹茶奶绿

April 3, 2025

Exercise 1.

设

$$X \sim B(10, 0.9), \quad Y - 8 \sim B(10, 0.1)$$

I 分别画出 X 和 Y 的概率质量函数图.

II 计算并比较 X 和 Y 的均值、方差、中位数和众数.

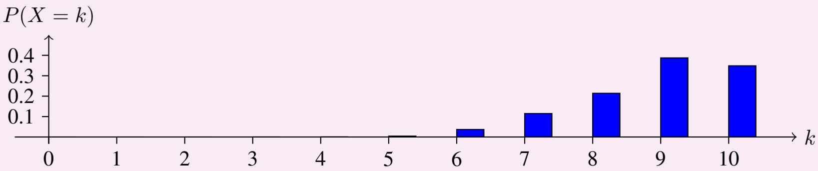
III 计算 X 和 Y 的偏度系数.

Solution.

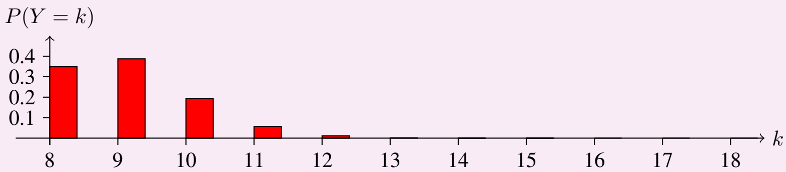
I 对于二项分布 $B(n,p)$，概率质量函数为

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

那么我们有



$X \sim B(10, 0.9)$ 的概率质量函数图



$Y - 8 \sim B(10, 0.1)$ 的概率质量函数图

II

$$E(X) = 10 \times 0.9 = 9, \quad E(Y) = E(Y-8) + 8 = 10 \times 0.1 + 8 = 9$$

$$Var(X) = 10 \times 0.9 \times (1-0.9) = 0.9, \quad Var(Y) = Var(Y-8) = 10 \times 0.1 \times (1-0.1) = 0.9$$

$$Median(X) = 9, \quad Median(Y) = 9$$

$$Mode(X) = 9, \quad Mode(Y) = 9$$

III 令 $I_1, I_2, \ldots, I_n$ 为相互独立的伯努利试验，其中

$$I_i = \begin{cases} 1, & \text{以概率} p, \\ 0, & \text{以概率} 1-p. \end{cases}$$

因此有

$$X = \sum_{i=1}^{n} I_i$$

对于每个 I_i，其中心化变量为 $I_i - p$. 计算其三阶中心矩：

$$E\left[(I_i - p)^3\right] = p\,(1-p)^3 + (1-p)(-p)^3 = p(1-p)(1-2p)$$

由于 I_i 是独立同分布的，因此有

$$E\big[(X-np)^3\big]=\sum_{i=1}^{n}E\big[(I_i-p)^3\big]=n\,p(1-p)(1-2p)$$

那么偏度系数为

$$\text{Skew}(X)=\frac{E\big[(X-np)^3\big]}{(\text{Var}(X))^{3/2}}=\frac{n\,p(1-p)(1-2p)}{\big(np(1-p)\big)^{3/2}}=\frac{1-2p}{\sqrt{np(1-p)}}$$

那么，由平移不变性

$$\text{Skew}(X)=\frac{1-2(0.9)}{\sqrt{10(0.9)(0.1)}}=-0.8433,\quad \text{Skew}(Y)=\text{Skew}(Y-8)=\frac{1-2(0.1)}{\sqrt{10(0.1)(0.9)}}=0.8433$$

Exercise 2.

I 分别计算 $Exp(1)$、$P(4)$ 和 $U(0,1)$ 的偏度系数和峰度系数.

II 分别求出 $Exp(1)$、$P(4)$ 和 $U(0,1)$ 的矩母函数.

III 利用矩母函数计算上述三个分布的偏度系数和峰度系数.

Solution.

I a 对于指数分布 $Exp(1)$，其概率密度函数为

$$f(x)=e^{-x},\quad x>0$$

那么

$$E(X^k)=\int_0^{\infty}x^ke^{-x}dx=k!$$

简单计算知其偏度系数和峰度系数分别为

$$\text{Skew}(X)=2,\quad \text{Kurt}(X)=9$$

b 对于泊松分布 $P(4)$，其概率质量函数为

$$P(X=k)=\frac{4^ke^{-4}}{k!},\quad k=0,1,2,\dots$$

那么

$$E(X^k)=\sum_{m=0}^{\infty}m^k\frac{4^m}{m!}e^{-4}=4\sum_{i=0}^{k-1}\binom{k-1}{i}E(X^i)$$

由递推式简单计算知其偏度系数和峰度系数分别为

$$\text{Skew}(X)=\frac{1}{2},\quad \text{Kurt}(X)=\frac{13}{4}$$

c 对于均匀分布 $U(0,1)$，其概率密度函数为

$$f(x)=\begin{cases}1, & 0<x<1\\ 0, & x<0 \text{ or } x>1\end{cases}$$

那么

$$E(X^k)=\int_0^1x^kdx=\frac{1}{k+1}$$

简单计算知其偏度系数和峰度系数分别为

$$\text{Skew}(X) = 0, \quad \text{Kurt}(X) = \frac{9}{5}$$

II a $Exp(1)$

$$M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx}e^{-x}dx = \frac{1}{1-t}, \quad t < 1$$

b $P(4)$

$$M_X(t) = E[e^{tX}] = \exp(4(e^t - 1))$$

c $U(0,1)$

$$M_X(t) = \int_0^1 e^{tx}dx = \frac{e^t - 1}{t}, \quad t \neq 0, \quad M_X(0) = 1$$

III 分别计算矩母函数的 k 阶导数算出 $E(X^k)$ 即可，不再赘述.

Exercise 3.

计算具有下列矩母函数的连续随机变量 X 的概率密度函数：

$$M_X(t) = \frac{1}{3} \cdot \frac{2}{2-t} + \frac{2}{3} \cdot \frac{3}{3-t}$$

Solution.

这显然是指数分布的形式，于是

$$f(x) = \frac{2}{3}e^{-2x} + 2e^{-3x}, \quad x > 0$$

经检验符合.

Exercise 4.

设 $X \sim N(\mu, \sigma^2)$，令 $Y = e^X$（对数正态分布）

I 证明对数正态分布的矩母函数不存在.

II 利用正态分布的矩母函数计算 Y 的各阶原点矩.

Solution.

I

$$M_Y(t) = E[e^{tY}] = E[e^{te^X}] = \int_{\mathbb{R}} e^{te^x} \cdot \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

由于

$$\lim_{x\to\infty} exp(te^x - \frac{(x-\mu)^2}{2\sigma^2}) = \infty$$

那么该积分发散，即矩母函数不存在.

II 正态分布的矩母函数为

$$M_X(t) = exp\{\mu t + \frac{\sigma^2 t^2}{2}\}$$

那么

$$E(Y^n) = E(e^{nX}) = M_X(n) = exp\{\mu n + \frac{\sigma^2 n^2}{2}\}$$

Exercise 5.

设随机变量 X_1 和 X_2 相互独立，且均服从泊松分布 $P(\lambda_i)$，令 $Y = X_1 + X_2$. 利用矩母函数确定 Y 的分布.

Solution.

由于 X_1 和 X_2 相互独立，所以

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

即 $Y \sim P(\lambda_1 + \lambda_2)$.

Exercise 6.

将单位线段 $(0, 1)$ 随机断开，先扔掉左半段（包括左端点），再对剩下的线段随机断开后扔掉左半段，求余下那一段长度的期望值.

Solution.

记第一次断点为 $X \sim U(0, 1)$，第二次断点距该段左端点的距离为 $Y \sim U(0, 1-X)$
那么则余下部分的长度为

$$L = (1 - X) - Y$$

条件期望为

$$E[L \mid X = x] = (1 - x) - \frac{1-x}{2} = \frac{1-x}{2}$$

因此

$$E[L] = E\left[\frac{1-X}{2}\right] = \frac{1}{2}E[1 - X] = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

Exercise 7.

一个矿工在井下迷路，面前有三个门：从第一个门出，2 小时可达安全之处；从第二个门出，3 小时后会回到原地；从第三个门出，1 小时后会回到原地。假定该人随机选门且始终无法区分这三个门，求其到达安全之处的平均用时.

Solution.

设期望时间为 T，由条件期望意义容易看出

$$T = \frac{1}{3}\big[2 + (3 + T) + (1 + T)\big] = \frac{1}{3}(6 + 2T)$$

于是

$$T = 6 \text{（小时）}$$

Exercise 8.

证明：若 $E(Y|X) = X$，则 $Cov(X, Y) = Var(X)$.

Solution.

由全期望公式

$$E(Y) = E[E(Y|X)] = E(X)$$
$$E(XY) = E[XE(Y|X)] = E[X^2]$$

那么

$$Cov(X, Y) = E(XY) - E(X)E(Y) = E[X^2] - E(X)^2 = Var(X)$$

Exercise 9.

I 证明：若 X 与 Y 独立，则 $E[Y|X] = E[Y]$.

II 上述结论的反之是否成立？请说明理由.

Solution.

I 以连续性随机变量说明，其余同理，由条件期望的定义

$$E[Y|X] = \int_{\mathbb{R}} y f_{Y|X}(y|x) dy$$

由于 X 与 Y 独立，所以 $f_{Y|X}(y|x) = f_Y(y)$，那么

$$E[Y|X] = \int_{\mathbb{R}} y f_Y(y) dy = E[Y]$$

II 反之并不成立，例如

- $X \sim \mathrm{U}\{-1, 0, 1\}$
- Z 是与 X 独立的随机变量，且 $P(Z = 1) = P(Z = -1) = 0.5$
- 定义 $Y = X \cdot Z$

那么 $E[Y|X] = E[Y] = 0$，但 X 与 Y 并不独立.

Exercise 10.

令 $\hat{Y} = E[Y|X], \quad \tilde{Y} = Y - \hat{Y}$. 证明 $\mathrm{Var}(Y) = \mathrm{Var}(\hat{Y}) + \mathrm{Var}(\tilde{Y})$, 并给出直观解释.

Solution.

由全期望公式有

$$E[Y] = E[E[Y|X]] = E[\hat{Y}]$$

因此

$$\begin{aligned}
\mathrm{Var}(Y) &= E\Big[\Big(Y - E[Y]\Big)^2\Big] \\
&= E\Big[\Big(\hat{Y} - E[\hat{Y}] + \tilde{Y}\Big)^2\Big] \\
&= E\Big[\Big(\hat{Y} - E[\hat{Y}]\Big)^2\Big] + 2\,E\Big[\Big(\hat{Y} - E[\hat{Y}]\Big)\tilde{Y}\Big] + E\Big[\tilde{Y}^2\Big]
\end{aligned}$$

下面证明交叉项为零，注意到

$$E[\tilde{Y} \mid X] = E[Y - \hat{Y} \mid X] = E[Y|X] - \hat{Y} = 0$$

利用全期望公式，有

$$E\left[\left(\hat{Y} - E[\hat{Y}]\right)\tilde{Y}\right] = E\left[E\left[\left(\hat{Y} - E[\hat{Y}]\right)\tilde{Y} \,\middle|\, X\right]\right] = E\left[\left(\hat{Y} - E[\hat{Y}]\right) E[\tilde{Y} \mid X]\right] = 0$$

因此

$$\text{Var}(Y) = E\left[\left(\hat{Y} - E[\hat{Y}]\right)^2\right] + E\left[\tilde{Y}^2\right] = \text{Var}(\hat{Y}) + \text{Var}(\tilde{Y})$$

直观解释：该公式说明总体的波动性（方差）可以拆分为两部分：1. $\text{Var}(E[Y|X])$ 表示由 X 可解释的部分的波动性，即“可预测部分”的方差；2. $\text{Var}(Y - E[Y|X])$ 表示由于 X 之外的不确定性（即误差）的波动性，即“预测误差”的方差. 这正是“全方差定律”的一个重要等价形式，反映了总体不确定性来源于可预测性与随机误差两部分的贡献.

Exercise 11.

定义条件方差为

$$\text{Var}(Y|X) = E\left[(Y - E[Y|X])^2|X\right]$$

证明：

I $\text{Var}(Y|X) = \left[\text{E}[Y^2|X]\right] - \left(E^2[Y|X]\right)$.

II $\text{Var}(Y) = E\left[\text{Var}(Y|X)\right] + \text{Var}\left(E[Y|X]\right)$.

Solution.

I

$$\begin{aligned}
\text{Var}(Y|X) &= E\left[(Y - E[Y|X])^2|X\right] \\
&= E\left[Y^2 - 2YE[Y|X] + E[Y|X]^2|X\right] \\
&= E[Y^2|X] - 2E[Y|X]E[Y|X] + E[Y|X]^2 \\
&= E[Y^2|X] - E[Y|X]^2
\end{aligned}$$

II

$$E[Var(Y|X)] = E[E[(Y - E[Y|X])^2|X]] = E[E[Y^2|X] - E[Y|X]^2] = E[Y^2] - E[E[Y|X]^2]$$

$$Var(E[Y|X]) = E[E[Y|X]^2] - E[Y]^2$$

那么

$$Var(Y) = E[Y^2] - E[E[Y|X]^2] + E[E[Y|X]^2] - E[E[Y|X]]^2 = E[Var(Y|X)] + Var(E[Y|X])$$

Exercise 12.
点 (X,Y) 在单位圆盘上半部分均匀分布, 若观测到 $X=0.5$，则在均方误差意义下 Y 的最优预测值是？

Solution.
给定 $X=x$，(X,Y) 均匀分布半圆盘满足

$$y \in \left[0, \sqrt{1-x^2}\right]$$

因此

$$E[Y|X=x] = \frac{\sqrt{1-x^2}}{2}$$

取 $x=0.5$ 得

$$E[Y|X=0.5] = \frac{\sqrt{1-0.25}}{2} = \frac{\sqrt{3}}{4}$$

那么，由

$$E[(Y-g(X))^2] \geq E[(Y-E[Y|X])^2]$$

即知最优预测值为 $\dfrac{\sqrt{3}}{4}$

Exercise 13.
在实际应用中常基于 X 的观测对 Y 进行预测。设已知

$$E[X]=\mu_1,\quad E[Y]=\mu_2,\quad \mathrm{Var}(X)=\sigma_1^2,\quad \mathrm{Var}(Y)=\sigma_2^2,\quad \mathrm{Corr}(X,Y)=\rho.$$

(1) 在均方误差意义下求 Y 的最优线性预测 $\hat{Y}=a+bX$（确定 a 和 b）.
(2) 给出该最优线性预测对应的最小均方误差，并指出其何时趋近于 0.
(3) 验证：若 (X,Y) 服从双变量正态分布，则最优线性预测恰为条件期望 $E[Y|X]$.

Solution.
I 令目标函数

$$L(a,b) = E\left[(Y-a-bX)^2\right]$$

那么

$$\frac{\partial L}{\partial a} = -2E[Y-a-bX]$$

于是

$$E[Y]-a-bE[X]=0 \quad\implies\quad a=E[Y]-bE[X]=\mu_2-b\mu_1$$

将 $a=\mu_2-b\mu_1$ 代入目标函数，得

$$\begin{aligned}
L(b) &= E\left[\Big(Y-(\mu_2-b\mu_1)-bX\Big)^2\right] \\
&= E\left[\Big((Y-\mu_2)-b(X-\mu_1)\Big)^2\right] \\
&= E\left[(Y-\mu_2)^2-2b(Y-\mu_2)(X-\mu_1)+b^2(X-\mu_1)^2\right] \\
&= \mathrm{Var}(Y)-2b\,\mathrm{Cov}(X,Y)+b^2\,\mathrm{Var}(X)
\end{aligned}$$

故

$$\frac{dL(b)}{db} = -2\,\text{Cov}(X,Y) + 2b\,\text{Var}(X)$$

于是

$$-2\,\text{Cov}(X,Y) + 2b\,\text{Var}(X) = 0 \implies b\,\text{Var}(X) = \text{Cov}(X,Y)$$

因此

$$b = \rho\,\frac{\sigma_2}{\sigma_1}$$

II 由第一问

$$E\Big[(Y-\hat{Y})^2\Big] = \text{Var}(Y) - \frac{\text{Cov}(X,Y)^2}{\text{Var}(X)} = \sigma_2^2 - \rho^2\,\sigma_2^2 = \sigma_2^2(1-\rho^2)$$

当 $|\rho| \to 1$ 时，均方误差趋于 0，说明 X 能很好地预测 Y.

III 当 (X,Y) 服从二重正态分布时，条件分布 $Y|X$ 为正态分布，其条件期望为

$$E[Y|X] = \mu_2 + \frac{\text{Cov}(X,Y)}{\text{Var}(X)}(X-\mu_1)$$

这与我们求得的最优线性预测

$$\hat{Y} = a + bX = \mu_2 - b\mu_1 + bX$$

完全一致，即

$$\hat{Y} = E[Y|X]$$

因此在二重正态分布下，最优线性预测正好就是条件期望.

Exercise 14.

设 $\{X_i\}_{i\geq 1}$ 独立同分布且公共期望为 μ，令 $Y = X_1 + X_2 + \cdots + X_N$，其中 N 为取正整数值的随机变量，分布为 $P(N=n) = p_n$，且与 $\{X_i\}$ 相互独立.

I 假设 $E(N) = a$，求 $E[Y]$.

II 求 N 的矩母函数 $M_N(t)$.

III 若 $M_X(t)$ 为 X_i 的矩母函数，求 Y 的矩母函数 $M_Y(t)$.

IV 假设 $X_i \sim Exp(\lambda)$，且 N 服从几何分布，即 $P(N=n) = p(1-p)^{n-1}$，求 Y 的分布.

V 比较 (4) 中所得 Y 的分布与 $X_1 + X_2$ 的分布是否相同.

Solution.

I 由全期望公式

$$E[Y] = E[E[Y|N]] = E[N]E[X] = a\mu$$

II

$$M_N(t) = \sum_{n=1}^{\infty} p_n e^{tn}$$

III

$$M_Y(t) = E(M_X(t)^N) = M_N(\ln(M_X(t))) = \sum_{n=1}^{\infty} p_n M_X(t)^n$$

IV 取 $X_i \sim Exp(\lambda)$，故

$$M_X(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda$$

又设 N 服从几何分布 $P(N=n)=p(1-p)^{n-1}$，其矩母函数为

$$M_N(t)=\frac{pe^t}{1-(1-p)e^t},\quad e^t<(1-p)^{-1}$$

因此，

$$M_Y(t)=\frac{p\,M_X(t)}{1-(1-p)M_X(t)}=\frac{p\lambda/(\lambda-t)}{1-(1-p)\lambda/(\lambda-t)}=\frac{p\lambda}{p\lambda-t}$$

这正是参数为 $p\lambda$ 的指数分布的矩母函数，故 $Y\sim Exp(p\lambda)$.

V 由于 X_1+X_2 的分布为指数分布的卷积，得到的分布为伽马分布，即 $X_1+X_2\sim\Gamma(2,\lambda)$，而其矩母函数为 $\left(\dfrac{\lambda}{\lambda-t}\right)^2$ 显然与 $M_Y(t)=\dfrac{p\lambda}{p\lambda-t}$ 一般不同，故二者分布不相同.

Exercise 15.
构造一个随机个数的独立正态随机变量之和却不是正态分布的例子，并加以说明.

Solution.

- $X_i\sim N(0,1)$，$i=1,2,\dots,n$
- $N\sim G(p)$
- $Y=X_1+X_2+\cdots+X_N$

那么

$$M_Y(t)=E[M_X(t)^N]=M_N(\ln(M_X(t)))=\frac{p\,M_X(t)}{1-(1-p)M_X(t)}=\frac{pe^{t^2/2}}{1-(1-p)e^{t^2/2}}$$

显然不符合正态分布的矩母函数形式.

Exercise 16.
设 $\{Y_i\}_{i=1}^n$ 为独立同分布随机变量，其分布满足

$$P(Y_i=1)=P(Y_i=-1)=\frac{1}{2}.$$

令

$$X_n=\sum_{i=1}^n Y_i,$$

视 $Y_i=1$ 为股票价格上涨一元，$Y_i=-1$ 为股票价格下降一元，X_n 为第 n 个时间点的股票价格。

I 求 $E[X_n]$ 和 $\mathrm{Var}(X_n)$.

II 模拟 X_n 并绘出 X_n 对于 $n=1,\dots,10000$ 的图形，讨论随机序列的趋势及不同模拟结果的差别，并用第一问结论解释.

Solution.
由于每个 Y_i 均有

$$E[Y_i]=0,\quad \mathrm{Var}(Y_i)=1$$

故

$$E[X_n]=\sum_{i=1}^n E[Y_i]=0,\quad \mathrm{Var}(X_n)=\sum_{i=1}^n \mathrm{Var}(Y_i)=n.$$

模拟代码如下：

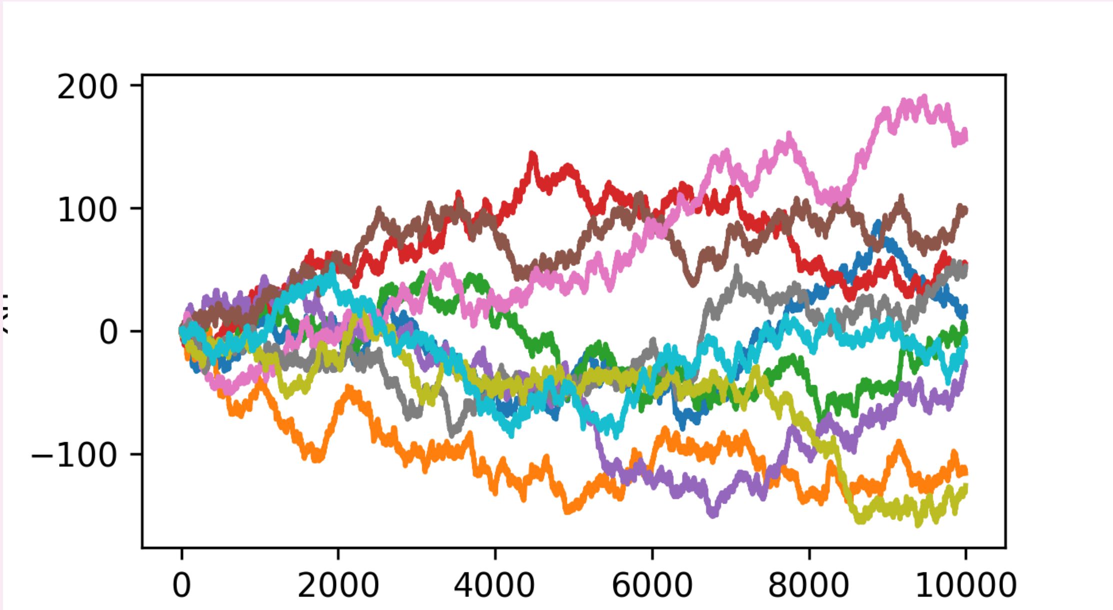
```python
from random import randint
from matplotlib import pyplot as plt

plt.figure(dpi=300)
plt.ylabel("Xn")
plt.xlabel("n")

def generate():
    y = [1 if randint(0, 1) else -1 for i in range(0, 10000)]
    x = [sum(y[0:ii + 1]) for ii in range(0, 10000)]
    plt.plot(range(0, 10000), x)

for i in range(0, 10):
    generate()

plt.show()
```



可以看出随着 n 的增大，X_n 的波动性也在增大（第一问的方差为 n 正说明了这个结果），且不同模拟结果的差别也在增大，随机序列并没有明显的趋势.