PROBABILITY AND STATISTICS I
HOMEWORK V

Jianyu Dong 2019511017

March, 29 2021

1

Since X_1 and X_2 are i.i.d. and each follows the uniform destribution on [0,1], we could get the p.d.f. of X_1 is

$$f_{X_1}(x_1) = 1,\ for\ 0 \leqslant x_1 \leqslant 1$$

$f_{X_1}(x_1) = 0$, elsewhere. The p.d.f. of X_2 is

$$f_{X_2}(x_2) = 1,\ for\ 0 \leqslant x_2 \leqslant 1$$

$f_{X_2}(x_2) = 0$, elsewhere. By the definition we get the p.d.f. of $Y = X_1 + X_2$ is

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X_1}(x) f_{X_2}(y-x)\,dx = \begin{cases} 0, & for\ y \leqslant 0, \\ y, & for\ 0 < y \leqslant 1, \\ 2-y, & for\ 1 < y \leqslant 2, \\ 0, & for\ y > 2. \end{cases}$$

2

Since $y_1 = x_1^2 + x_2^2,\ y_2 = x_2$, we could get $x_1 = \pm\sqrt{y_1 - y_2^2},\ x_2 = y_2$.
Let $\mathcal{S} = \{(x_1, x_2) : 0 < x_1^2 + x_2^2 < 1\}$, then the support for Y_1 and Y_2 is

$$\mathcal{T} = \{(y_1, y_2) : -1 < y_2 < 1, y_2^2 < y_1 < 1\}$$

Next by taking the partial derivatives, we could determine the Jacobian is

$$J = \det \begin{bmatrix} \pm\frac{1}{2}\frac{1}{\sqrt{y_1 - y_2^2}} & \mp\frac{y_2}{\sqrt{y_1 - y_2^2}} \\ 0 & 1 \end{bmatrix} = \pm\frac{1}{2\sqrt{y_1 - y_2^2}}.$$

Thus, by the theorem, the joint p.d.f. of $\mathbf{Y}$ is

$$f_{\mathbf{Y}}(y_1, y_2) = \frac{1}{\pi}\frac{1}{2\sqrt{y_1 - y_2^2}} = \frac{1}{2\pi\sqrt{y_1 - y_2^2}},$$

for $(y_1, y_2) \in \mathcal{T}$; otherwise $f_{\mathbf{Y}}(y_1, y_2) = 0$.

3

Since $y_1 = \frac{1}{2}(x_1 - x_2),\ y_2 = x_2$, we could get $x_1 = 2y_1 + y_2,\ x_2 = y_2$.
Let $\mathcal{S} = \{(x_1, x_2) : 0 < x_1 < \infty,\ 0 < x_2 < \infty\}$, then the support for Y_1 and Y_2 is

$$\mathcal{T} = \{(y_1, y_2) : -\frac{y_2}{2} < y_1 < \infty,\ 0 < y_2 < \infty\}$$

Next by taking the partial derivatives, we could determine the Jacobian is

$$J = \det \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} = 2$$

Thus, by the theorem, the joint p.d.f. of Y_1 and Y_2 is

$$f_{\mathbf{Y}}(y_1, y_2) = \frac{1}{2}e^{-(y_1+y_2)},$$

for $(y_1, y_2) \in \mathcal{T}$; otherwise $f_{Y_1,Y_2}(y_1, y_2) = 0$
The marginal p.d.f. of Y_1 is

$$f_{Y_1} = \int_{-\infty}^{+\infty} f_{Y_1,Y_2}(y_1, y_2)\,dy_2 = \begin{cases} \int_0^{+\infty} \frac{1}{2}e^{-(y_1+y_2)}\,dy_2 = \frac{1}{2}e^{-y_1},\ for\ y_1 > 0, \\ \int_{-2y_1}^{+\infty} \frac{1}{2}e^{-(y_1+y_2)}\,dy_2 = \frac{1}{2}e^{y_1},\ for\ y_1 \leqslant 0. \end{cases}$$

4

To determine the p.d.f. of Y, we could calculate the c.d.f. of Y.

Since $X_1, X_2, \ldots, X_n$ are random sample of n observations from the uniform distribution on the interval [0, 1], so the p.d.f. of X_i is $f_{X_i}(x_i) = 1$, for $x_i \in [0, 1]$; $f_{X_i}(x_i) = 0$ otherwise, for $i \in \{1, 2, ..., n\}$. The c.d.f. of Y is

$$\begin{aligned} F_Y(y) = P(at\ least\ n-1\ observations \leqslant y) &= \binom{n}{n-1} \left(\int_0^y 1\, dy\right)^{n-1} \int_y^1 1\, dy + \left(\int_0^y 1\, dy\right)^n \\ &= ny^{n-1}(1-y) + y^n, \end{aligned}$$

$for\ 0 \leqslant y \leqslant 1$.When y<0, $F_Y(y) = 0$, when y>1, $F_Y(y) = 1$. Then we could determine the p.d.f. of Y is

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \begin{cases} 0, & for\ y < 0, \\ n(n-1)y^{n-2}(1-y), & for\ 0 \leqslant y \leqslant 1, \\ 0, & for\ y > 1. \end{cases}$$

5

By the definition, we have the conditional p.d.f. of X given Y

$$g_1(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{3x^2}{y^3},\ for\ 0 < x < y$$

and $g_1(x \mid y) = 0$, otherwise. We also have the support of the marginal p.d.f. of Y is $(0, \infty)$, which means $\int_0^{+\infty} f_Y(y)\, dy = 1$. We also have the marginal p.d.f. of Y is $f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dx$

Since Z=X/Y and 0<x<y, we could get 0<z<1 and the c.d.f. of Z is

$$F_Z(z) = P\left(\frac{x}{y} \leqslant z\right) = \int_0^{+\infty} dy \int_0^{yz} f_{X,Y}(x,y)\, dx = z^3,\ for\ 0 < z < 1,$$

$F_Z(z) = 1,\ for\ z \geqslant 1,\ F_Z(z) = 0,\ for\ z \leqslant 0.$

The joint c.d.f. of Y and Z is

$$F_{Y,Z}(u,v) = P(y \leqslant u, z \leqslant v) = \int_0^u dy \int_0^{vy} f_{X,Y}(x,y)\, dx = v^3 \int_0^u f_Y(y)\, dy,\ for\ 0 < v < 1,\ u > 0.$$

Which means

$$F_{Y,Z}(y,z) = \begin{cases} F_Y(y), & for\ y > 0,\ z > 1, \\ z^3 F_Y(y), & for\ y > 0,\ 0 < z < 1, \\ 0, & elsewhere. \end{cases}$$

So we get $F_{Y,Z}(y,z) = F_Y(y)F_Z(z)$, which shows Z and Y are independent. The marginal p.d.f. of Z is $f_Z(z) = \frac{dF_Z(z)}{dz} = 3z^2,\ for\ 0 < z < 1$; $f_Z(z) = 0$,elsewhere.

6

6.1

By the definition, we have

$$P(X_1 < X_2 \mid X_1 < 2X_2) = \frac{P\left((X_1 < X_2) \cap (X_1 < 2X_2)\right)}{P(X_1 < 2X_2)} = \frac{P(X_1 < X_2)}{P(X_1 < 2X_2)}.$$

Since X_1, X_2, X_3 be i.i.d. with the p.d.f. $f(x) = e^{-x}$which is continuous, for 0<x<∞, zero elsewhere, we could get

$$P(X_1 < X_2) = \int_0^{x_2} e^{-x}\,dx = 1 - e^{-x_2},\ P(X_1 < 2X_2) = \int_0^{2x_2} e^{-x}\,dx = 1 - e^{-2x_2}$$

Thus we get

$$P(X_1 < X_2 \mid X_1 < 2X_2) = \frac{1 - e^{-x_2}}{1 - e^{-2x_2}} = \frac{1}{1 + e^{-x_2}},\ for\ x_2 > 0$$

and $P(X_1 < X_2 \mid X_1 < 2X_2) = 0$, elsewhere.

By the definition, we have

$$P(X_1 < X_2 < X_3 \mid X_3 < 1) = \frac{P\left((X_1 < X_2 < X_3) \cap (X_3 < 1)\right)}{P(X_3 < 1)}.$$

Since X_1, X_2, X_3 be i.i.d. with the p.d.f. $f(x) = e^{-x}$which is continuous, for 0<x<∞, zero elsewhere, we could get

$$\begin{aligned}P((X_1 < X_2 < X_3) \cap (X_3 < 1)) &= \int_0^1 dx_1 \int_{x_1}^1 dx_2 \int_{x_2}^1 e^{-(x_1+x_2+x_3)}\,dx_3 \\ &= \frac{1}{6}\left(1 - \frac{1}{e}\right)^3 \\ P(X_3 < 1) &= \int_0^1 e^{-x}\,dx = 1 - \frac{1}{e}.\end{aligned}$$

Thus, we get

$$P(X_1 < X_2 < X_3 \mid X_3 < 1) = \frac{1}{6}\left(1 - \frac{1}{e}\right)^2$$

6.2

Let $Y_1 = X_1/X_2,\ Y_2 = X_3/(X_1+X_2),\ Y_3 = X_1 + X_2$, we could determine that

$$X_1 = \frac{Y_1 Y_3}{1 + Y_1},\ X_2 = \frac{Y_3}{1 + Y_1},\ X_3 = Y_2 Y_3.$$

Let $\mathcal{S} = \{(x_1, x_2, x_3) : 0 < x_1, x_2, x_3 < \infty\}$, then the support for Y_1, Y_2, Y_3 is

$$\mathcal{T} = \{(y_1, y_2, y_3) : 0 < y_1 < \infty, 0 < y_2 < \infty, 0 < y_3 < \infty\}$$

Next by taking the partial derivatives, we could determine the Jacobian is

$$J = \det\begin{bmatrix} \frac{y_3}{(1+y_1)^2} & 0 & \frac{y_1}{1+y_1} \\ -\frac{y_3}{(1+y_1)^2} & 0 & \frac{1}{1+y_1} \\ 0 & y_3 & y_2 \end{bmatrix} = -\frac{y_3^2}{(1+y_1)^2}$$

By the definition, we have the joint p.d.f. of X_1, X_2, X_3 is $f_{\mathbf{X}}(\mathbf{x}) = e^{-(x_1+x_2+x_3)}$, for $0 < x_1 < \infty, 0 < x_2 < \infty, 0 < x_3 < \infty$, zero elsewhere. Thus the joint p.d.f. of Y_1, Y_2, Y_3 is

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{y_3^2}{(1+y_1)^2} e^{-(1+y_2)y_3},\ for\ 0 < y_1 < \infty, 0 < y_2 < \infty, 0 < y_3 < \infty,$$

zero elsewhere.

The c.d.f. of Y_1 is

$$F_{Y_1}(y_1) = P\left(\frac{x_1}{x_2} \leqslant y_1\right) = \int_0^{+\infty} dx_2 \int_0^{x_2 y_1} e^{-(x_1+x_2)}\,dx_1 = \frac{y_1}{1+y_1},\ for\ y_1 > 0,$$

zero elsewhere.

Then we could calculate the p.d.f. of Y_1 is

$$f_{Y_1}(y_1) = \frac{dF_{Y_1}(y_1)}{dy_1} = \frac{1}{(1+y_1)^2},\ for\ y_1 > 0,$$

zero elsewhere.

The c.d.f. of Y_2 is

$$\begin{aligned} F_{Y_2}(y_2) = P\left(\frac{x_3}{x_1+x_2} \leqslant y_2\right) &= \int_0^{+\infty} dx_1 \int_0^{+\infty} dx_2 \int_0^{y_2(x_1+x_2)} e^{-(x_1+x_2+x_3)}\, dx_3 \\ &= 1 - \frac{1}{(1+y_2)^2},\ for\ y_2 > 0, \end{aligned}$$

zero elsewhere.

Then we could calculate the p.d.f. of Y_2 is

$$f_{Y_2}(y_2) = \frac{dF_{Y_2}(y_2)}{dy_2} = \frac{2}{(1+y_2)^3},\ for\ y_2 > 0,$$

zero elsewhere.

The c.d.f. of Y_3 is

$$F_{Y_3}(y_3) = P\left(x_1 + x_2 < y_3\right) = \int_0^{y_3} dx_2 \int_0^{y_3-x_2} e^{-(x_1+x_3)}\, dx_1 = 1 - (y_3+1)e^{-y_3},\ for\ y_3 > 0,$$

zero elsewhere.

Then we could calculate the p.d.f. of Y_3 is

$$f_{Y_3}(y_3) = \frac{dF_{Y_3}(y_3)}{dy_3} = y_3 e^{-y_3},\ for\ y_3 > 0,$$

zero elsewhere.

It is clear that

$$f_{\mathbf{X}}(\mathbf{x}) \neq F_{Y_1}(y_1)F_{Y_2}(y_2)F_{Y_3}(y_3),$$

so Y_1, Y_2, Y_3 are not independent.

7

7.1

Let A=The student is late on a certain day, B=The student will be on time on each of the next three days. By the definition of Markov chain, we have

$$P(B \mid A) = 0.8 \times (1-0.7)^2 = 0.072.$$

So if the student is late on a certain day, the probability that he will be on time on each of the next three days is 0.072.

7.2

Let C=The student is on time on a certain day, D=The student will be late on each of the next three days. By the definition of Markov chain, we have

$$P(D \mid C) = 0.7 \times (1-0.8)^2 = 0.028$$

So if the student is on time on a certain day, the probability that he will be late on each of the next three days is 0.028.

7.3

Let the probability that the student will be on time on the fourth day of class is p_4 and the probability that the student will be on time on the third day of class is p_3, so the probability that the student will be late on the third day is (1-p_3). By the definition of Markov chain, we have

$$p_4 = 0.3p_3 + 0.8(1 - p_3) = 0.8 - 0.5p_3,$$

and so on, we have $p_3 = 0.8 - 0.5p_2,\ p_2 = 0.8 - 0.5p_1$. We have that $p_1 = 1$, so we get $p_4 = 0.475$. So the probability that the student will be on time on the fourth day of class is 0.475.

7.4

Let the probability that the student will be on time on the day n of class is p_n and the probability that the student will be on time on the day (n-1) of class is p_{n-1}, so the probability that the student will be late on the day (n-1) is (1-p_{n-1}). By the definition of Markov chain, we have

$$p_n = 0.8(1 - p_{n-1}) + (1 - 0.7)p_{n-1} = 0.8 - 0.5p_{n-1}.$$

So we could determine that

$$p_n = 0.8\sum_{k=0}^{n-2}\left(-\frac{1}{2}\right)^k + \left(-\frac{1}{2}\right)^{n-1} p_1 = \frac{8}{15}\left[1 - \left(-\frac{1}{2}\right)^{n-1}\right] + \left(-\frac{1}{2}\right)^{n-1} p_1$$

Let the probability that he is on time on the first day of the class is $p_1 = p$, we could determine the probability that he will be on time on the seventh day of the class is $p_7 = 0.525 + 0.0156p$.

8

Since all tosses are independent, with following exception: Whenever either three heads or three tails have been obtained on three successive tosses, then the outcome of the next toss is always of the opposite type. Which shows the truth that the conditional distribution of all X_{n+j} for $j \geqslant 1$ given $X_1, \ldots, X_n$ depend only on X_n and ont on the earlier states $X_1, \ldots, X_{n-1}$. Then this process is a Markov chain.

Let $X_n = 1$ be {hhh}, $X_n = 2$ be {hht}, $X_n = 3$ be {hth}, $X_n = 4$ be {thh}, $X_n = 5$ be {htt}, $X_n = 6$ be {tht}, $X_n = 7$ be {tth}, $X_n = 8$ be {ttt}. (t means tail, h means head)

Thus we could easily get the transition matrix is

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

9

By the definition, we have the probability mass function of the geometric distribution with parameter p is

$$p_X(x) = P(X = x) = (1 - p)^{x-1}p,\ for\ x = 1, 2, \ldots.$$

Thus, the expectation is

$$E(X) = \sum_{x=1}^{\infty} x p_X(x) = p\sum_{x=1}^{\infty} x(1 - p)^{x-1} = \frac{1}{p}.$$

10

Since X has a Bernoulli distribution with parameter p, the probability mass function is

$$p_X(1) = P(X = 1) = p,\ p_X(0) = P(X = 0) = 1 - p.$$

Let $Y = 2^X$, so the support for Y is 1,2, and the probability mass function is

$$p_Y(1) = p_X(0) = 1 - p,\ p_Y(2) = p_X(1) = p.$$

So the expectation is

$$E(Y) = E(2^X) = 1(1 - p) + 2p = 1 + p$$

11

Let X be a discrete finite random variable, and the p.m.f. of X is $a_i = p_X(x_i) = P(X = x_i),\ for\ i \in 1, 2, \ldots, n$. So $\sum_{i=1}^{n} a_i = 1$.

Then we could determine the expectation of X is

$$E(X) = \sum_{i=1}^{n} x_i a_i.$$

The expectation of $\varphi(x)$ is

$$E(\varphi(x)) = \sum_{i=1}^{n} \varphi(x_i) a_i$$

Thus, we only need to prove the inequality

$$\varphi\left(\sum_{i=1}^{n} a_i x_i\right) \leqslant \sum_{i=1}^{n} a_i \varphi(x_i)$$

Since $\varphi(x)$ is a convex function, by the definition, we could get

$$\varphi\left(\sum_{i=1}^{n} a_i x_i\right) = \varphi\left((1 - a_n)\left(\sum_{i=1}^{n-1} \frac{a_i}{1 - a_n} x_i\right) + a_n x_n\right) \leqslant (1 - a_n)\varphi\left(\sum_{i=1}^{n-1} \frac{a_i}{1 - a_n} x_i\right) + a_n \varphi(x_n)$$

And so on, using $\varphi(kx_i) \leqslant k\varphi(x_i)$, we could get

$$\varphi\left(\sum_{i=1}^{n} a_i x_i\right) \leqslant \sum_{i=1}^{n} a_i \varphi(x_i).$$

Which means

$$\varphi\left(E(X)\right) \leqslant E\left(\varphi(X)\right)$$

Let $\varphi(x) = x^2$, a convex function, we could easily get that

$$E(X)^2 \leqslant E(X^2)$$

12

12.1

Since X_1, X_2, X_3 are three independent random variables following the geometric distribution with parameter p, we could determine the joint p.m.f. of X_1, X_2, X_3 is

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = \frac{p^3}{(1 - p)^3}(1 - p)^{x_1 + x_2 + x_3},\ for\ x_1, x_2, x_3 \in N^*.$$

Then the p.m.f. of $Z = X_1 + X_2 + X_3$ is

$$f_Z(k) = P(x_1 + x_2 + x_3 = k) = p^3(1-p)^{k-3}\sum_{i=1}^{k-2} i = \frac{(k-1)(k-2)}{2}p^3(1-p)^{k-3}$$

12.2

By the definition, we have

$$E(X_1^2) = \sum_{x_1=1}^{\infty} p x_1^2 (1-p)^{x_1-1} = \sum_{x_1=3}^{\infty} p(x_1-2)^2(1-p)^{x_1-3}$$

$$E(X_1) = \sum_{x_1=1}^{\infty} p x_1 (1-p)^{x_1-1} = \sum_{x_1=3}^{\infty} p(x_1-2)(1-p)^{x_1-3}$$

So we get that

$$E(X_1^2) + E(X_1) = \sum_{x_1=3}^{\infty} p(x_1-1)(x_1-2)(1-p)^{x_1-3}$$

Using the fact that

$$1 = \sum_{z=3}^{\infty} f_Z(z) = \frac{p^2}{2}\sum_{z=3}^{\infty} p(z-1)(z-2)(1-p)^{z-3} = \frac{p^2}{2}\left(E(X_1^2) + E(X_1)\right),$$

we get

$$p^2\left(E(X_1^2) + E(X_1)\right) = 2$$

12.3

Since we have

$$p^2(E(X^2) + E(X)) = 2, \quad E(X) = \frac{1}{p},$$

we could easily get that

$$E(X^2) = \frac{2}{p^2} - \frac{1}{p} = \frac{2-p}{p^2} \textit{ and } E(X^2) - E(X)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$