Probability & Statistics for EECS: Homework #10

Due on April 23, 2023 at 23:59

Name: **Zhou Shouchen**
Student ID: 2021533042

Problem 1

(a) X discrete, Y discrete:

According to the defination of the marginal PMF, we can get that $P(X = x) = \sum\limits_{y} P(X = x, Y = y)$.

And from the defination of the conditional PMF, we can get that $P(X = x|Y = y) = \frac{P(X=x,Y=y)}{P(Y=y)}$.

So $P(X = x, Y = y) = P(X = x|Y = y)P(Y = y)$.

Combine them, we can get that $P(X = x) = \sum\limits_{y} P(X = x|Y = y)P(Y = y)$.

So above all, when X discrete, Y discrete, $P(X = x) = \sum\limits_{y} P(X = x|Y = y)P(Y = y)$.

(b) X discrete, Y continuous:

Let us prove that inversely.

$P(X = x|Y = y) = \lim\limits_{\epsilon \to 0} P(X = x|Y \in (y - \epsilon, y + \epsilon))$.

From the Bayes' Rule, we can get that $P(X = x|Y \in (y-\epsilon, y+\epsilon)) = \dfrac{P(Y \in (y - \epsilon, y + \epsilon)|X = x)P(X = x)}{Y \in (y - \epsilon, y + \epsilon)}$.

So we can get that $P(X = x|Y = y) = \lim\limits_{\epsilon \to 0} \dfrac{P(Y \in (y - \epsilon, y + \epsilon)|X = x)P(X = x)}{Y \in (y - \epsilon, y + \epsilon)}$

$= \lim\limits_{\epsilon \to 0} \dfrac{\dfrac{P(Y \in (y - \epsilon, y + \epsilon)|X = x)}{2\epsilon}P(X = x)}{\dfrac{Y \in (y - \epsilon, y + \epsilon)}{2\epsilon}} = \dfrac{f_Y(y|X = x)}{f_Y(y)}P(X = x)$.

So $P(X = x|Y = y)f_Y(y) = f_Y(y|X = x)P(X = x)$.

When can integrating y on both sides simultaneously, i.e.

$$\int_{-\infty}^{+\infty} P(X = x|Y = y)f_Y(y)dy = \int_{-\infty}^{+\infty} f_Y(y|X = x)P(X = x)dy$$

And for the right-hand side, we can get that

$$\int_{-\infty}^{+\infty} f_Y(y|X = x)P(X = x)dy = P(X = x)\int_{-\infty}^{+\infty} f_Y(y|X = x)dy$$

And since conditional PDF is also a valid PDF, so $\int_{-\infty}^{+\infty} f_Y(y|X = x)dy = 1$.

So we can get that

$$\int_{-\infty}^{+\infty} P(X = x|Y = y)f_Y(y)dy = P(X = x)$$

i.e.

$$P(X = x) = \int_{-\infty}^{+\infty} P(X = x|Y = y)f_Y(y)dy$$

So above all, when X discrete, Y continuous, $P(X = x) = \int_{-\infty}^{+\infty} P(X = x|Y = y)f_Y(y)dy$.

(c) X continuous, Y discrete:

Since X is continuous, so the PDF of X is $f_X(x) = \lim\limits_{\epsilon \to 0} \dfrac{P(X \in (x - \epsilon, x + \epsilon))}{2\epsilon}$.

And from (a) we can get that $P(X \in (x - \epsilon, x + \epsilon)) = \sum\limits_{y} P(X \in (x - \epsilon, x + \epsilon)|Y = y)P(Y = y)$.

Combine them, we can get that

$f_X(x) = \lim\limits_{\epsilon \to 0} \dfrac{\sum\limits_{y} P(X \in (x - \epsilon, x + \epsilon)|Y = y)P(Y = y)}{2\epsilon} = \lim\limits_{\epsilon \to 0} \dfrac{\sum\limits_{y} P(X \in (x - \epsilon, x + \epsilon)|Y = y)}{2\epsilon}P(Y = y) = \sum\limits_{y} f_X(x|Y = y)P(Y = y)$.

2

So above all, when X continuous, Y discrete, $f_X(x) = \sum_y f_X(x|Y=y)P(Y=y)$.

(d) X continuous, Y continuous:

From the defination of marginal PDF, we can get the $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$.

And from the conditional PDF, we can get that $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$.

i.e. $f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y)$.

And combine them we can get that $f_X(x) = \int_{-\infty}^{+\infty} f_{X|Y}(x|y)f_Y(y)dy$.

So above all, when X continuous, Y continuous, $f_X(x) = \int_{-\infty}^{+\infty} f_{X|Y}(x|y)f_Y(y)dy$.

Problem 2

Because bus from Blissville to the city arrives every 15 minutes, it have periodicity.

And bus from Blotchville to the next is $Expo(\frac{1}{15})$, it have memoryless property.

So we only need to consider the time range between two buses's interval of Blissville, i.e. during the 15 minutes' interval.

And we can regard that the Fred arrives at the bus stop at time 0.

Since Fred arrives at a uniformly random time, so we can regard that the bus of Blissville arrives at a uniformly random time.

Let $U \sim Unif(0, 15)$, and $X \sim Expo(\frac{1}{15})$.

So we can use U to be the time that bus from Blissville arrives, and X be the time that the bus from Blotchville arrives.

All time units are in minutes.

(a) Since $X \sim Expo(\frac{1}{15})$, so $F_X(x) = 1 - e^{-\frac{1}{15}x}, x > 0$. And $U \sim Unif(0, 15)$, so $f_U(u) = \frac{1}{15}$.

If the Blotchville company bus arrives first, which means that $X < U$.

With LOTP, we can get that

$P(X < U) = \int_0^{15} P(X < U|U = u)f_U(u)du = \int_0^{15} \frac{1}{15}P(X < u)du$

$= \frac{1}{15}\int_0^{15}(1 - e^{-\frac{1}{15}u})du = \frac{1}{15}(u + 15e^{-\frac{1}{15}u})|_0^{15} = \frac{1}{15}(15 + 15e^{-1} - 15) = \frac{1}{e}$.

So above all, the probability that the Blotchville company bus arrives first is $\frac{1}{e}$.

(b) Let T be Fred's waiting time, then $T = min(U, X)$.

The generation function of T is that $G_T(t) = P(T > t) = P(min(U, X) > t) = P(U > t, X > t)$.

Since the arrival of two bus are independent, i.e. U and X are independent.

The CDF of X is $F_X(x) = 1 - e^{-\frac{1}{15}x}$, so $P(X > t) = 1 - (1 - e^{-\frac{1}{15}t}) = e^{-\frac{1}{15}t}$.

The CDF of U is $F_U(u) = \frac{u}{15}$, so $P(U > t) = 1 - \frac{t}{15}$.

So $P(U > t, X > t) = P(U > t)P(X > t) = e^{-\frac{1}{15}t}(1 - \frac{t}{15})$.

So the CDF of T is that $F_T(t) = 1 - G_T(t) = 1 - e^{-\frac{1}{15}t}(1 - \frac{t}{15}), t \in (0, 15)$.

And when $t \leq 0$, $F_T(t) = 0$. When $t \geq 15$, $F_T(t) = 1$.

So above all, the CDF of Fred's waiting time for a bus is that $F_T(t) = \begin{cases} 0, & t \leq 0 \\ 1 - e^{-\frac{1}{15}t}(1 - \frac{t}{15}), & 0 < t < 15 \\ 1, & t \geq 15 \end{cases}$

Problem 3

Since $N \sim Pois(\lambda)$, so $P(N = n) = \dfrac{e^{-\lambda}\lambda^n}{n!}$.
And let $q = 1 - p$.

(a)The joint PMF of N, X, Y is that
$P(N = n, X = x, Y = y) = P(X = x, Y = y|N = n)P(N = n)$
If $x + y = n$, then $P(X = x, Y = y|N = n)P(N = n) = \dfrac{e^{-\lambda}\lambda^n}{n!} \cdot \binom{n}{x} \cdot p^x \cdot q^{n-x}$.
If $x + y \neq n$, then $P(X = x, Y = y|N = n)P(N = n) = 0$.

Since for the support of the joint PMF, $x + y = n$ must be satisfied, so the domain of X, Y, N are couple, which means that X, Y, N are not independent.

So above all, the joint PMF of N, X, Y is that
$$P(N = n, X = x, Y = y) = \begin{cases} 0, & x + y \neq n \\ \binom{n}{x} \cdot p^x \cdot q^{n-x} \cdot \dfrac{e^{-\lambda}\lambda^n}{n!}, & x + y = n \end{cases}$$
Where $q = 1 - p$.
And X, Y, N are not independent.

(b) Similary with (a), we can get that the PMF of N, X is that
When $x \leq n$, $P(N = n, X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x} \cdot \dfrac{e^{-\lambda}\lambda^n}{n!} = e^{-\lambda}\lambda^n q^n \cdot \dfrac{p^x}{q^x \cdot x!} \cdot \dfrac{1}{(n-x)!}$
When $x > n$, $P(N = n, X = x) = 0$

Firstly, we can not seperate $(n - x)!$ into the functions of n and x, secondly, for the support of the joint PMF, we must make sure that $x \leq n$, which means that the domain couple, so X, N are not independent.

So above all, $P(N = n, X = x) = \begin{cases} 0, & x > n \\ \binom{n}{x} \cdot p^x \cdot q^{n-x} \cdot \dfrac{e^{-\lambda}\lambda^n}{n!}, & x \leq n \end{cases}$
Where $q = 1 - p$.
And X, N are not independent.

(c) Similarly with (b), we can get that the PMF of X, Y is that
$P(X = x, Y = y) = P(X = x, N = x + y) = \dfrac{e^{-\lambda}\lambda^n}{n!} \cdot \binom{n}{x} \cdot p^x \cdot q^{n-x}$
$= \dfrac{e^{-\lambda}\lambda^{x+y}}{(x+y)!} \cdot \dfrac{(x+y)!}{x!y!}p^x q^y = e^{-\lambda} \cdot \dfrac{\lambda^x p^x}{x!} \cdot \dfrac{\lambda^y p^y}{y!}$.
Since $p + q = 1$, so $e^{-\lambda} = e^{-\lambda(p+q)} = e^{-\lambda p}e^{-\lambda q}$.
So $P(X = x, Y = y) = \dfrac{e^{-\lambda p}(\lambda p)^x}{x!} \cdot \dfrac{e^{-\lambda q}(\lambda q)^y}{y!}$.

So we can write the joint PMF into $g(x) = \dfrac{e^{-\lambda p}(\lambda p)^x}{x!}$, and $h(y) = \dfrac{e^{-\lambda q}(\lambda q)^y}{y!}$, and $g(x), h(y)$ domain have no couple.
So we can say that they are independent,
and X, Y with PMF $P_X(x) = \dfrac{e^{-\lambda p}(\lambda p)^x}{x!}, P_Y(y) = \dfrac{e^{-\lambda q}(\lambda q)^y}{y!}$.

So above all, the joint PMF of X, Y is $P(X, Y) = g(x)h(y) = \dfrac{e^{-\lambda p}(\lambda p)^x}{x!} \cdot \dfrac{e^{-\lambda q}(\lambda q)^y}{y!}$,

5

where $q = 1 - p$.
And X, Y are independent.

(d) From (c), we can get that the PDF of X is $P_X(x) = g(X) = \dfrac{e^{-\lambda p}(\lambda p)^x}{x!}$.

And the PDF of Y is $P_Y(y) = h(Y) = \dfrac{e^{-\lambda q}(\lambda q)^y}{y!}$.

So $X \sim Pois(\lambda p)$, and $Y \sim Pois(\lambda q)$.

So $Var(X) = \lambda p$, and $Var(Y) = \lambda q$.

Similarly, since $N \sim Pois(\lambda)$, so $Var(N) = \lambda$.

And the correlation between N and X is that

$Corr(N, X) = \dfrac{Cov(N, X)}{\sqrt{Var(N)Var(X)}}$.

Since $N = X + Y$, so $Cov(N, X) = Cov(X + Y, X) = Cov(X, X) + Cov(X, Y)$.

From the property of covariance, we can get that $Cov(X, X) = Var(X)$.

And from (c), we can get that X, Y are independent, so $Cov(X, Y) = 0$.

So $Cov(N, X) = Cov(X + Y, X) = Cov(X, X) + Cov(X, Y) = Var(X) + 0 = Var(X)$

So $Corr(N, X) = \dfrac{Cov(N, X)}{\sqrt{Var(N)Var(X)}} = \dfrac{Var(X)}{\sqrt{Var(N)Var(X)}} = \dfrac{\lambda p}{\sqrt{\lambda \cdot \lambda p}} = \sqrt{p}$.

So above all, the correlation between N and X is $\sqrt{p}$.

Problem 4

Let X, Y be the i.i.d standard Normal random variables.

i.e. $X \sim N(0,1)$, $Y \sim N(0,1)$.

And let L be the larger one, i.e. $L = max(X, Y)$.

Let S be the smaller one, i.e. $S = min(X, Y)$.

So the correlation between the larger and smaller one is that

$Corr(L, S) = \dfrac{Cov(L, S)}{\sqrt{Var(L)Var(S)}}.$

1. For $Cov(L, S)$:

From the defination of covariance, we can get that $Cov(L, S) = E(LS) - E(L)E(S)$.

Also, we could discover that $L \cdot S = X \cdot Y$, so $E(LS) = E(XY)$.

Since $X, Y \sim N(0,1)$,so $E(X) = E(Y) = 0$,

also X, Y are independent, so $E(XY) = E(X)E(Y) = 0$.

And we can also discover that $L + S = X + Y$, $L - S = |X - Y|$,

so $L = \dfrac{1}{2}(X + Y + |X - Y|)$, $S = \dfrac{1}{2}(X + Y - |X - Y|)$.

So $E(L) = E(\dfrac{1}{2}(X + Y + |X - Y|)) = \dfrac{1}{2}(E(X) + E(Y) + E(|X + Y|)) = \dfrac{1}{2}E(|X - Y|)$.

Similary, $E(S) = -\dfrac{1}{2}E(|X - Y|)$.

So we just need to find out the value of $E(|X - Y|)$.

Since $X, Y \sim N(0,1)$, so $-Y \sim N(0,1)$, so $X - Y \sim N(0,2)$.

Let $Z \sim N(0,1)$, so $X - Y \sim \sqrt{2}Z$.

So from LOTUS, we can get that

$E(|Z|) = \int_{-\infty}^{+\infty} |z| \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 2 \int_{0}^{+\infty} z \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$

$= \dfrac{1}{\sqrt{2\pi}} \int_{0}^{+\infty} e^{-\frac{1}{2}z^2} dz^2 = \dfrac{-2}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \Big|_{z^2=0}^{+\infty} = \dfrac{-2}{\sqrt{2\pi}}(0 - 1) = \sqrt{\dfrac{2}{\pi}}.$

So $E(|X - Y|) = E(\sqrt{2}|Z|) = \sqrt{2}E(|Z|) = \sqrt{2}\sqrt{\dfrac{2}{\pi}} = \dfrac{2}{\sqrt{\pi}}.$

So we can get that $E(L) = \dfrac{1}{2}E(|X - Y|) = \dfrac{1}{2} \cdot \dfrac{2}{\sqrt{\pi}} = \dfrac{1}{\sqrt{\pi}}.$

And $E(S) = -\dfrac{1}{2}E(|X - Y|) = -\dfrac{1}{2} \cdot \dfrac{2}{\sqrt{\pi}} = -\dfrac{1}{\sqrt{\pi}}.$

So combine all we get above, we can get that $Cov(L, S) = E(LS) - E(L)E(S) = 0 - \dfrac{1}{\sqrt{\pi}} \cdot (-\dfrac{1}{\sqrt{\pi}}) = \dfrac{1}{\pi}.$

2. for $Var(L)$:

Since $X, Y \sim N(0,1)$, so with the symmetry of Normal distribution, we can get that $-X, -Y \sim N(0,1)$.

And since $L = max(X, Y)$, and $S = -max(-X, -Y)$.

So with symmetry, we could say that $Var(L) = Var(S)$.

And we can find that $L + S = X + Y$, so $Var(L + S) = Var(X + Y)$.

Since X, Y are independent, so $Var(X + Y) = Var(X) + Var(Y) = 2$.

And from the property of variance, we can get that $Var(L + S) = Var(L) + Var(S) + 2Cov(L, S)$.

So combine them we can get that $Var(L) + Var(S) + 2Cov(L, S) = 2$.

From above, we have get that $Var(L) = Var(S)$, and $Cov(L, S) = \dfrac{1}{\pi}$.

7

So we can get that $Var(L) = Var(S) = 1 - \dfrac{1}{\pi}$.

Combine 1. and 2., we can get that

$$Corr(L, S) = \frac{Cov(L, S)}{\sqrt{Var(L)Var(S)}} = \frac{\dfrac{1}{\pi}}{\sqrt{(1 - \dfrac{1}{\pi})^2}} = \frac{1}{\pi - 1}.$$

So above all, the correlation between the larger and smaller of the values is $\dfrac{1}{\pi - 1}$.

Problem 5

(a) From the defination of expectation, we can get that $E(X) = \sum\limits_{i=1}^{n} \dfrac{1}{n} x_i = \bar{x}$,

similarly, $E(Y) = \bar{y}$.

From the defination of covariance, we can get that

$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E[(X - \bar{x})(Y - \bar{y})]$.

Since (X, Y) are chosen uniformly at random, so the expectation can be calculate through defination,

i.e. $E[(X - \bar{x})(Y - \bar{y})] = \sum\limits_{i=1}^{n} \dfrac{1}{n}(x_i - \bar{x})(y_i - \bar{y}) = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = r$.

So $Cov(X, Y) = r$.

So above all, the $Cov(X, Y)$ is equal to the sample correlation coefficient r.

(b)

1. Since all x_i are distinct, and all y_i are distinct,

so for the signed area contributed by (x_i, y_i), and (x_j, y_j) is that $S_{i,j} = (x_i - x_j) \cdot (y_i - y_j)$.

And for a pair of points, we only need to calculate one time for the signed area they contributed to avoid repeat.

So the total signed area is that $S = \sum\limits_{i<j}(x_i - x_j)(y_i - y_j)$

As for $E[(X - \tilde{X})(Y - \tilde{Y})]$, consider it in two ways.

For the first way, we can consider it as the average signed area, their are total n^2 uniformly random choices for choosing the points (X, Y) and $(\tilde{X}, \tilde{Y})$.

If we pick two same points, then the signed area they contributed is 0. And their are total n situations.

As for the other $\binom{n}{2}$ situations, they must not be the same point, but each pair are calculate twice as $(X, Y) = (x_i, y_i), (\tilde{X}, \tilde{Y}) = (x_j, y_j)$ and opposite.

So we can get the expectation with defination:

$E[(X - \tilde{X})(Y - \tilde{Y})] = \dfrac{1}{n^2}[0 \cdot n + 2 \cdot \sum\limits_{i<j}(x_i - x_j)(y_i - y_j)] = \dfrac{2S}{n^2}$.

So we can get that $S = \dfrac{n^2}{2} E[(X - \tilde{X})(Y - \tilde{Y})]$.

i.e. the total signed area of the rectangles is the constant times $E[(X - \tilde{X})(Y - \tilde{Y})]$, and the constant is that $\dfrac{n^2}{2}$.

2. For the second way, with property of expectation, we can get that

$E[(X - \tilde{X})(Y - \tilde{Y})] = E(XY) - E(X\tilde{Y}) - E(\tilde{X}Y) + E(\tilde{X}\tilde{Y})$.

Since (X, Y) is independent with $(\tilde{X}, \tilde{Y})$, so $E(X\tilde{Y}) = E(X)E(\tilde{Y})$, and $E(\tilde{X}Y) = E(\tilde{X})E(Y)$.

Also, since (X, Y) have same distribution with $(\tilde{X}, \tilde{Y})$, so $X \sim \tilde{X}$, and $Y \sim \tilde{Y}$.

So $E(\tilde{X}) = E(X)$, $E(\tilde{Y}) = E(Y)$, $E(\tilde{X}\tilde{Y}) = E(XY)$.

With these, we can get that

$E[(X - \tilde{X})(Y - \tilde{Y})] = E(XY) - E(X)E(Y) - E(X)E(Y) + E(XY)$

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

Since the two parts are equal, so we can get that $2Cov(X, Y) = \dfrac{2S}{n^2}$, i.e. $Cov(X, Y) = \dfrac{S}{n^2}$.

And from (a), we have get that $Cov(X, Y) = r$, where r is the sample covariance of the data.

So $r = \dfrac{S}{n^2}$.

So we can get that the sample covariance of the data is the constant times the total signed area of the

---

9

rectangles, and the constant is that $\dfrac{1}{n^2}$.

So above all, the total signed area of the rectangles is the constant times $E[(X - \tilde{X})(Y - \tilde{Y})]$, and the constant is that $\dfrac{n^2}{2}$.

And the sample covariance of the data is the constant times the total signed area of the rectangles, and the constant is that $\dfrac{1}{n^2}$.

(c)Since the covariance is a constant$(\dfrac{1}{n^2})$ times the total signed area of the rectangles,

So we can use the total signed area to analyze the covariance.

And we can regard the first dimension as the x-axis, and the second dimension as the y-axis.

(i) $Cov(W_1, W_2) = Cov(W_2, W_1)$

If we swap the order of the x-axis and the y-axis, their is no effect on the signed area of the rectangles.

So if we swap the order of the two random variables, their is no effect on their covariance.

(ii) $Cov(a_1 W_1, a_2 W_2) = a_1 a_2 Cov(W_1, W_2)$

If a_1 or a_2 is negative, it means that it shrinking $a_i (i = 1, 2)$ times on its axis.

If a_1 or a_2 is positive, it means that it stretching $a_i (i = 1, 2)$ times on its axis.

So if we multiply a constant on the x-axis or the y-axis, it will have a shrink or stretch effect on its axis.

And since the signed area of the rectangles is the multiply of the two axis, so it will have a shrink or stretch effect on the signed area of the rectangles with the coefficient of $a_1 a_2$.

So the will have an effect on the covariance with the coefficient of $a_1 a_2$.

(iii) $Cov(W_1 + a_1, W_2 + a_2) = Cov(W_1, W_2)$

This means that the all points are shifting along the x-axis for the distance of a_1, and shifting along the y-axis for the distance of a_2.

Since shifting has no effect on the signed area of the rectangles, so it has no effect on the covariance.

(iv) $Cov(W_1, W_2 + W_3) = Cov(W_1, W_2) + Cov(W_1, W_3)$

We can split one rectangle into two rectangles, the newly two rectangles have the same length on the x-axis with the original rectangle,

And the sum of their length on the y-axis is the length of the original rectangle.

And we need to regulate that if the signed area of the original rectangle is negative, then the two new rectangles are also with negative areas.

If the signed area of the original rectangle is positive, then the two new rectangles are also with positive areas.

With this regulation, we can say that the signed area of the origin is the same as the sum of the two new rectangles. This because the new rectangles are split from the origin one.

So we can say that the covariance of the origin is the same as the sum of the covariance of the two new ones.