A2-DS203

September 2020

Q1

(2.5 marks)

Let $Z = \min(X, Y)$. So, for $a > 0$,

$$P(Z > a) = P(\min(X, Y) > a) = P(X > a, Y > a) \tag{1}$$

But, X and Y are independent. So,

$$= P(X > a) \cdot P(Y > a) = e^{-a(\lambda_1+\lambda_2)} \tag{2}$$

So,

$$\boxed{F_z(a) = 1 - P(Z > a) = 1 - e^{-a(\lambda_1+\lambda_2)}} \tag{3}$$

(2.5 marks)

Let $Z = \max(X, Y)$. So, for $a > 0$,

$$P(Z < a) = P(\max(X, Y) < a) = P(X < a, Y < a) \tag{4}$$

But, X and Y are independent. So,

$$= P(X < a) \cdot P(Y < a) = (1 - e^{-a\lambda_1}) \cdot (1 - e^{-a\lambda_2}) = 1 - e^{-a\lambda_1} - e^{-a\lambda_1} + e^{-a(\lambda_1+\lambda_2)} \tag{5}$$

So,

$$\boxed{F_z(a) = 1 - e^{-a\lambda_1} - e^{-a\lambda_1} + e^{-a(\lambda_1+\lambda_2)}} \tag{6}$$

Q2

1(Equation)+1($X = 1|Y = 3$)+1($X = 2|Y = 3$)+1($X = 3|Y = 3$)+1(Final) mark

$$E\{X \mid Y = 3\} = 0 \cdot P\{X = 0 \mid Y = 3\} + 1 \cdot P\{X = 1 \mid Y = 3\} \tag{7}$$

$$+ 2 \cdot P\{X = 2 \mid Y = 3\} + 3 \cdot P\{X = 3 \mid Y = 3\} \tag{8}$$

$$= 1 \cdot \frac{4}{9} + 2 \cdot \frac{2}{9} + 3 \cdot \frac{1}{27} \tag{9}$$

$$= 1 \tag{10}$$

Q3

Assuming, $0 \le m \le n_1 + n_2$.
Let random variable Z be defined as $Z = X + Y$.

$$\begin{aligned}
P(Z = k) &= P(X + Y = k) \\
&= \sum_{i=0}^{k} P(X = i, Y = k - i) \\
&= \sum_{i=0}^{k} P(X = i)P(Y = k - i) \\
&= \sum_{i=0}^{k} \binom{n_1}{i} p^i (1-p)^{n_1 - i} \binom{n_2}{k-i} p^{k-i} (1-p)^{n_2 - k + i} \\
&= p^k (1-p)^{n_1 + n_2 - k} \sum_{i=0}^{k} \binom{n_1}{i} \binom{n_2}{k-i} \\
&= \binom{n_1 + n_2}{k} p^k (1-p)^{n_1 + n_2 - k}
\end{aligned}$$

Hence, we see that Z follows a binomial distribution, with parameters $n_1 + n_2$, and p.

$$\begin{aligned}
P(X = k | Z = m) &= \frac{P(X = k, Z = m)}{P(Z = m)} \\
&= \frac{P(X = k, Y = m - k)}{P(Z = m)} \\
&= \frac{P(X = k)P(Y = m - k)}{P(Z = m)} \\
&= \frac{\binom{n_1}{k} p^k (1-p)^{n_1 - k} \binom{n_2}{m-k} p^{m-k} (1-p)^{n_2 - m + k}}{\binom{n_1 + n_2}{m} p^m (1-p)^{n_1 + n_2 - m}} \\
&= \frac{\binom{n_1}{k} \binom{n_2}{m-k}}{\binom{n_1 + n_2}{m}}
\end{aligned}$$

(2 marks for finding the PDF of Z, 2 marks for correctly calculating the conditional PDF of X, 1 mark for final answer)

Q4

Consider a random variable U following uniform distribution in $[0, 2\pi]$. Define random variables $X = \sin(U)$, and $Y = \cos(U)$. Then
$E(XY) - E(X)E(Y) = 0$, implying they are uncorrelated. However $X^2 + Y^2 = 1$, implying they are not independent.
(1 mark for a correct example. 2 marks for showing X and Y are uncorrelated. 2 marks for showing X and Y are not independent.)

Q5

The probability is given by-

$$\begin{aligned} P\{X = n\} &= \int_{-\infty}^{\infty} P(X = n \cap \lambda \in (y, y + dy)) \\ &= \int_{\lambda} f(X = n|\lambda) f(\lambda) \\ &= \int_{0}^{\infty} \left(e^{-y} \frac{y^n}{n!}\right) \left(e^{-y}\right) dy \\ &= \int_{0}^{\infty} \frac{y^n e^{-2y}}{n!} dy \end{aligned}$$

[getting this form 4m]

Setting $t = 2y$ and hence $dt = 2dy$

$$= \frac{1}{2^{n+1} * n!} \int_{0}^{\infty} t^n e^{-t} dt$$

[substitute 2y by t 0.5m]

$\int_{0}^{\infty} t^n e^{-t} dt$ is an integral called the gamma function $\Gamma(n+1)$ which for integers evaluates to $n!$

$$= \left(\frac{1}{2}\right)^{n+1}$$

[substitute gamma 0.5m]

Q6

Given the joint distribution function $f_{X,Y}(x, y)$ as

$$f_{X,Y}(x, y) = \begin{cases} c(1 + xy) & \text{if } 2 \leq x \leq 3 \text{ and } 1 \leq y \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

Part 1 To find c:
Since f is the probability density function jointly defined over x and y:

$$\int_{\mathbb{R}^2} f_{X,Y} = 1$$

$$\begin{aligned} &\implies \int_{2}^{3} \int_{1}^{2} c(1 + xy) dy dx = 1 \\ &\implies \int_{2}^{3} c\left(1 + \frac{3x}{2}\right) = 1 \\ &\implies c\frac{19}{4} = 1 \\ &\implies c = \frac{4}{19} \end{aligned}$$

[1m for getting c, -0.5 for calc errors]

Part 2 To find f_X and f_Y:
We find f_X by marginalizing over all y

$$\begin{aligned} f_X &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \\ &= \int_1^2 \frac{4}{19}(1+xy)dy \\ &= \frac{4+6x}{19} \end{aligned}$$

[2m for getting c, -0.5 for calc errors]

Similarly we find f_Y by marginalizing over all x

$$\begin{aligned} f_Y &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \\ &= \int_2^3 \frac{4}{19}(1+xy)dx \\ &= \frac{4+10y}{19} \end{aligned}$$

[2m for getting c, -0.5 for calc errors]

Q7

The probability is given by-

$$\begin{aligned} P\{X=n\} &= \int_{-\infty}^{\infty} P(X=n \cap \lambda \in (y, y+dy)) && (11) \\ &= \int_0^{\infty} \left(e^{-y}\frac{y^n}{n!}\right)\left(ye^{-y}\right) dy && (12) \\ &= \int_0^{\infty} \frac{y^{n+1}e^{-2y}}{n!} dy && (13) \\ &= \frac{n+1}{2^{n+2}} && (14) \end{aligned}$$

Marking scheme:

- Equation (11) : 2.5 Marks
- Correct final answer: 2.5 Marks

Q8

Let the number of men and women visiting the academy on a day be represented by the RVs x & Y. $X+Y$ is a Poisson Distribution. Now, by Total Probability

Theorem,

$$P(X=m,Y=n)=\sum_{i=0}^{\infty}P\{X=m,Y=n\mid X+Y=i\}P(X+Y=i) \tag{15}$$
$$=P\{X=m,Y=n\mid X+Y=m+n\}e^{-\lambda}\frac{\lambda^{n+m}}{(n+m)!} \tag{16}$$
$$=\binom{m+n}{n}p^n(1-p)^m e^{-\lambda}\frac{\lambda^{n+m}}{(n+m)!} \tag{17}$$

Marking scheme:

- Equation 15 or 16: 2.5 Marks
- Correct final answer: 2.5 Marks
- Directly writing (16) is also fine.

Q9

1. (2.5 marks)

$$\begin{aligned}
Cov(aX_1+b,cX_2+b) &= E[(aX_1+b-E[aX_1+b])*(cX_2+b-E[cX_2+b])] \\
&= E[(aX_1+b-E[aX_1]+b)*(cX_2+b-E[cX_2]+b)] \\
&= E[(aX_1-aE[X_1])*(cX_2-cE[X_2])] \\
&= E[a(X_1-E[X_1])*c(X_2-E[X_2])] \\
&= acE[(X_1-E[X_1])*(X_2-E[X_2])] \\
&= acCov(X_1,X_2)
\end{aligned}$$

2. (2.5 marks)

$$\begin{aligned}
Cov(X_1+X_2,X_3) &= E[(X_1+X_2-E[X_1+X_2])*(X_3-E[X_3])] \\
&= E[(X_1+X_2-E[X_1]-E[X_2])*(X_3-E[X_3])] \\
&= E[((X_1-E[X_1])+(X_2-E[X_2]))*(X_3-E[X_3])] \\
&= E[((X_1-E[X_1])*(X_3-E[X_3])+(X_2-E[X_2])*(X_3-E[X_3]))] \\
&= E[(X_1-E[X_1])*(X_3-E[X_3])]+E[(X_2-E[X_2])*(X_3-E[X_3])] \\
&= Cov(X_1,X_3)+Cov(X_2,X_3)
\end{aligned}$$

Q10

We know that :

$$\epsilon=\sqrt{\frac{1}{n}\log(\delta/2)}$$

Plugging in n as 100, and δ as 0.05 (the probability of error < 0.05) we get ϵ as 0.192. Hence the confidence interval $(\mu - \epsilon, \mu + \epsilon)$ is $(0.258, 0.642)$ (3 marks)

Since the confidence interval is inversely proportional to the root of number of samples $\epsilon \propto \sqrt{\frac{1}{n}}$, to decrease the interval by half we need to increase the number of samples by a factor of 4 i.e. 4n = 400 (2 marks)