DS203 Assigment 2

Exercise 1

PDF for X $f_X(x) = \begin{cases} \lambda_1 e^{-\lambda_1 x}, & x \geq 0 \\ 0, & x \leq 0 \end{cases}$

PDF for Y $f_Y(y) = \begin{cases} \lambda_2 e^{-\lambda_2 y}, & y \geq 0 \\ 0, & y < 0 \end{cases}$

(a) let function $g = \min(X, Y)$.

CDF for g, $F_g(z) = P(\min(X, Y) \leq z)$

$P(\min(X, Y) \leq z) = 0 \quad \forall z < 0$ as X and Y take positive values

$$P(\min(x,y) \leq z) = \sum_{\forall j \in (0, z)} P(X = j, j \leq Y < \infty) + \sum_{\forall j \in (0, z)} P(j \leq X < \infty, Y = j)$$

if $z > 0$

Page No
Date

$$= \int_0^z \int_j^\infty \lambda_1 \lambda_2 e^{-\lambda_1 j} e^{-\lambda_2 y} \, dy \, dj + \int_0^z \int_j^\infty \lambda_1 \lambda_2 e^{-\lambda_1 x} e^{-\lambda_2 j} \, dx \, dj$$

$$= \lambda_1 \int_0^z e^{-(\lambda_1+\lambda_2)j} \, dj + \lambda_2 \int_0^z e^{-(\lambda_1+\lambda_2)j} \, dj$$

$$= 1 - e^{-(\lambda_1+\lambda_2) z} \quad \text{if } z \geq 0$$

(b) $g = \max(X, Y)$

CDF for g, $F_g(z) = P(\max(X,Y) \leq z)$

$P(\max(X,Y) \leq z) = 0 \quad \forall z < 0$ as X and Y take positive values.

$$P(\max(X,Y) \leq z) = \sum_{\forall j \in (0,z)} P(X=j, 0 \leq Y \leq j) + \sum_{\forall j \in (0,z)} P(0 \leq x \leq j, Y=j)$$

$$= \int_0^z \int_0^j \lambda_1 \lambda_2 e^{-\lambda_1 j} e^{-\lambda_2 y} \, dy \, dj + \int_0^z \int_0^j \lambda_1 \lambda_2 e^{-\lambda_1 x} e^{-\lambda_2 j} \, dx \, d$$

Page No
Date

$$= \lambda_1 \int_0^z e^{-\lambda_1 j} (1 - e^{-\lambda_2 j})\, dj + \lambda_2 \int_0^z e^{-\lambda_2 j} (1 - e^{-\lambda_1 j})\, dj$$

$$= (1 - e^{-\lambda_1 z}) + (1 - e^{-\lambda_2 z}) - (1 - e^{-(\lambda_1 + \lambda_2) z})$$

$$= 1 - e^{-\lambda_1 z} - e^{-\lambda_2 z} + e^{-(\lambda_1 + \lambda_2) z}$$

if $z \geq 0$

Exercise 2 3 white. 6 red. 5 blue.

Total 6 selections

X = number of white balls selected
Y = number of blue balls selected

$$E[X \mid Y=3] = 1 \cdot P(X=1 \mid Y=3) + 2 \cdot P(X=2 \mid Y=3) + 3 \cdot P(X=3 \mid Y=3)$$

$$= \frac{P(X=1 \cap Y=3)}{P(Y=3)} + 2 \cdot \frac{P(X=2 \cap Y=3)}{P(Y=3)} + 3 \cdot \frac{P(X=3 \cap Y=3)}{P(Y=3)}$$

In the 6 selections available, if we have to pick m white and n blue balls, then the total number of ways are :

selecting n blue from remaining $(6-m)$ selections

$$= {}^{6}C_{m} \cdot \left(\frac{3}{14}\right)^{m} \cdot {}^{(6-m)}C_{n} \left(\frac{5}{14}\right)^{n} \cdot {}^{6-m-n}C_{6-m-n} \left(\frac{6}{14}\right)^{6-m-n} \quad - (1)$$

ways of having m white selections out of 6

Probability of selecting m white balls

$= 1$

Here, I have assumed $m+n \leq 6$ to demonstrate this case.

Using (1),

$$E[X \mid Y=3] = \frac{{}^{6}C_{1} \cdot \left(\frac{3}{14}\right) \cdot {}^{5}C_{3} \left(\frac{5}{14}\right)^{3} \cdot \left(\frac{6}{14}\right)^{2} + 2 \cdot {}^{6}C_{2} \cdot \left(\frac{3}{14}\right)^{2} \cdot {}^{4}C_{3} \left(\frac{5}{14}\right)^{3} \cdot \left(\frac{6}{14}\right) + 3 \cdot {}^{6}C_{3} \left(\frac{3}{14}\right)^{3} \cdot {}^{3}C_{3} \cdot \left(\frac{5}{14}\right)^{3}}{{}^{6}C_{3} \left(\frac{5}{14}\right)^{3} \cdot {}^{3}C_{3} \left(\frac{9}{14}\right)^{3}}$$

$$= \frac{{}^{6}C_{1} \cdot {}^{5}C_{3} \cdot 3 \cdot 6^{2} + {}^{6}C_{2} \, {}^{4}C_{3} \, 3^{2} \cdot 6 \cdot 2 + {}^{6}C_{3} \, 3^{3} \cdot 3}{{}^{6}C_{3} \cdot 9^{3}}$$

$$= \frac{6480 + 6480 + 1620}{14580} = \underline{\underline{1}}$$

Exercise 3 $\quad X_1 \sim Bin(n_1, p), \quad X_2 \sim Bin(n_2, p)$

$$P(X_1 = i \mid X_1 + X_2 = m) = \frac{P(X_1 = i \cap X_1 + X_2 = m)}{P(X_1 + X_2 = m)}$$

$$= \frac{{}^{n_1}C_i \, p^i (1-p)^{n_1-i} \; {}^{n_2}C_{m-i} \, p^{m-i} (1-p)^{n_2-m+i}}{\sum_{j=0}^{m} {}^{n_1}C_j \, p^j (1-p)^{n_1-j} \cdot {}^{n_2}C_{m-j} \, p^{m-j} (1-p)^{n_2-m+j}}$$

$$= \frac{{}^{n_1}C_i \cdot {}^{n_2}C_{m-i} \cdot p^m \cdot (1-p)^{n_1+n_2-m}}{p^m (1-p)^{n_1+n_2-m} \sum_{j=0}^{m} {}^{n_1}C_j \, {}^{n_2}C_{m-j}}$$

$$= \frac{{}^{n_1}C_i \cdot {}^{n_2}C_{m-i}}{\sum_{j=0}^{m} {}^{n_1}C_j \cdot {}^{n_2}C_{m-j}}$$

The denominator is the coefficient of x^m in $(1+x)^{n_1} \cdot (1+x)^{n_2} = (1+x)^{n_1+n_2} = {}^{n_1+n_2}C_m$.

$$P(X_1 = i \mid X_1 + X_2 = m) = \frac{{}^{n_1}C_i \cdot {}^{n_2}C_{m-i}}{{}^{n_1+n_2}C_m}$$

This formula is however, only valid for $n_1 + n_2 \geq m$.

If $n_1 + n_2 < m$, the denominator does not make any sense, hence, the PMF of X_1 is identically 0.

If $n_1 + n_2 = m$, then the only possible values taken by X_1, X_2 s.t $X_1 + X_2 = m$ is $X_1 = n_1$, $X_2 = n_2$.

Hence, the PMF of X_1 given $X_1 + X_2 = m$ will be :

$$P(X_1 = n_1 \mid X_1 + X_2 = m) = \frac{{}^{n_1}C_{n_1} \cdot {}^{n_2}C_{n_2}}{{}^{n_1+n_2}C_{n_1+n_2}}$$

$$= 1.$$

In all other cases with $n_1 + n_2 > m$, the formula derived earlier is valid and gives the PMF of X_1 given $X_1 + X_2 = m$.

Date

Exercise 4

For X and Y to be uncorrelated, it is sufficient that $\operatorname{Cov}(X, Y) = 0$

$$\operatorname{Cov}(X, Y) = E[XY] - E[X]\,E[Y] = 0$$
$$\Rightarrow E[XY] = E[X]\,E[Y]$$

Consider $X \sim U(-1, 1)$
$Y = X^2$

Clearly, X and Y are dependent.

But, $E[X] = 0$
and since $XY = X^3 \sim U(-1, 1)$,
$E[XY] = 0$.
Hence, $\operatorname{Cov}(X, Y) = 0$.

So, X and Y are uncorrelated but not independent

Exercise 5

$$X \sim \text{Poi}(\lambda).$$
$$\lambda \sim \text{Exp}(1)$$

$$P(X=i, \lambda) = \frac{e^{-\lambda} \lambda^{i}}{i!}, \quad i \geq 0.$$

PDF of λ,

$$f_{\lambda}(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

$$P\{X=n\} = \int_{-\infty}^{\infty} P(X=n, \lambda=x) \cdot f_{\lambda}(x)\, dx$$

$$= \int_{0}^{\infty} \frac{e^{-x} x^{n}}{n!} e^{-x}\, dx$$

$$= \frac{1}{n!} \int_{0}^{\infty} x^{n} e^{-2x}\, dx$$

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

$$= \frac{\Gamma(n+1)}{n!\, 2^{n+1}} = \boxed{\frac{1}{2^{n+1}}}$$

Page No
Date

Exercise 6

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

1. $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y) = 1$$

$$\Rightarrow \int_{1}^{2}\int_{2}^{3} c(1+xy)\,dx\,dy = 1.$$

$$c\left[1 \cdot 1 + \frac{3}{2} \cdot \frac{5}{2}\right] = 1.$$

$$c = \frac{4}{19}$$

$$f_X = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy.$$

$$= \int_{1}^{2} \frac{4}{19}(1+xy)\,dy = \frac{4}{19}\left(1 + \frac{3x}{2}\right).$$

Date

$$f_y = \int_{-\infty}^{\infty} f_{x,y}(x,y)\, dx$$

$$= \int_{2}^{3} \frac{4}{19}(1+xy)\, dx = \frac{4}{19}\left(1+\frac{5}{2}y\right)$$

Exercise 7

let X = number of accidents random variable.

$$X \sim Poi(\lambda_1)$$

where λ_1 is a random variable s.t. P.D.F of λ_1 is

$$g(\lambda) = \lambda e^{-\lambda}, \quad \lambda \geq 0$$

$$P(X=n, \lambda_1 = p) = \frac{e^{-p} p^n}{n!}, \quad n \geq 0.$$

$$P(X=n) = \int_{-\infty}^{\infty} P(X=n, \lambda_1 = p) \cdot g(p)\, dp.$$

$$= \int_{0}^{\infty} \frac{e^{-p} p^n}{n!} \, p e^{-p}\, dp$$

Date

$$= \frac{1}{n!\,2^{n+2}} \int_0^\infty t^{n+1} e^{-t}\, dt$$

$$= \frac{1}{n!\,2^{n+2}} \Gamma(n+2).$$

$$= \boxed{\frac{n+1}{2^{n+2}}}$$

Exercise 8

let X be number of people visiting for yoga each day.

$$X \sim Poi(\lambda).$$

$$P(X=n) = \frac{e^{-\lambda} \lambda^n}{n!}, \quad n \geq 0.$$

If n women & m men visit today, total number visiting $= (m+n)$.

$$P(X=(m+n)) = \frac{e^{-\lambda} \lambda^{m+n}}{(m+n)!}$$

$$P(n \text{ women}, m \text{ men}) = P(X = (m+n)) \cdot {}^{m+n}C_{n}\, p^{n} \cdot {}^{m}C_{m} (1-p)^{m}$$

$$= \frac{e^{-\lambda} \lambda^{m+n}}{(m+n)!} \cdot \frac{(m+n)!}{m!\, n!} \cdot p^{n} \cdot (1-p)^{m}$$

$$= \frac{e^{-\lambda} \lambda^{m+n} p^{n} (1-p)^{m}}{m!\, n!}$$

<u>Exercise 9</u>

(a) $\operatorname{cov}(aX_1 + b, cX_2 + b)$

$$= E[(aX_1 + b)(cX_2 + b)] - E[aX_1 + b]\, E[cX_2 + b]$$

$$= E[acX_1X_2 + abX_1 + bcX_2 + b^2] - (aE[X_1] + bE[1])(cE[X_2] + bE[1])$$

$$= ac\, E[X_1X_2] + ab\, E[X_1] + bc\, E[X_2] + b^2 - ac\, E[X_1]E[X_2] - ab\, E[X_1] - bc\, E[X_2] - b^2$$

$(\because E[1] = 1)$

$$= ac(E[X_1X_2] - E[X_1]E[X_2])$$

$$= ac \operatorname{Cov}(X_1, X_2)$$

(b) $\text{Cov}(X_1 + X_2, X_3)$

$$= E[(X_1 + X_2) X_3] - E[X_1 + X_2] E[X_3]$$

$$= E[X_1 X_3] + E[X_2 X_3] - (E[X_1] + E[X_2]) E[X_3]$$

$$= (E[X_1 X_3] - E[X_1] E[X_3]) + (E[X_2 X_3] - E[X_2] E[X_3])$$

$$= \text{Cov}(X_1, X_3) + \text{Cov}(X_2, X_3)$$

Exercise 10

(a) number of samples $n = 100$.

Estimated mean $\hat{p} = 0.45$

let the true mean after taking large number of samples be p.

let $|p - \hat{p}| \geq \varepsilon$ for some $\varepsilon > 0$.

Then, we know,

$$P(|p - \hat{p}| \geq \varepsilon) \leq 2e^{-n\varepsilon^2}$$

Thus, $\delta = 2e^{-n\varepsilon^2}$ gives the maximum probability with which p lies outside $\hat{p} - \varepsilon$ and $\hat{p} + \varepsilon$.

Since we want p to lies between $\hat{p} - \varepsilon$ and $\hat{p} + \varepsilon$ with probability at least 0.95.

$$\Rightarrow \quad \delta = 1 - 0.95 = 0.05$$

$$\varepsilon = \sqrt{\frac{-1}{n} \ln\left(\frac{\delta}{2}\right)}$$

$$= \sqrt{\frac{-1}{100} \ln\left(\frac{0.05}{2}\right)}$$

$$= 0.192$$

So, the confidence interval is $(\hat{p} - \varepsilon, \hat{p} + \varepsilon)$

$$= (0.45 - 0.192, 0.45 + 0.192)$$

$$= (0.258, 0.642)$$

(b) For confidence interval of $(\hat{p} - \varepsilon, \hat{p} + \varepsilon)$ to shrink by half $\left(\hat{p} - \frac{\varepsilon}{2}, \hat{p} + \frac{\varepsilon}{2}\right)$, ε must shrink by half.

let $\varepsilon = \sqrt{\frac{1}{n_1} \ln\left(\frac{\delta}{2}\right)}$, $\frac{\varepsilon}{2} = \sqrt{\frac{1}{n_2} \ln\left(\frac{\delta}{2}\right)}$

$$\Rightarrow \quad n_2 = 4n_1$$

Since $n_1 = n$

$\Rightarrow \quad n_2 = 4n.$

So, we need $(n_2 - n_1) = 3n$ more samples.