SHIVAM PATEL 200070077 DS203 Assignment 2

Ans 1. X and Y are independent RV with exponential distribution, and parameters λ_1 & λ_2.

(a) Distribution of min (X, Y)

$$P(\min(X,Y) \le x) = 1 - P(\min(X,Y) > x)$$

$$= 1 - P((X > x) \cap (Y > x))$$

$$= 1 - P(X > x)\, P(Y > x)$$

$$= 1 - e^{-\lambda_1 x} \times e^{-\lambda_2 x}$$

$$\boxed{= 1 - e^{-(\lambda_1 + \lambda_2) x}.}$$

(b) Distribution of max (x, y)

$$P(\max(X,Y) \le x) = P(X \le x \text{ and } Y \le x)$$

$$= P(X \le x) \cdot P(Y \le x)$$

$$= (1 - e^{-\lambda_1 x}) \cdot (1 - e^{-\lambda_2 x})$$

$$\boxed{= 1 - e^{-\lambda_1 x} - e^{-\lambda_2 x} + e^{-(\lambda_1 + \lambda_2) x}.}$$

[$\because$ CDF of exp(x) = $1 - e^{-\lambda x}$]

Ans 2. 3 white, 6 Red & 5 blue balls.

Selection with Replacement.

$X \to$ No. of white balls $\quad Y \to$ No. of blue balls.

$E[X|Y=3] \to$ Expectation of getting white ball when 3 blue balls are already chosen.

$$E[X|Y=3] = \sum x \, P[x|Y=3]$$

$$= \frac{{}^3C_3 \times {}^6C_0 \times 3}{{}^9C_3} + \frac{{}^3C_2 \times {}^6C_1 \times 2}{{}^9C_3}$$

$$+ \frac{{}^3C_1 \times {}^6C_2 \times 1}{{}^9C_3} + \cancel{\frac{{}^3C_0 \, {}^6C_3 \times 0}{{}^9C_3}}^{\,0}$$

$$= \frac{45}{84} + \frac{36}{84} + \frac{3}{84} = 1.$$

Hence, $\boxed{E[X|Y=3] = 1.}$

Ans 3. X_1, X_2 are independent $Bin(n_1, p)$ and $Bin(n_2, p)$.

We need to find the conditional prob. mass f^n of X_1 given that $X_1 + X_2 = m$

$$\text{PMF}(X_1 \mid X_1 + X_2 = m) = \frac{P_1(X_1)\, P_2(X_2)}{\Sigma P_1(X_1)\, P_2(X_2)}$$

$$= \frac{P_1(X_1)\, P_2(m - X_1)}{P_1(0)\, P_2(m-0) + P_1(1)\, P_2(m-1) \cdots P_1(m)\, P_2(0)}$$

$$= \frac{\left({}^{n_1}C_{x_1}\, p^{x_1} (1-p)^{n_1 - x_1}\right) \times \left({}^{n_2}C_{m-x_1} (1-p)^{n_2 - m + x}\, p^{m-x}\right)}{\sum_{i=0}^{m} {}^{n_1}C_i\, p^i (1-p)^{n_1 - i} \times {}^{n_2}C_{m-i} (1-p)^{n_2 - m + i}\, p^{m-i}}$$

$$\text{Denominator} = p^m (1-p)^{n_1 + n_2 - m} \times \left[{}^{n_1}C_0\, {}^{n_2}C_m + {}^{n_1}C_1\, {}^{n_2}C_{m-1} + \cdots {}^{n_1}C_m\, {}^{n_2}C_0 \right]$$

Simplifying as a product of bino-series,

we have

$$D^r = p^m (1-p)^{n_1 + n_2 - m} \times \left({}^{n_1 + n_2}C_m\right)$$

Ans 3 contd

$\therefore$ PMF $(X_1 \mid X_1 + X_2 = m)$

$$= \frac{\left({}^{n_1}C_{x_1} \; {}^{n_2}C_{m-x_1} \right) p^m (1-p)^{n_1+n_2-m}}{p^m (1-p)^{n_1+n_2-m} \left({}^{n_1+n_2}C_m \right)}$$

$$\boxed{\text{PMF}\left(X_1 \mid X_1 + X_2 = m\right) = \frac{{}^{n_1}C_{x_1} \; {}^{n_2}C_{m-x_1}}{{}^{n_1+n_2}C_m}}$$

Ans 4. Let us construct a ~~d~~ f^n with discrete domain, given ~~as~~ by the probability density /mass table as →

P_i

$x \backslash y$	2	1	0	1	2
1	1/5	0	0	0	0
2	0	1/5	0	0	0
3	0	0	1/5	0	0
4	0	0	0	1/5	0
5	0	0	0	0	1/5

Here, we have two dependent r.v.s x and y, such that

~~x[i]~~

$y[i] = |x[i] - 3|$

~~So~~ Also, prob. of occuring of each $x[i]$ is uniform, i.e. $\frac{1}{5}$. Similarly as y is completely dependent on X, $P[y[i]] = \frac{1}{5}$ $\forall i \in \{1, 2, 3, 4, 5\}$

Now, we find the covariance.

$$COV(X, Y) = E(XY) - E(X)\,E(Y)$$

$$E(XY) = \Sigma x_i y_i P(i) = 2\times1\times\frac{1}{5} + 1\times2\times\frac{1}{5} + 3\times0\times\frac{1}{5} + 4\times1\times\frac{1}{5} + 5\times2\times\frac{1}{5} = \frac{18}{5}$$

$$E(X) = \Sigma x\, P(i) = \frac{1+2+3+4+5}{5} = 3.$$

$$E(Y) = \Sigma y\, P(i) = \frac{2+1+0+1+2}{5} = \frac{6}{5}.$$

~~COV(X,Y) = E(XY)~~

Ans 4 contd.

$$\therefore \; cov(X,Y) = E(XY) - E(X)\,E(Y)$$

$$= \frac{18}{5} - (3)\left(\frac{6}{5}\right) = 0.$$

Hence, X and Y are not correlated. But they are completely dependent.

Without Loss of generality, this argument can be extended to real nos also, where

$x \in [0,6]$ and $y = |x-3|$

with x being a uniformly distributed RV.

This shows that

$$\boxed{\text{non correlated} \not\Rightarrow \text{Independence of RVs}}.$$

Ans 5. X is Poisson RV with mean λ.

$\therefore$ $X(i) = \dfrac{\lambda^i e^{-\lambda}}{i!}$

But λ itself is exponential with mean 1, i.e.

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

By applying the Law of the Unconcious Statistician, we have

$$P\{X=n\} = \int_0^\infty \frac{t^n e^{-t}}{n!} \times e^{-t}\, dt$$

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

Let $2t = \alpha$

$\therefore$ $2dt = d\alpha$

$$\therefore P\{X=n\} = \int_0^\infty \frac{e^{-\alpha}\alpha^n}{n! \times 2^n} \times \frac{d\alpha}{2} \qquad \left[\because \int_0^\infty x^n e^{-x} dx = n!\right]$$

$$= \frac{n!}{n!} \times \frac{1}{2^{n+1}} = \frac{1}{2^{n+1}}$$

$$\therefore \boxed{P\{X=n\} = 1/2^{n+1}}$$

Ans 6. $f_{x,y}(x,y) = c(1+xy)$

By finding marginal PDFs, we can obtain $f_X(x)$ and $f_Y(y)$ in terms of c. And integrating the PDF will give us the value of c.

$$\therefore \int_1^2 f_{x,y}(x,y)\,dy = f_x(x)$$

$$\therefore f_x(x) = \int_1^2 c(1+xy)\,dy = c\left(1+\frac{3x}{2}\right)$$

$$\int_2^3 f_x(x)\,dx = 1$$

$$\therefore \int_2^3 c\left(1+\frac{3x}{2}\right)dx = 1$$

$$\therefore c\left(1+\frac{15}{4}\right) = 1$$

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

(i) Hence $c = 4/19$.

(ii)

Ans 6 contd.

(2). $f_X(x) = \int_1^2 f_{(X,Y)}(x,y)\,dy$

$= c\left(1+\frac{3x}{2}\right) = \frac{4}{19}\left(1+\frac{3x}{2}\right)$

Similarly $f_Y(y) = \int_2^3 f_{(X,Y)}(x,y)\,dx$

$= c\left(1+\frac{5xy}{2}\right)$ ~~$= \frac{4}{19}(\;)$~~

$= \frac{4}{19}\left(1+\frac{5xy}{2}\right)$.

Ans 7. No of accidents is modelled as a Poisson distribution.

$$P(x=i) = \frac{\lambda^i e^{-\lambda}}{(i)!}$$

But mean has a gamma distribution as $\Rightarrow$

$$g_\lambda(t) = te^{-t} \quad , \quad t \geq 0$$

By applying the Law of the Unconcious Statisticians, we get

$$P(x=n) = \int_0^\infty \frac{\lambda^n e^{-\lambda}}{n!} \times \lambda e^{-\lambda} \, d\lambda$$

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

Let $2\lambda = \alpha \Rightarrow 2d\lambda = d\alpha$.

$$\therefore P(x=n) = \int_0^\infty \frac{1}{n!} \times \frac{\alpha^{n+1} e^{-\alpha} \, d\alpha}{2^{n+2}}$$

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

$\therefore$ $\boxed{P(X=n) = \frac{n+1}{2^{n+2}}}$ is the prob. of randomly chosen person having n accidents next yr.

Ans 8. No. of people visiting the studio → Poisson RV, λ.

$$= \frac{\lambda^i e^{-\lambda}}{i!}$$

Each person who visits is male with prob $1-p$ & female with prob p.

Joint probability that exactly n women & m men visit the academy at any day.

We know that the distributions (Poisson and binomial) in this case are jointly continuous.

So, we can say that

$$P\begin{pmatrix} m \text{ men} \\ n \text{ women} \end{pmatrix} = \text{Prob of } (m+n) \text{ people visiting} \times \text{Prob of } m \text{ men and } n \text{ women out of } m+n \text{ people}$$

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

Hence, the joint probability is →

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

Ans 9. $\mathrm{cov}(X_1, X_2) = E((X_1 - E(X_1))(X_2 - E(X_2)))$

$$= E(X_1 X_2 - E(X_1) X_2 - X_1 E(X_2) + E(X_1) E(X_2))$$

$$= E(X_1 X_2) - E(X_1) E(X_2) - E(X_1) E(X_2) + E(X_1) E(X_2)$$

$$= E(X_1 X_2) - E(X_1) E(X_2)$$

We also know that E operator is linear

(a) $\mathrm{cov}(aX_1 + b, cX_2 + b) = ac\, \mathrm{cov}(X_1, X_2)$

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

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

$$= ac E(X_1 X_2) + ab E(X_1) + bc E(X_2) + b^2 - ac E(X_1) E(X_2) - ab(E(X_1)) - bc E(X_2) - b^2$$

$$= ac [E(X_1 X_2) - E(X_1) E(X_2)] = ac\, \mathrm{cov}(X_1, X_2) = RHS$$

Hence proved.

(b) $\mathrm{cov}(X_1 + X_2, X_3) = \mathrm{cov}(X_1, X_3) + \mathrm{cov}(X_2, X_3)$

$\mathrm{cov}(X_1 + X_2, X_3) = E[(X_1+X_2)(X_3)] - E(X_1+X_2) E(X_3)$

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

$$\cancel{= E(X_1 X_3) + E(X_2 X_3) - E(X_1)E(X_3) - E(X_2)X_3)}$$

$$\cancel{= E(X_1 X_3) - E}$$

Ans 9 contd.

$$= E(X_1X_3) + E(X_2X_3) - (E(X_1) + E(X_2))\, E(X_3)$$

$$= E(X_1X_3) + E(X_2X_3) - (E(X_1)E(X_3) + E(X_2)E(X_3))$$

$$= (E(X_1X_3) - E(X_1)E(X_3)) + (E(X_2X_3) - E(X_2)E(X_3))$$

$$= cov(X_1, X_3) + cov(X_2, X_3) = RHS$$

Hence Proved.

Ans10. $n = 100$ i.i.d samples

$\hat{\mu} = 0.45$ avg (estimated)

We know that

$$P((\hat{\mu} - \mu) > \varepsilon) \leq 2e^{-n\varepsilon^2}$$

(a) So, to get interval where prob of finding it is atleast 0.95, we have

$$0.95 = 2e^{-100\varepsilon^2}$$

$\therefore$ $\varepsilon = 0.0863$

$\therefore$ our confidence interval is 0.45 ± 0.0863, i.e. the confidence interval is ~~36~~ $(-0.3637, 0.5362)$

(b) For confidence interval to shrink to half, we need to reduce the value of ε by half. But from the above ~~ara~~ method, we see that

$\varepsilon \propto \frac{1}{\sqrt{n}}$ where n is No. of samples.

To shrink ε by half, we need n to become 4 times

Hence, the number of samples needed to reduce the confidence interval to half is $\underline{n = 400}$.
So we will need $\boxed{300}$ more samples.
new confidence interval $\rightarrow$ $(-0.4068, 0.4931)$

(If is estimate is 0.45).

Or more generally, confidence interval $\rightarrow$ $(\hat{\mu} - 0.04315, \hat{\mu} + 0.04315)$ ~~$\hat{\mu}$ 0.45~~