DS203 - Assignment 2

210020120

SAUMYA SHETH

Exercise 1

$X_1 \rightarrow (n_1, p) \quad X_2 \rightarrow (n_2, p)$

Conditional probability mass function (pmf) of X_1 given that $X_1 + X_2 = m$

$$P(X_1 = k \mid X_1 + X_2 = m)$$

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

$$= \frac{P(X_1 = k, X_2 = m-k)}{P(X_1 + X_2) = m}$$

$$= \frac{P(X_1 = k)\, P(X_2 = m-k)}{P(X_1 + X_2) = m} \quad (\because X_1, X_2 \text{ are independent})$$

$$= \frac{{}^{n_1}C_k \, p^k (1-p)^{n_1-k} \cdot {}^{n_2}C_{m-k} \, p^{m-k} (1-p)^{n_2-m+k}}{{}^{n_1+n_2}C_m \, p^m (1-p)^{n_1+n_2-m}}$$

$$= \frac{{}^{n_1}C_k \cdot {}^{n_2}C_{m-k} \cdot \cancel{p^m} \, \cancel{(1-p)^{n_1+n_2-m}}}{{}^{n_1+n_2}C_m \cdot \cancel{p^m} \, \cancel{(1-p)^{n_1+n_2-m}}}$$

$$= \boxed{\frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n_1+n_2}{m}}}$$

Ans The conditional pmf of X_1 given that $X_1 + X_2 = m$ is

$$P(X_1 = k \mid X_1 + X_2 = m) = \frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n_1+n_2}{m}}$$

Exercise 2

Two Random variables X_1 and X_2 are correlated if $|cov(X_1, X_2)| > 0$.

For two random variable X, Y to be uncorrelated ~~$cov(X_1, X_2) = 0$~~

$$cov(X, Y) = 0$$

$$cov(X, Y) = E(XY) - E(X)E(Y)$$

Let's define x such that $E(X) = 0$

Let X be a discrete random variable and

$X \in \{-1, 0, 2\}$ with

$P(X = -1) = 1/2$, $P(X=0) = 1/4$, $P(X=2) = 1/4$

$E(X) = \sum x_i P(x_i) = -1 \times \frac{1}{2} + 0 \times \frac{1}{4} + 2 \times \frac{1}{4} = 0$

Now to make $cov(X, Y) = 0$, $E(XY)$ has to be 0.

We can define a Y such that XY is always zero, and in that case we will get $E(XY) = 0$

Let $Y = \begin{cases} 3 & \text{when } X = 0 \\ 0 & \text{otherwise} \end{cases}$

For this case, XY would always be 0 and hence $E(XY) = 0$

$\therefore cov(X, Y) = E(XY) - E(X)E(Y) = 0$

Hence X and Y are uncorrelated.

Now $P(X=0 \cap Y=0)$ i.e. both X and Y are zero is 0 ~~$(P(X=0 \cap Y=0) = 0)$~~ $[P(X=0 \cap Y=0) = 0]$ and $P(X=0)$ is $1/4$ and even $P(Y=0)$ is also non-zero. $\therefore P(X=0 \cap Y=0) \neq P(X=0)\,P(Y=0)$

$\therefore X$ and Y are not independent.

Ans: $\boxed{\begin{array}{l} \text{Example of two random variables } X, Y \text{ that are uncorrelated} \\ \text{but not independent is } X \in \{-1, 0, 2\} \text{ (discrete random variable)} \\ \text{with } P(X=-1) = 1/2,\ P(X=0) = 1/4,\ P(X=2) = 1/4 \text{ and} \\ Y = \begin{cases} 3 & \text{when } X = 0 \\ 0 & \text{otherwise} \end{cases} \end{array}}$

Exercise 3

1. $f_{X,Y}(x, y) = c(1+xy)$ if $2 \le x \le 3$ & $1 \le y \le 2$

$= 0$ otherwise.

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

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

$$\int_{1}^{2} c\left(x + y\frac{x^2}{2}\right)\Big|_2^3 \,dy = 1$$

$$\int_{1}^{2} c\left(3 + \frac{9y}{2} - (2 + 2y)\right) dy = 1$$

$$\int_{1}^{2} c\left(1 + \frac{5y}{2}\right) dy = 1$$

$$c\left(y + \frac{5y^2}{4}\right)\Big|_1^2 = 1$$

$$c\left(2 + 5 - (1 + 5/4)\right) = 1$$

$$c\left(6 - 5/4\right) = 1$$

$$c\left(\frac{19}{4}\right) = 1$$

$$\boxed{c = 4/19}$$

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

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

$= c(y + xy^2/2)\Big|_1^2$

$= c(2 + 2x - 1 - x/2)$

$= c(1 + 3x/2) = 4/19\,(1 + 3x/2)$

$= 4/19 + 6x/19 = \dfrac{6x+4}{19}$

~~$f_Y(y)$~~ $\boxed{f_X(x) = \dfrac{6x+4}{19}}$

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

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

$= c(x + yx^2/2)\Big|_2^3$

$= c(3 + 9y/2 - 2 - 2y)$

$= 4/19\,(1 + 5y/2)$

$= 4/19 + 10y/19$

$= \dfrac{10y+4}{19}$

$\boxed{f_Y(y) = \dfrac{10y+4}{19}}$

Ans :- 1. $c = 4/19$

2. $f_X = \dfrac{6x+4}{19}$; $f_Y = \dfrac{10y+4}{19}$

Exercise 4

No. of accidents in a year $\rightarrow$ Poisson distributed ~~with mean~~ Poisson mean has a gamma distribution with density function,

$$g(\lambda) = \lambda e^{-\lambda}.$$

$X \rightarrow$ No. of accidents random policy holder has next year.

$Y \rightarrow$ Poisson mean number of accidents.

$$P(X=n) = \int_0^\infty P\{X=n \mid Y=\lambda) g(\lambda)\, d\lambda$$

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

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

$X \to \text{Gamma}(\alpha, \lambda)$

$$f_X(x) = \begin{cases} \dfrac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx = (\alpha-1)!$

For $\alpha = n+2$ & $\lambda = 2$

$$f_X(x) = \frac{2^{n+2} x^{n+1} e^{-2x}}{\Gamma(n+2)} = \frac{2(2x)^{n+1} e^{-2x}}{(n+1)!}$$

$$f_X(\lambda) = \begin{cases} \dfrac{2(2\lambda)^{n+1} e^{-2\lambda}}{(n+1)!} & \to x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Integral of density function is 1

$$\therefore \int \frac{2(2\lambda)^{n+1} e^{-2\lambda}}{(n+1)!} d\lambda = 1$$

$$\frac{2^{n+2}}{(n+1)!} \int \lambda^{n+1} e^{-2\lambda} d\lambda = 1$$

$$\therefore P(X=n) = \frac{1}{n!} \int_0^{\infty} \lambda^{n+1} e^{-2\lambda} d\lambda \quad (\text{found before})$$

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

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

Ans: Probability that a randomly chosen policy holder has exactly n accidents next year is $(n+1)/2^{n+2}$

Exercise 5

Number of people who visit a yoga studio each day → Poisson random variable → mean λ.

female → p ; male → $1-p$.

n women, m men visit

Total $m+n$ people.

$$\underbrace{\frac{e^{-\lambda}\lambda^{m+n}}{(m+n)!}}_{\text{Poisson, Probability that total no. of people who visited} = m+n} \cdot \underbrace{{}^{m+n}C_n\, p^n (1-p)^m}_{\text{Probability that out of } m+n \text{ people, } n \text{ are female and rest } m \text{ are male.}}$$

Ans Joint probability that exactly n women and m men visit the academy today.

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

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

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

Exercise 6

- $\text{Cov}(aX_1 + b, cX_2 + b) = ac\,\text{Cov}(X_1, X_2)$

$\text{Cov}(X_1, X_2) = E((X_1 - E(X_1))(X_2 - E(X_2)))$

$\therefore \text{Cov}(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$

$\therefore \text{Cov}(aX_1 + b, cX_2 + b) = E((aX_1 + b)(cX_2 + b)) - E(aX_1 + b)\,E(cX_2 + b)$

$= E(acX_1X_2 + abX_1 + bcX_2 + b^2) - E(aX_1 + b)\,E(cX_2 + b)$

$= E(acX_1X_2) + ab\,E(X_1) + bc\,E(X_2) + E(b^2) - (aE(X_1) + E(b))(cE(X_2) + E(b))$

$= ac\,E(X_1X_2) + ab\,E(X_1) + bc\,E(X_2) + E(b^2) - (ac\,E(X_1)E(X_2) + a\,E(b)E(X_1) + c\,E(b)E(X_2) + E(b)E(b))$

Since b is a constant, expectation value of b and b^2 is b and b^2 respectively

i.e $E(b) = b$, $E(b^2) = b^2$)

$= ac\,E(X_1X_2) + \cancel{ab\,E(X_1)} + \cancel{bc\,E(X_2)} + \cancel{b^2} - ac\,E(X_1)E(X_2) - \cancel{ab\,E(X_1)} - \cancel{bc\,E(X_2)} - \cancel{b^2}$

$= ac\,(E(X_1X_2) - E(X_1)E(X_2))$

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

~~$\therefore \text{Cov}(X_1, X$~~

$\therefore \boxed{\text{Cov}(aX_1 + b, cX_2 + b) = ac\,\text{Cov}(X_1, X_2)}$

<u>LHS = RHS</u> <u>Hence Proved</u>

- $Cov(X_1 + X_2, X_3) = Cov(X_1, X_3) + Cov(X_2, X_3)$

LHS:- $Cov(X_1 + X_2, X_3)$

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

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

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

$= \underbrace{E(X_1 X_3) - E(X_1)E(X_3)}_{Cov(X_1, X_3)} + \underbrace{E(X_2 X_3) - E(X_2)E(X_3)}_{Cov(X_2, X_3)}$

$= Cov(X_1, X_3) + Cov(X_2, X_3)$

$=$ R.H.S

$\therefore$ L.H.S $=$ R.H.S   <u>Hence Proved</u>

Exercise 7

n RVs $\rightarrow X_1, X_2, \cdots X_n$ each $X_i \rightarrow Ber(p)$

so each X_i takes binary values i.e $\{0, 1\}$

and $P(X_i = 1) = p$ and $P(X_i = 0) = 1-p$

$$X = \sum_{i=1}^{n} X_i$$

PMF: $P(X = k) \rightarrow P(\sum_{i=1}^{n} X_i = k) \rightarrow$ which means out of n $X_1, X_2, \cdots X_n$, k of them have to be 1 and rest $n-k$ are 0.

So $P(X=k) = \underbrace{{}^nC_k}_{\text{choosing } k \text{ out of } n\ X_is \text{ which would be } 1} \underbrace{p^k}_{\text{probability of } k\ X_is \text{ to be } 1} \underbrace{(1-p)^{n-k}}_{\text{probability of rest } X_is \text{ to be } 0}$

$$\therefore \boxed{P(X = k) = {}^nC_k\, p^k (1-p)^{n-k}}$$

$\therefore$ This is a Binomial distribution & $X \sim Bin(n, p)$

Hence Proved

Exercise 8

- $n = 100$ iid samples
  outcome lies in $[0, 1]$
  estimate of mean $\hat{\mu} = 0.45$
  True mean $\rightarrow \mu$
  let the interval around $\hat{\mu}$ be $(\hat{\mu}-\varepsilon, \hat{\mu}+\varepsilon)$
  True mean lies in this interval with probability at least 0.95

$$P(|\hat{\mu}-\mu| \le \varepsilon) \ge 0.95$$

$$\therefore P(|\hat{\mu}-\mu| > \varepsilon) \le 0.05$$

Using $P(|\hat{\mu}-\mu| > \varepsilon) \le 2\exp(-n\varepsilon^2)$

we get $2e^{-n\varepsilon^2} = 0.05$

$$\frac{2}{0.05} = e^{n\varepsilon^2}$$

$$40 = e^{n\varepsilon^2}$$

$$\ln(40) = n\varepsilon^2$$

$$\varepsilon^2 = \frac{\ln(40)}{100} = 0.0368888$$

$$\varepsilon = 0.1920645582639$$

Confidence interval $\rightarrow [\hat{\mu}-\varepsilon, \hat{\mu}+\varepsilon]$

$$= \boxed{[0.258, 0.642]}$$

<u>Ans</u> The confidence interval is $[0.258, 0.642]$

- Confidence interval shrinks by half.
  New confidence interval $\rightarrow \{\hat{\mu}-\varepsilon/2, \hat{\mu}+\varepsilon/2\}$
  New total number of samples $= n'$

(found in 1st part) $40 = e^{n\varepsilon^2} = e^{n'(\varepsilon/2)^2}$

$$\therefore n\varepsilon^2 = n'\varepsilon^2/4 \Rightarrow n' = 4n$$

$$n' = 4 \times 100 = 400$$

So to halve the confidence interval, we need $\boxed{\text{total 400 samples}}$

<u>Ans</u> We need $\boxed{\text{300 more samples}}$ (400-100). As there were $\boxed{\text{already 100 samples}}$, and $\boxed{\text{total samples required is 400}}$