MM 225 – AI and Data Science

Day 13: Joint Random Variable 1

Instructors: Hina Gokhale, MP Gururajan, N. Vishwanathan

27 AUGUST 2024

Joint Random Variables

Some examples

- X = person with lung cancer
- Y = person is a smoker

- X = Fracture Toughness of an alloy
- Y = Fatigue life of the alloy

- X = Height of an adult male
- Y = country of his residence

All the above random variables X and Y are related to each other.

- They vary together in some sense

These are known as **Joint Random** Variables

Example : Discrete case

A product is classified according to the number of defects it contains and the factory that produces it. Let X and Y be random variables.

- X = number of defects (= 0, 1, 2, 3,…)
- Y = factory where product is produced (say, 1 and 2)

Probability mass function is given by

X \ Y	**1**	**2**
0	1/8	1/16
1	1/16	1/16
2	3/16	1/8
3	1/8	1/4

Example: Continuous case

The joint density function for X and Y is given by

$$f(x, y) = \begin{cases} 2\ for\ 0 < x < y\ and\ 0 < y < 1 \\ 0\ otherwise \end{cases}$$

Joint Random Variables

(X, Y) is a joint RV, then CDF of (X, Y) is defined as

- $F_{X,Y}(x, y) = P[X \leq x, Y \leq y]$

Marginal CDF of X and Y are defined as

- $F_X(x) = F_{X,Y}(x, \infty)$
- $F_Y(y) = F_{X,Y}(\infty, y)$

Joint Discrete RV

(X, Y) is discrete then pmf of (X, Y) is

- $f_{X,Y}(x_i, y_j) = P[X = x_i,\ Y = y_j], i = 1, 2, 3, \ldots and\ j = 1, 2, 3, \ldots$

Hence,

- CDF $F_{X,Y}(a, b) = \sum_{x_i \leq a} \sum_{y_j \leq b} f_{X,Y}(x_i, y_j),\ i = 1, 2, 3, \ldots and\ j = 1, 2, 3, \ldots$

Joint continuous RV

pdf of (X, Y) is $f_{X,Y}(x, y)$ is such that

- $P[(X, Y) \in C] = \iint_C \ f_{X,Y}(x, y)dxdy$
- $P[X \in A, Y \ \in B] = \int_B \ \int_A \ f_{X,Y}(x, y)dxdy$

CDF can be given by

- $F_{X,Y}(a, b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f_{X,Y}(x, y)dxdy$

Marginal Distribution

DISCRETE RV

Marginal pmf

$$f_X(x_i) = \sum_{j=1}^{\infty} f_{X,Y}(x_i, y_j)$$

Conditional pmf:

$$f_{X|Y}(x_i|y_j) = \frac{f_{X,Y}(x_i, y_j)}{f_Y(y_j)}$$

CONTINUOUS RV

Marginal pdf

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

Conditional pdf

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Example : Discrete case

A product is classified according to the number of defects it contains and the factory that produces it. Let X and Y be random variables.

- X = number of defects (= 0, 1, 2, 3,…)
- Y = factory where product is produced (say, 1 and 2)

Probability mass function is given by

X \ Y	1	2
0	1/8	1/16
1	1/16	1/16
2	3/16	1/8
3	1/8	1/4

Find marginal pmf

Find E(X), E(Y), Var(X), Var(Y)

Cov(X,Y)

Solution

X \ Y	1	2	$f(X = x)$
0	1/8	1/16	1/8 +1/16 = **(3/16)**
1	1/16	1/16	1/16 +1/16 = **(2/16)**
2	3/16	1/8	3/16 + 1/8 **= (5/16)**
3	1/8	1/4	1/8 + 1/4 = **(3/8)**
$f(Y = y)$	**1/2**	**1/2**	**1**

Why is it called "Marginal"?

Expected Values

DISCRETE RV

$$E(X) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i f_{X,Y}(x_i, y_j)$$

$$E(XY) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j f_{X,Y}(x_i, y_j)$$

CONTINUOUS RV

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

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

Covariance and Correlation coefficient

Covariance between X and Y are defined as

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

Correlation Coefficient is given by

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

RV X and Y are independent iff

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

X \ Y	1	2	$f(X=x)$	E(X)	$E(X^2)$
0	1/8	1/16	1/8 +1/16 = (3/16)	= 0*3/16 = 0	= 0*(3/16)
1	1/16	1/16	1/16 +1/16 = (2/16)	= 1* (2/16) =2/16	=1*(2/16)
2	3/16	1/8	3/16 + 1/8 = (5/16)	= 2*(5/16) = 10/16	=4*(10/16)
3	1/8	1/4	1/8 + 1/4 = (3/8)	= 3*(3/8) = 9/8	=9*(9/8)
$f(Y=y)$	1/2	1/2	**1**	**=30/16**	**=76/16**

$$\mathbf{Var(X)} = \mathbf{E}\left(X^2\right) - E(X)^2 = \left(\frac{76}{16}\right) - \left(\frac{900}{256}\right) \approx 1.23$$

X \ Y	1	2	$f(X = x)$
0	1/8	1/16	1/8 +1/16 = (3/16)
1	1/16	1/16	1/16 +1/16 = (2/16)
2	3/16	1/8	3/16 + 1/8 = (5/16)
3	1/8	1/4	1/8 + 1/4 = (3/8)
$f(Y = y)$	**1/2**	**1/2**	**1**
E(Y)	**=1*(1/2)**	**=2*(1/2)**	**=3/2**
$E(Y^2)$	**=1*(1/2)**	**=4*(1/2)**	**=5/2**

$$\mathrm{Var(Y)} = \mathrm{E}(Y^2) - E(Y)^2 = \left(\frac{5}{2}\right) - \left(\frac{9}{4}\right) = 0.25$$

X \ Y	1	2
0	1/8	1/16
1	1/16	1/16
2	3/16	1/8
3	1/8	1/4

E(XY) = 0*(1/8) + 0*(1/16) + 1*(1/16) +2*(1/16) +2*(3/16)+4*(1/8) + 3*(1/8) + 6*(1/4) = 47/16

Cov(X,Y) = E(XY) – E(X)E(Y) = $\frac{47}{16} - \left(\frac{30}{16}\right)\left(\frac{3}{2}\right) = \frac{4}{32} = 0.125$

Corr(X,Y) = $\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{0.125}{\sqrt{1.23 * 0.25}} = 0.225$

Example : Continuous case

The joint density function for X and Y is given by

$$f(x,y) = \begin{cases} xe^{-(x+y)} & x > 0, y > 0 \\ 0 \ otherwise \end{cases}$$

a. Compute density of X
b. Compute density of Y
c. Are they independent?

Solution:

a. $f(x) = \int_{y=0}^{\infty} xe^{-(x+y)}dy = xe^{-x}, for\ x > 0$

b. $f(y) = \int_{x=0}^{\infty} xe^{-(x+y)}dx = \ e^{-y}, for\ y > 0$

c. f(x,y) = f(x)*f(y) ...they are independent

Example: Continuous case

The joint density function for X and Y is given by

$$f(x, y) = \begin{cases} 2\ for\ 0 < x < y\ and\ 0 < y < 1 \\ 0\ otherwise \end{cases}$$

a. Compute density of X
b. Compute density of Y
c. Are they independent?

Solution:

a. $f(x) = \int_{y=x}^{1} 2\,dy = 2(1 - x), for\ 0 < x < 1$

b. $f(y) = \int_{x=0}^{y} 2dx = 2y, for\ 0 < y < 1$

c. f(x,y) ≠ f(x)*f(y) …they are NOT independent

Thank you…