# STATISTICS 2

NOTES BY
GAGNEET
KAUR

September 10, 2021

# WEEK 1

# Multiple Random Variables

## SECTION 1 : 2 random variables
Joint, Marginal, Conditional PMFs

## JOINT PMF

Suppose $X$ and $Y$ are discrete random variables defined in the same probability space. Let the range of $X$ and $Y$ be $T_X$ and $T_Y$, resp. The Joint PMF of $X$ and $Y$, denoted $f_{XY}$, is a function from $T_X \times T_Y$ to $[0,1]$ defined as

$$\boxed{f_{XY}(t_1, t_2) = P(X = t_1 \text{ and } Y = t_2) \quad , t_1 \in T_X, \ t_2 \in T_Y}$$

- Joint PMF is usually written as table or a matrix.
- $P(X = t_1$ and $Y = t_2)$ is denoted $P(X = t_1, Y = t_2)$

**EXAMPLES** : Toss a fair coin twice

(1) Let $X_i = 1$ if i-th toss is heads and $X_i = 0$ if i-th toss is tails, $i = 1, 2$.

$\rightarrow$ $f_{X_1X_2}(0,0) = P(X_1=0, X_2=0) = 1/2 \cdot 1/2 = 1/4$

$f_{X_1X_2}(0,1) = P(X_1=0, X_2=1) = 1/2 \cdot 1/2 = 1/4$

| $t_2$ \ $t_1$ | 0 | 1 |
|---|---|---|
| 0 | 1/4 | 1/4 |
| 1 | 1/4 | 1/4 |

~ Joint PMF ~

# Each entry : b/w 0 and 1
# Sum of all entries = 1

---

(2) Random 2-digit number
X = units place , Y = number modulo 4

$f_{XY}(0,0) = P(X=0, Y=0)$
= P(number ends in 0 and multiple of 4)
$= P(\{00, 20, 40, 60, 80\}) = 5/100 = 1/20$

$f_{XY}(1,0) = P(X=1, Y=0)$
= P(no. ends in 1 and multiple of 4)
= 0

$f_{XY}(4,2) = P(X=4, Y=2)$
= P(no. ends in 4 and remainder as 2)
$= P(\{14, 34, 54, 74, 94\}) = 5/100 = 1/20$

---

# MARGINAL PMF

Suppose X and Y are jointly distributed discrete random variables with joint PMF $f_{XY}$. The PMF of the individual random variables X and Y are called as marginal PMFs. It can be shown that

$$f_X(t) = P(X=t) = \sum_{t' \in T_Y} f_{XY}(t,t')$$

$$f_Y(t) = P(Y=t) = \sum_{t' \in T_X} f_{XY}(t',t)$$

where $T_X$ and $T_Y$ are the ranges of X and Y, resp.

→ Note that the marginal PMF is simply a PMF.

---

EXAMPLE : Toss a fair Coin Twice

| $t_2 \backslash t_1$ | 0 | 1 | $f_{X_2}(t_2)$ |
|---|---|---|---|
| 0 | 1/4 | 1/4 | 1/2 |
| 1 | 1/4 | 1/4 | 1/2 |
| $f_{X_1}(t_1)$ | 1/2 | 1/2 | |

- Marginal PMF of $X_1$
  - → $f_{X_1}(0) = f_{X_1X_2}(0,0) + f_{X_1X_2}(0,1)$
  - → $f_{X_1}(1) = f_{X_1X_2}(1,0) + f_{X_1X_2}(1,1)$
- Marginal PMF of $X_2$
  - → $f_{X_2}(0) = f_{X_1X_2}(0,0) + f_{X_1X_2}(1,0)$
  - → $f_{X_2}(1) = f_{X_1X_2}(0,1) + f_{X_1X_2}(1,1)$

---

# CONDITIONAL DISTRIBUTION OF A RANDOM VARIABLE GIVEN AN EVENT

Suppose $X$ is a discrete random variable with range $T_X$, and $A$ is an event in the same probability space. The conditional PMF of $X$ given $A$ is defined as the PMF

$$Q(t) = P(X=t \mid A) \quad , \quad t \in T_X$$

We will use the notation $f_{X|A}(t)$ for the above conditional PMF, and $(X|A)$ to denote the conditional random variable.

$$f_{X|A}(t) = \frac{P((X=t) \cap A)}{P(A)}$$

★ Important : Range of $(X|A)$ can be different from $T_X$ and will depend on $A$.

---

# CONDITIONAL DISTRIBUTION OF ONE R.V. GIVEN ANOTHER R.V.

Suppose $X$ and $Y$ are jointly distributed discrete random variables with joint PMF $f_{XY}$. The conditional PMF of $Y$ given $X=t$ is defined as the PMF

$$Q(t') = P(Y=t' \mid X=t) = \frac{P(Y=t', X=t)}{P(X=t)} = \frac{f_{XY}(t,t')}{f_X(t)}$$

$$\boxed{\sum_{t' \in T_Y} f_{Y|X=t}(t') = 1}$$

We will use the notation $f_{Y|X=t}(t')$ for the above conditional PMF, and $(Y|X=t)$ to denote the 'conditional' random variable.

\# $\underbrace{f_{XY}(t,t')}_{\text{Joint PMF}} = \underbrace{f_{Y|X=t}(t')}_{\text{Conditional PMF}} \cdot \underbrace{f_X(t)}_{\text{Marginal PMF}} = f_{X|Y=t'}(t) \cdot f_Y(t')$

## SOME APPLICATIONS

(1) Throw a die and toss a coin as many times as the number shown on die. Let X be the no. shown on die. Let Y be the no. of heads. What is the joint PMF of X and Y?

Solution: $X \sim \text{Uniform}(\{1,2,3,4,5,6\})$

$Y|X=t \sim \text{Binomial}(t, 1/2)$

$\therefore \quad f_X(t) = \frac{1}{6}$

and $f_{Y|X=t}(t') = {}^tC_{t'} \left(\frac{1}{2}\right)^{t'} \left(\frac{1}{2}\right)^{t-t'} = {}^tC_{t'} \left(\frac{1}{2}\right)^{t}$

Thus, $\underbrace{f_{XY}(t,t')}_{\text{[Joint PMF]}} = f_X(t) \cdot f_{Y|X=t}(t') = \frac{1}{6} \cdot {}^tC_{t'} \left(\frac{1}{2}\right)^{t}$

$\text{Range}(Y|X=t) = \{0, 1, \ldots, t\}$

classmate

Date
Page

(2) Let $N \sim$ Poisson $(\lambda)$. Given $N = n$, toss a fair coin $n$ times and denote the no. of heads obtained by $X$. What is the distribution of $X$?

Solution : $X \rightarrow$ no. of heads obtained

$$f_N(n) = e^{-\lambda} \frac{\lambda^n}{n!} \quad , \quad n = 0, 1, 2, \ldots$$

$$(X \mid N = n) \sim \text{Binomial}\left(n, 1/2\right)$$

$$f_{X|N=n}(k) = {}^nC_k \left(\frac{1}{2}\right)^n \quad , \quad k = 0, 1, 2, \ldots$$

Now,

$$f_{NX}(n, k) = e^{-\lambda} \frac{\lambda^n}{n!} \cdot {}^nC_k \left(\frac{1}{2}\right)^n$$

$$\therefore \quad f_X(k) = \sum_{n=k}^{\infty} f_{NX}(n, k)$$

$$= \sum_{n=k}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \, {}^nC_k \left(\frac{1}{2}\right)^n$$

$$= \sum_{n=k}^{\infty} \frac{e^{-\lambda} \lambda^n}{k! \, (n-k)!} \left(\frac{1}{2}\right)^n$$

$$= \sum_{n=k}^{\infty} \frac{\lambda^{n-k}}{(n-k)! \, 2^{n-k}} \times \frac{e^{-\lambda} \lambda^k}{k! \, 2^k}$$

$$\Rightarrow \quad f_X(k) = e^{\lambda/2} \times \frac{e^{-\lambda} \lambda^k}{k! \, 2^k}$$

$$\Rightarrow \quad f_X(k) = \frac{e^{-\lambda/2}}{k!} \left(\frac{\lambda}{2}\right)^k \quad \Rightarrow \quad X \sim \text{Poisson}\left(\frac{\lambda}{2}\right)$$

Date September 11, 2021

# SECTION 2 : More than 2 discrete R.V

## JOINT PMF

Suppose $X_1, X_2, \ldots, X_n$ are discrete random variables defined in the same probability space. Let the range of $X_i$ be $T_{X_i}$. The joint PMF of $X_i$, denoted $f_{X_1 \ldots X_n}$, is a $f^n$ from $T_{X_1} \times T_{X_2} \times \ldots \times T_{X_n}$ to $[0,1]$ defined as

$$\boxed{f_{X_1 \ldots X_n}(t_1, t_2, \ldots, t_n) = P(X_1 = t_1 \text{ and } \ldots \text{ and } X_n = t_n), \; t_i \in T_{X_i}}$$

→ $P(X_1 = t_1 \text{ and } \ldots \text{ and } X_n = t_n)$ is denoted by $P(X = t_1, \ldots, X_n = t_n)$

## MARGINAL PMF

Suppose $X_1, X_2, \ldots, X_n$ are jointly distributed random variables with joint PMF $f_{X_1 \ldots X_n}$. The PMF of the individual random variables $X_1, X_2, \ldots, X_n$ are called as marginal PMFs. It can be shown that

$$f_{X_1}(t) = P(X_1 = t) = \sum_{t_2' \in T_{X_2}, t_3' \in T_{X_3}, \ldots, t_n' \in T_{X_n}} f(t, t_2', t_3', \ldots t_n').$$

$\vdots$

$$f_{X_n}(t) = P(X_n = t) = \sum_{t_1' \in T_{X_1}, t_2' \in T_{X_2}, \ldots, t_{n-1}' \in T_{X_{n-1}}} f(t_1', t_2' \ldots, t_{n-1}', t)$$

where $T_{X_i}$ is the range of $X_i$.

- Marginalisation

Suppose $X_1, X_2, X_3 \sim f_{X_1X_2X_3}$ and $X_i \in T_{X_i}$

We have discussed the marginal PMF $f_{X_i}$ of the individual random variables $X_1, X_2, X_3$.

What about $f_{X_1X_2}$, the joint PMF of $X_1$ & $X_2$?
What about $f_{X_1X_3}$, $f_{X_2X_3}$?

$$f_{X_1X_2}(t_1, t_2) = P(X_1 = t_1, X_2 = t_2) = \sum_{t_3' \in T_{X_3}} f_{X_1X_2X_3}(t_1, t_2, t_3')$$

$$f_{X_1X_3}(t_1, t_3) = P(X_1 = t_1, X_3 = t_3) = \sum_{t_2' \in T_{X_2}} f_{X_1X_2X_3}(t_1, t_2', t_3)$$

$$f_{X_2X_3}(t_2, t_3) = P(X_2 = t_2, X_3 = t_3) = \sum_{t_1' \in T_{X_1}} f_{X_1X_2X_3}(t_1', t_2, t_3)$$

# Marginalisation : Sum over everything you do not want.

---

# CONDITIONING WITH MUTIPLE R.V.

Suppose $X_1, X_2, X_3, X_4 \sim f_{X_1X_2X_3X_4}$ and $X_i \in T_{X_i}$

A wide variety of conditioning is possible when there are many random variables.

$$(X_1 \mid X_2 = t_2) \sim f_{X_1 \mid X_2 = t_2}(t_1) = \frac{f_{X_1 X_2}(t_1, t_2)}{f_{X_2}(t_2)}$$

$$(X_1, X_2 \mid X_3 = t_3) \sim f_{X_1 X_2 \mid X_3 = t_3}(t_1, t_2) = \frac{f_{X_1 X_2 X_3}(t_1, t_2, t_3)}{f_{X_3}(t_3)}$$

$$(X_1 \mid X_2 = t_2, X_3 = t_3) \sim f_{X_1 \mid X_2 = t_2, X_3 = t_3}(t_1) = \frac{f_{X_1 X_2 X_3}(t_1, t_2, t_3)}{f_{X_2 X_3}(t_2, t_3)}$$

→ Conditioning and factors of Joint PMF

suppose $X_1, X_2, X_3, X_4 \sim f_{X_1 X_2 X_3 X_4}$ and $X_i \in T_{X_i}$

$$f_{X_1 \dots X_4}(t_1, \dots, t_4) = P(X_1 = t_1, X_2 = t_2, X_3 = t_3, X_4 = t_4)$$

In factor form, it is written as,

$$\boxed{\underbrace{f_{X_1 \dots X_4}(t_1, \dots, t_4)}_{\text{Joint PMF}} = f_{X_1 \mid X_2 = t_2, X_3 = t_3, X_4 = t_4}(t_1) \cdot f_{X_2 \mid X_3 = t_3, X_4 = t_4}(t_2) \cdot f_{X_3 \mid X_4 = t_4}(t_3) \cdot f_{X_4}(t_4)}$$

→ Factoring can be done in any sequence

$$\therefore f_{X_1 \dots X_4}(t_1, \dots, t_4) = f_{X_4 \mid X_3 = t_3, X_2 = t_2, X_1 = t_1}(t_4) \cdot f_{X_3 \mid X_2 = t_2, X_1 = t_1}(t_3) \cdot f_{X_2 \mid X_1 = t_1}(t_2) \cdot f_{X_1}(t_1)$$

.... etc.