BSCMA1004

STATISTICS II NOTES

WEEK 1 NOTES

IITM B.S Degree

PREPARED BY

Vehaan Handa, IIT Madras

Statistics II → Week I Notes

★ Multiple Random Variables

Eg. Toss a coin thrice.

A fair coin is tossed thrice. Naturally there can be 3 random variables.
Let $X_i = 1$ if i^{th} toss is heads
$X_i = 0$ if i^{th} toss is tails. $i = 1, 2, 3$.

i.e Repeated trials (Bernoulli).

Together, the 3 random variables completely specify outcome of the experiment.

Event $X_1 = 1$ is independent of $X_2 = 1$ and $X_3 = 1$

Eg. A 2 digit no. from 00 to **99** is selected at random. Partial info available about the number as 2 random variables. X be digit in units place. Y be remainder obtained when num ÷ 4.

$$X \in \text{Uniform}(\{0,1,2,3,4,5,6,7,8,9\})$$
$$Y \in \text{Uniform}(\{0,1,2,3\})$$

Suppose $X = 1$ occurred. What abt event $Y = 0$?
Not possible.

When 2 random variables are defined in the same probability space, value of one can influence the value of the other.

Eg. X = runs in over   Y = wickets in over

Y = 0, Y = 1, Y = 2

If Y = 0, X takes larger values (expected) than when Y = 1
If Y = 2, X is expected to take lower values.

In complex experiments, such relationships b/w random variables are useful in modelling.

★ Joint, Marginal and Conditional PMFS : 2 RV's

① Joint PMF

Suppose X and Y are drv's defined in same probability space. Let the range of X and Y be T_X and T_Y. Joint PMF of X and Y, f_{XY}, is a function from $T_X \times T_Y$ to $[0,1]$ defined as :→

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

Usually written as a table/matrix.
$P(X = t_1 \text{ and } Y = t_2)$ is denoted $P(X = t_1, Y = t_2)$

Eg. Let $X_i = 1$, i^{th} toss heads, $X_i = 0$, i^{th} toss tails $i = 1, 2$.
$i = 1, 2$ $f_{X_1X_2}(0,0) = P(X_1 = 0, X_2 = 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

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

$f_{X_1X_2}(t_1, t_2)$

$t_2 \backslash t_1$ →	0	1
0	1/4	1/4
1	1/4	1/4

- Each entry $\in [0,1]$
- Σentry $= 1$

DATE

② Marginal PMFs

Suppose X, Y are jointly distributed drv's with joint PMF f_{XY}. PMF of individual random variables X and Y are called marginal PMFs.

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

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

where T_X, T_Y are ranges of X, Y resp.

Proof: Suppose $T_Y = \{y_1, y_2, y_3, \dots, y_k\}$

$(X=t) = (X=t \text{ and } Y=y_1) \text{ or } \dots \text{ or } (X=t \text{ and } Y=y_k)$

$P(X=t) = P(X=t, Y=y_1) + \dots + P(X=t, Y=y_k)$

Marginal PMF is simply a PMF.

The above 2 box eqns are called marginalisation equations.

Eg. Table of $f_{X_1X_2}(t_1,t_2)$

$t_2 \backslash t_1$	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	

DATE

Marginal PMF of X_1: → add over columns of joint table.
Marginal PMF of X_2: → add over rows of joint table.

$$f_{X_1}(0) = f_{X_1X_2}(0,0) + f_{X_1X_2}(0,1)$$
$$f_{X_1}(1) = f_{X_1X_2}(1,0) + f_{X_1X_2}(1,1)$$

$$f_{X_2}(0) = f_{X_1X_2}(0,0) + f_{X_1X_2}(1,0)$$
$$f_{X_2}(1) = f_{X_1X_2}(0,1) + f_{X_1X_2}(1,1)$$

★ Same marginal PMF from different joint PMFs

Case 1

$t_2 \backslash t_1$	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	

Ppl wrongly assume that joint PMF = product of marginal PMFs
Which is a valid answer, but not the only joint PMFs.

Case 2

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

Joint $\xrightarrow{\text{Unique}}$ Marginal
Marginal $\xrightarrow[\text{Unique}]{\text{Not}}$ Joint

For every x b/w 0 and 1/2, we get a joint PMF resulting in same marginal.

DATE

★ Conditional Distribution of Random Variable given an event

Suppose X is a drv with range T_X, A is an event in the same probability space. Conditional PMF of X given A is defined as the PMF:

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

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

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

Here range of $(X|A)$ can be different from T_X, will depend on A.

★ Conditional Distribution of One Random Variable given another

Suppose X and Y are jointly distributed drv's with joint PMFs f_{XY}, the conditional PMF of Y given $X = t$ is defined as :→

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

$$= \frac{f_{XY}(t, t')}{f_X(t)}$$

$f_{Y|X=t}(t')$ for PMF.

$(Y|X=t)$ for "conditional" rv.

DATE

$$\boxed{f_{XY}(t,t') = f_{Y|X=t}(t')\, f_X(t)}$$

Range of $(Y|X=t)$ can be diff. from T_Y, and will depend on t.

Eg. Joint PMF $f_{XY}(t_1,t_2)$

$t_2 \backslash t_1$	0	1	2	$f_Y(t_2)$
0	1/4	1/8	1/8	1/2
1	1/8	1/8	1/4	1/2
$f_X(t_1)$	3/8	1/4	3/8	

T_X i.e. $X \in \{0,1,2\}$. $Y \in \{0,1\}$

$Y|X=0 \in \{0,1\}$ $\quad f_{Y|X=0}(0) = \dfrac{f_{XY}(0,0)}{f_X(0)} = \dfrac{1/4}{3/8} = \boxed{2/3}$

$$f_{Y|X=0}(1) = \frac{f_{XY}(0,1)}{f_X(0)} = \frac{1/8}{3/8} = \boxed{\frac{1}{3}}$$

Or just divide the zero column with the number below it.

$X|Y=1 \in \{0,1,2\}$ We get $\dfrac{1/8}{1/2}, \dfrac{1/8}{1/2}, \dfrac{1/4}{1/2}$

$$= \boxed{1/4}, \boxed{1/4}, \boxed{1/2}$$

Check $\boxed{\begin{aligned} f_{XY}(t_1,t_2) &= f_{Y|X=t_1}(t_2)\, f_X(t_1) \\ &= f_{X|Y=t_2}(t_1) . f_Y(t_2) \end{aligned}}$ $\quad * \displaystyle\sum_{t' \in T_Y} f_{Y|X=t}(t') = 1$

DATE

(1) Throw a die & toss a coin as many times as number shown on die. Let Y be number of heads. Joint PMF of X and Y?

$X \sim \text{Uniform}(1,2,3,4,5,6) \quad f_X(t) = \frac{1}{6}, \; 1 \leq t \leq 6$

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

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

$$= {}^{t}C_{t'} \left(\frac{1}{2}\right)^{t}, \quad t' = 0, 1, 2 \ldots, t$$

$$f_{XY}(t, t') = f_X(t) f_{Y|X=t}(t') = \frac{1}{6} \, {}^{t}C_{t'} \left(\frac{1}{2}\right)^{t}$$

$t' \backslash t$	1	2	3	4	5	6
0	$1/6 \cdot 1/2$	$1/6(1/2)^2$				
1	$1/6 \cdot 1/2$	$1/6 \cdot 2 \cdot \left(\frac{1}{2}\right)^2$				
2	0	$1/6\left(\frac{1}{2}\right)^2$				
3	0	0				
⋮	⋮	⋮				

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

DATE

② Poisson Number of Coin Tosses

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

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

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

Asked $f_X(x)$?

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

$$f_X(k) = \sum_{n=0}^{\infty} f_{NX}(n,k)$$ [we don't want N, so add up everything over N]

$$f_X(k) = \sum_{n=0}^{\infty} f_{NX}(n,k) = \sum_{n=0}^{\infty} e^{-\lambda}\frac{\lambda^n}{n!}\cdot {}^nC_k\left(\frac{1}{2}\right)^n$$

Got k heads.

i.e $n=k$

If $n<k$, whole exp. goes to 0.

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

$$= \frac{e^{-\lambda}}{k!}\frac{\lambda^k}{2^k}\underbrace{\sum_{n=k}^{\infty}\frac{\lambda^{n-k}}{(n-k)!\,2^{n-k}}}_{=\boxed{e^{\lambda/2}}}$$

So $f_X(k) = e^{-\lambda/2}\dfrac{(\lambda/2)^k}{k!}$ $\quad$ Interesting, so $X \sim \text{Poisson}(\lambda/2)$

classmate $\quad$ PAGE

DATE

③ IPL Powerplay

Let X = no. of runs in over, Y = no. of wickets in over.

Assume :→

$$Y \sim \overset{13/16}{\{0}, \overset{1/8}{1}, \overset{1/16}{2\}}$$

$X|Y=0 \sim \text{Uniform}\{6,7,8,9,10,11,12\}$

$X|Y=1 \sim \text{Uniform}\{2,3,4,5,6,7,8\}$

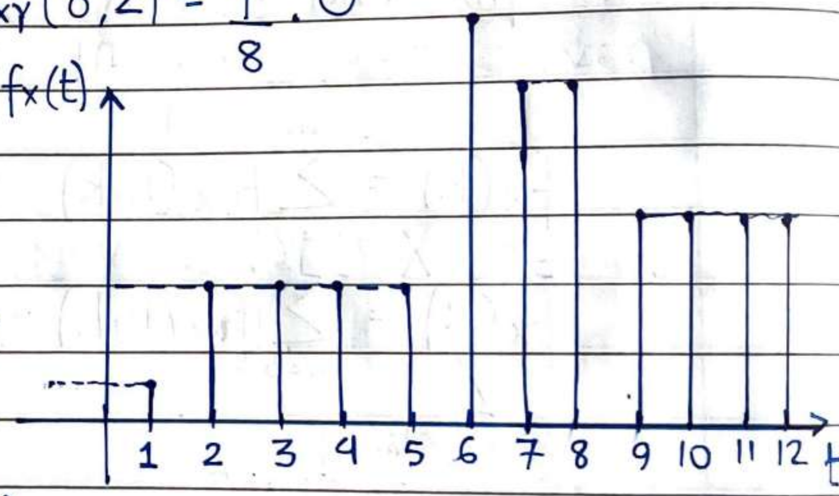
$X|Y=2 \sim \text{Uniform}\{0,1,2,3,4,5,6\}$

$$* \quad f_{XY}(t,t') = f_Y(t')\, f_{X|Y=t'}(t) \quad *$$

$f_{XY}(6,1) = \frac{1}{8} \times \frac{1}{7}$ $\qquad f_{XY}(8,2) = \frac{1}{8} \cdot 0$

$f_X(t) = ?$

$t = 0,1,2 \ldots, 12.$



$f_X(0) = f_{XY}(0,2) = 1/16 \times 1/7$

$= f_X(1)$

$f_X(2) = f_{XY}(2,2) + f_{XY}(2,1)$

$= \frac{3}{16} \times \frac{1}{7}$

$f_X(6) = f_{XY}(6,2) + f_{XY}(6,1) + f_{XY}(6,0) = 1/7$

DATE

★ PMF and conditioning on multiple RV's.

1. Joint PMF

Suppose $X_1, X_2, \ldots, X_n$ are drv's defined in the same probability space. Let the range of X_i be T_{X_i}. The joint PMF of X_i, denoted $f_{X_1 X_2 \ldots X_n}$ is a function from $T_{X_1} \times \ldots \times T_{X_n}$ to $[0,1]$ defined as :→

$$f_{X_1 \ldots X_n}(t_1, \ldots, t_n) = P(X_1 = t_1, X_2 = t_2, \ldots, X_n = t_n),\ t_i \in T_{X_i}$$

Written as a table.

Eg. Let $X_i = 1$, i^{th} toss heads, $X_i = 0$ if i^{th} toss tails. $i = 1, 2, 3$.

t_1	t_2	t_3	$f_{X_1 X_2 X_3}(t_1, t_2, t_3)$
0	0	0	1/8
0	0	1	1/8
0	1	0	1/8
0	1	1	1/8
1 ~~0~~	0	0	1/8
1	0	1	1/8
1	1	0	1/8
1	1	1	1/8

DATE

Eg. Random 3 digit no. 000 to 999.

$X = 1^{st}$ digit from left $\quad X = \{0, 1, \ldots, 9\}$

$Y =$ num modulo 2 $\quad Y = \{0, 1\}$

$Z = 1^{st}$ digit from right. $\quad Z = \{0, 1, 2, \ldots, 9\}$

$$f_{XYZ}(0,0,0) = P(\text{starts } 0, \text{ num even, ends with } 0)$$
$$= \frac{10}{1000} = \frac{1}{100}$$

$$f_{XYZ}(1,1,1) = \frac{1}{100} \qquad f_{XYZ}(1,0,1) = 0$$

$$f_{XYZ}(t_1, t_2, t_3) = \begin{cases} 0 & \text{if } t_2 = 0,\ t_3 \in \text{odd} \\ 0 & \text{if } t_2 = 1,\ t_3 \in \text{even} \\ 1/100 & \text{otherwise} \end{cases}$$

Eg. Suppose over has 6 deliveries. Let X_i be no. of runs scored in i^{th} delivery.

$$X_i \in \{0, 1, 2, \ldots, 8\} \qquad i = 1, 2, \ldots, 6$$

Very complicated possibilities.

DATE

2. Marginal PMF (Individual)

Suppose $X_1, X_2 \ldots, X_n$ are jointly distributed drv's with joint PMF $f_{X_1 \ldots X_n}$. PMF of individual rv's $X_1, X_2 \ldots, X_n$ are called as marginal PMFs. It can be shown :→

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

$$f_{X_2}(t) = P(X_2 = t) = \sum_{t_1' \in T_{X_1}, t_3' \in T_{X_3} \ldots t_n' \in T_{X_n}} f_{X_1 \ldots X_n}(t_1', t, t_3' \ldots, t_n')$$

$\vdots$

$f_{X_n}(t) = \ldots\ldots = \ldots\ldots$

where T_{X_i} is range of X_i.

Eg. Same 1^{st} eg of 3 coin tosses.

$$f_{X_1}(0) = f_{X_1X_2X_3}(0,0,0) + f_{X_1X_2X_3}(0,0,1) + f_{X_1X_2X_3}(0,1,0) + f_{X_1X_2X_3}(0,1,1)$$

$$= 1/8 + 1/8 + 1/8 + 1/8 = 1/2$$

$$f_{X_1}(1) = f_{X_1X_2X_3}(1,0,0) + f_{X_1X_2X_3}(1,0,1) + f_{X_1X_2X_3}(1,1,0) + f_{X_1X_2X_3}(1,1,1)$$

$$= 1/8 + 1/8 + 1/8 + 1/8 = 1/2$$

DATE

Eg. 2^{nd} eg. 000 to 999.

- $X \sim$ Uniform $\{0,1,2,3,4,5,6,7,8,9\}$
- $Y \sim$ Uniform $\{0,1\}$
- $Z \sim$ Uniform $\{0,1,2,3,4,5,6,7,8,9\}$

Eg. IPL Powerplay

$X_i \in \{0,1,2,3,4,5,6,7,8\}$. How to assign probabilities?

Ball 1 : 0 - 957 matches
1 - 429 matches
2 - 57 matches
3 - 5 matches (out of 1598)
4 - 138 matches
5 - 8 matches
6 - 4 matches

Assign probabilities in the same proportion as data.

★ Marginalisation

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

~~To~~ Marginal PMF for X_1, X_2, X_3 done ✓
What about $f_{X_1X_2}$? $f_{X_1X_3}$? $f_{X_2X_3}$?

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

and so on.....

Keep what you want, sum over everything you don't want.

Marginal PMF (General)

Suppose $X_1, X_2, \ldots\ldots, X_n$ are jointly distributed drv's with joint PMF $f_{X_1\ldots X_n}$. The joint PMF of rv's $X_{i_1}, X_{i_2}, \ldots\ldots, X_{i_k}$, denoted $f_{X_{i_1}\cdots X_{i_k}}$, given:→

$$f_{X_{i_1}\ldots X_{i_k}}(t_{i_1}, t_{i_2}\ldots, t_{i_k}) = \sum_{\substack{t_1\ldots t_{i_1}'-1,\\ t'_{i_1+1}\ldots\ldots t'_{i_k-1},\\ t'_{i_k+1}\ldots, t_n}} f_{X_1,\ldots,X_n}(t_1,\ldots, t_{i_1}'-1, t_{i_1}, t_{i_1+1}\ldots t_{i_k-1}, t_{i_k}, t_{i_k+1}\ldots t_n)$$

3. Conditioning with multiple drv's

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

$$(X_1 | X_2 = t_2) \sim f_{X_1|X_2=t_2}(t_1) = \frac{f_{X_1X_2}(t_1,t_2)}{f_{X_2}(t_2)} \begin{matrix}\rightsquigarrow \text{Joint} \\ \rightsquigarrow \text{Marginal}\end{matrix}$$

$$(X_1, X_2 | X_3 = t_3) \sim f_{X_1X_2|X_3=t_3}(t_1,t_2) = \frac{f_{X_1X_2X_3}(t_1,t_2,t_3)}{f_{X_3}(t_3)}$$

$$(X_1 | X_2 = t_2, X_3 = t_3) \sim f_{X_1|X_2=t_2, X_3=t_3}(t_1) = \frac{f_{X_1X_2X_3}(t_1,t_2,t_3)}{f_{X_2X_3}(t_2,t_3)}$$

DATE

t_1	t_2	t_3	t_4	$f_{X_1 \dots X_4}(t_1, \dots, t_4)$
0	0	0	0	1/12
0	0	0	1	1/12
0	0	1	1	1/12
0	0	2	0	1/12
0	1	1	0	1/12
0	1	1	1	1/12
0	1	2	0	1/12
1	0	0	1	1/12
1	0	2	0	1/12
1	0	2	1	1/12
1	1	0	1	1/12
1	1	1	0	1/12

$f_{X_1|X_2=0}$. Range:- $T_{X_1} = T_{X_2} = T_{X_4} = \{0,1\}$

$T_{X_3} = \{0,1,2\}$

$(X_1 | X_2 = 0) \sim \{0, 1\}$

$\frac{4/12}{7/12}$ $\frac{3/12}{7/12}$

$(X_1 | X_3 = 0, X_4 = 1) \sim \{0, 1\}$

$1/3$ $2/3$

$X_3 X_4 | X_1 = 0.$

$t_4 \backslash t_3$	0	1	2
0	1/7	1/7	2/7
1	1/7	2/7	0

classmate

DATE

★ Conditioning and Factors of Joint PMF

$X_1, X_2, \ldots, X_4 \sim f_{X_1X_2X_3X_4}, \; X_i \in T_{X_i}$

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

$$= P(X_1 = t_1 \mid (X_2 = t_2, X_3 = t_3, X_4 = t_4) \times P(X_2 = t_2, X_3 = t_3, X_4 = t_4)$$

$$= P(X_1 = t_1 \mid X_2 = t_2, X_3 = t_3, X_4 = t_4) \times P(X_2 = t_2 \mid X_3 = t_3, X_4 = t_4) \times P(X_3 = t_3, X_4 = t_4)$$

$$= P(X_1 = t_1 \mid X_2 = t_2, X_3 = t_3, X_4 = t_4) \, P(X_2 = t_2 \mid X_3 = t_3, X_4 = t_4) \, P(X_3 = t_3 \mid X_4 = t_4) \times P(X_4 = t_4)$$

Or read it in reverse: $\rightarrow f_{X_1\ldots X_4}(t_1 \ldots t_4) =$

$$= P(X_4 = t_4) \cdot P(X_3 = t_3 \mid X_4 = t_4) \cdot P(X_2 = t_2 \mid X_3 = t_3, X_4 = t_4) \cdot P(X_1 = t_1 \mid X_2 = t_2, X_3 = t_3, X_4 = t_4)$$