Week 3: Multiple Random Variables

Example: Toss a coin thrice

A fair coin is tossed thrice. Naturally, there can be three random variables.

Let $X_i = 1$ if i-th toss is heads and $X_i = 0$ if i-th toss is tails, $i = 1, 2, 3$.

Example: Toss a coin thrice

A fair coin is tossed thrice. Naturally, there can be three random variables.

Let $X_i = 1$ if i-th toss is heads and $X_i = 0$ if i-th toss is tails, $i = 1, 2, 3$.

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

The event $X_1 = 1$ is independent of $X_2 = 1$ and $X_3 = 1$.

Example: Random 2-digit number

A 2-digit number from 00 to 99 is selected at random. Partial information is available about the number as two random variables. Let X be the digit in units place. Let Y be the reminder obtained when the number is divided by 4.

$23 \Rightarrow X = 3$
$Y = 3$
$48 \Rightarrow X = 8$
$Y = 0$

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

Example: Random 2-digit number

A 2-digit number from 00 to 99 is selected at random. Partial information is available about the number as two random variables. Let X be the digit in units place. Let Y be the reminder obtained when the number is divided by 4.

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

Suppose the event $X = 1$ has occurred. What about the event $Y = 0$?

Example: Random 2-digit number

A 2-digit number from 00 to 99 is selected at random. Partial information is available about the number as two random variables. Let X be the digit in units place. Let Y be the reminder obtained when the number is divided by 4.

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

Suppose the event $X = 1$ has occurred. What about the event $Y = 0$?

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

Example: IPL powerplay over

Let $X =$ number of runs in the over. Let $Y =$ number of wickets in the over.

Consider the events: $Y = 0$, $Y = 1$, $Y = 2$

Example: IPL powerplay over

Let $X =$ number of runs in the over. Let $Y =$ number of wickets in the over.

Consider the events: $Y = 0$, $Y = 1$, $Y = 2$

Given $Y = 0$, we expect X to take larger values than when $Y = 1$.

Given $Y = 2$, we expect X to take significantly lower values.

Example: IPL powerplay over

Let X = number of runs in the over. Let Y = number of wickets in the over.

Consider the events: $Y = 0$, $Y = 1$, $Y = 2$

Given $Y = 0$, we expect X to take larger values than when $Y = 1$.

Given $Y = 2$, we expect X to take significantly lower values.

In complex experiments, such relationships between random variables are useful in modeling.

Section 1

Two random variables: Joint, marginal, conditional PMFs

Two discrete random variables: Joint PMF

Definition (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, respectively. The joint PMF of X and Y, denoted f_{XY}, is a function from $T_X \times T_Y$ to $[0, 1]$ defined as

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

- Joint PMF is usually written as a table or a matrix
- $P(X = t_1 \text{ and } Y = t_2)$ is denoted $\underline{P(X = t_1, Y = t_2)}$

Example: Toss a fair coin twice

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

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

Example: Toss a fair coin twice

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

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

Table 1: $f_{X_1X_2}(t_1, t_2)$

$t_2 \setminus t_1$	0	1
0	1/4	1/4
1	1/4	1/4

$\rightarrow f_{X_1X_2}(1,1)$

Joint PMF $\Rightarrow$ { - Each entry is between 0 + 1; - Sum of all entries equals 1 }

Example: Random 2-digit number

X = units place, Y = number modulo 4 remainder when divided by 4

- $f_{XY}(0,0) = P(X = 0 \text{ and } Y = 0)$

  $= P(\text{number ends in 0 and multiple of 4})$

  $= P(\{00, 20, 40, 60, 80\}) = 5/100 = 1/20$

- $f_{XY}(1,0) = P(\text{number}\overset{X=1}{\ }\text{ends in 1 and multi}\overset{Y=0}{\ }\text{ple of 4}) = 0$

- $f_{XY}(4,2) = P(\text{number}\overset{X=4}{\ }\text{ends in 4 and }\overset{Y=2}{\underline{2 \text{ mod } 4}})$
  $= P(\{14, 34, 54, 74, 94\}) = 5/100 = 1/20$

Example: Random 2-digit number

$X =$ units place, $Y =$ number modulo 4

- $f_{XY}(0,0) = P(X = 0 \text{ and } Y = 0)$

  $= P(\text{number ends in 0 and multiple of 4})$

  $= P(\{00, 20, 40, 60, 80\}) = 5/100 = 1/20$

- $f_{XY}(1,0) = P(\text{number ends in 1 and multiple of 4}) = 0$

- $f_{XY}(4,2) = P(\text{number ends in 2 and 2 mod 4})$
  $= P(\{14, 34, 54, 74, 94\}) = 5/100 = 1/20$

Table of $f_{XY}(t_1, t_2)$

$t_2 \setminus t_1$	0	1	2	3	4	5	6	7	8	9
0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0
1	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$
2	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0
3	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$

Two random variables: Marginal PMF

Definition (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

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

Marginalisation

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

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

- Proof
  - Suppose $T_Y = \{y_1, \ldots, y_K\}$
  - $(\underline{X = t}) = (X = t \text{ and } Y = y_1) \text{ or } \cdots \text{ or } (X = t \text{ and } Y = y_K)$
  - $P(X = t) = P(X = t, Y = y_1) + \cdots + P(X = t, Y = y_K)$
- Note that the marginal PMF is simply a PMF

Example: Toss a fair coin twice

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

$t_2 \setminus t_1$	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: add over the columns of joint PMF table
  - $f_{X_1}(0) = f_{X_1X_2}(0,0) + f_{X_1X_2}(0,1)$
  - $f_{X_1}(1) = f_{X_1X_2}(1,0) + f_{X_1X_2}(1,1)$
- Marginal PMF of X_2: add over the rows of joint PMF table
  - $f_{X_2}(0) = f_{X_1X_2}(0,0) + f_{X_1X_2}(1,0)$
  - $f_{X_2}(1) = f_{X_1X_2}(0,1) + f_{X_1X_2}(1,1)$

Example: Marginal from joint PMF

Table of $f_{X_1 X_2}(t_1, t_2)$

$t_2 \setminus t_1$	0	1	$f_{X_2}(t_2)$
0	0.05	0.35	0.40
1	0.25	0.35	0.60
$f_{X_1}(t_1)$	0.30	0.70	

Verify: valid Joint PMF

$0.05 + 0.35 + 0.25 + 0.35 = 1$ ✓

Example: Same marginal PMF from different joint PMFs

Case 1

$t_2 \setminus t_1$	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	

Case 2

$t_2 \setminus t_1$	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	

- For every x between 0 and $1/2$, we get a joint PMF that results in the same marginal.

Example: Random 2-digit number

Table of $f_{XY}(t_1, t_2)$

$t_2 \setminus t_1$	0	1	2	3	4	5	6	7	8	9	$f_Y(t_2)$
0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$1/4$
1	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	$1/4$
2	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$1/4$
3	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	0	$\frac{1}{20}$	$1/4$
$f_X(t_1)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	

$X \sim \text{Uniform}\{0, 1, \ldots, 9\}$

$Y \sim \text{Uniform}\{0, 1, 2, 3\}$

$P(X=0, Y=0) = \Pr(X=0 \text{ and } Y=0) = 1/20$

$P(X=0)\,P(Y=0) = \frac{1}{10} \cdot \frac{1}{4} = \frac{1}{40}$

Not equal

Conditional distribution of a random variable given an event

Definition (Conditional distribution 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

$$Q(t) = P(X = t|A), \qquad t \in T_X.$$

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

Conditional distribution of a random variable given an event

Definition (Conditional distribution 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

$$Q(t) = P(X = t | A), \qquad t \in T_X.$$

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 random variable given another

Definition (Conditional distribution of Y given $X = t$)

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

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

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.

Conditional distribution of one random variable given another

Definition (Conditional distribution of Y given $X = t$)

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

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

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.

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

- Important: Range of $(Y|X = t)$ can be different from range of Y and will depend on t

Example: Compute conditional PMFs from joint PMF

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

$t_2 \setminus t_1$	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	

$X \in \{0, 1, 2\}$, $Y \in \{0, 1\}$

$(Y|X=0) \in \{0, 1\}$

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

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

$(X|Y=1) \in \{\underset{\frac{1/8}{1/2}=\frac{1}{4}}{0}, \underset{\frac{1/8}{1/2}=\frac{1}{4}}{1}, \underset{\frac{1/4}{1/2}=\frac{1}{2}}{2}\}$

Example: Compute marginal/conditional PMFs from joint PMF

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

$t_2 \setminus t_1$	0	1	2	$f_Y(t_2)$
0	1/12	0	3/12	1/3
1	2/12	1/12	0	1/4
2	3/12	1/12	1/12	5/12
$f_X(t_1)$	1/2	1/6	1/3	

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

$$Y|X=0 \sim \left\{ \overset{1/6}{0}, \overset{1/3}{1}, \overset{1/2}{2} \right\}$$

$$Y|X=1 \sim \left\{ \overset{1/2}{1}, \overset{1/2}{2} \right\}, \quad Y|X=2 \sim \left\{ \overset{3/4}{0}, \overset{1/4}{2} \right\}$$

$$X|Y=1 \sim \left\{ \overset{2/3}{0}, \overset{1/3}{1} \right\}$$

$$\boxed{f_{XY}(t_1, t_2) = f_{Y|X=t_1}(t_2)\, f_X(t_1) = f_{X|Y=t_2}(t_1) \cdot f_Y(t_2)}$$

Example: Throw a die and toss coins

Throw a die and toss a coin as many times as the number shown on die. Let X be the number shown on die. Let Y be the number of heads. What is the joint PMF of X and Y?

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

$(Y \mid X = t) \sim \text{Binomial}(t, 1/2)$ $\quad f_{Y|X=t}(t') = {}^tC_{t'} \left(\frac{1}{2}\right)^{t'} \cdot \left(\frac{1}{2}\right)^{t-t'}$

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

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

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

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

Exercise

Example: Poisson number of coin tosses

Let $N \sim \text{Poisson}(\lambda)$. Given $N = n$, toss a fair coin n times and denote the number of heads obtained by X. What is the distribution of X?

$f_X(x) = ?$

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

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

$$f_{NX}(n,k) = e^{-\lambda} \frac{\lambda^n}{n!} \cdot \binom{n}{k}\left(\frac{1}{2}\right)^n \qquad \begin{array}{l} n = 0, 1, 2, \ldots \\ k = 0, 1, \ldots, n \end{array}$$

$${}^nC_k = \binom{n}{k}$$

$$\underset{\substack{\downarrow \\ \text{Got } k \text{ heads}}}{f_X(k)} = \sum_{n=0}^{\infty} \underset{\substack{\downarrow \\ =0 \text{ if } n<k}}{f_{NX}(n,k)} = \sum_{n=k}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \cdot \binom{n}{k}\left(\frac{1}{2}\right)^n \qquad \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

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

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

$$\boxed{X \sim \text{Poisson}(\lambda/2)}$$

$$\boxed{f_X(k) = e^{-\lambda/2} \frac{(\lambda/2)^k}{k!}} \quad k = 0, 1, 2, \ldots$$

Example: IPL powerplay over

Let X = number of runs in the over. Let Y = number of wickets in the over. Assume the following:

$$Y \sim \{\overset{13/16}{0}, \overset{1/8}{1}, \overset{1/16}{2}\},$$
$$(X|Y=0) \sim \text{Uniform}\{6,7,8,9,10,11,12\},$$
$$(X|Y=1) \sim \text{Uniform}\{2,3,4,5,6,7,8\},$$
$$(X|Y=2) \sim \text{Uniform}\{0,1,2,3,4,5,6\}.$$

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

$f_X(t) = ?$

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

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

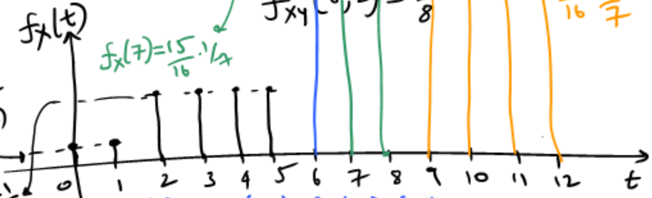
$f_X(2) = \underbrace{f_{XY}(2,2)}_{\frac{1}{16}\cdot\frac{1}{7}} + \underbrace{f_{XY}(2,1)}_{\frac{1}{8}\cdot\frac{1}{7}} = \frac{3}{16}\cdot\frac{1}{7}$

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

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

$\frac{13}{16}\cdot\frac{1}{7}$

$f_X(7) = \frac{15}{16}\cdot 1/7$



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

Section 2

More than two random variables: Joint, marginal, conditional PMFs

Multiple discrete random variables: Joint PMF

Definition (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 $f_{X_1 \cdots X_n}$, is a function from $T_{X_1} \times \cdots \times T_{X_n}$ to $[0, 1]$ defined as

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

- Joint PMF could be written as a table in small examples
- $P(X_1 = t_1 \text{ and } \cdots \text{ and } X_n = t_n)$ is denoted $P(X_1 = t_1, \ldots, X_n = t_n)$

Example: Toss a fair coin thrice

Let $X_i = 1$ if i-th toss is heads and $X_i = 0$ if i-th toss is tails, $i = 1, 2, 3$.

t_1	t_2	t_3	$f_{X_1X_2X_3}(t_1, t_2, t_3)$
0	0	0	1/8
0	0	1	1/8
0	1	0	1/8
0	1	1	1/8
1	0	0	1/8
1	0	1	1/8
1	1	0	1/8
1	1	1	1/8

Example: Random 3-digit number 000 to 999

$X =$ first digit from left, $Y =$ number modulo 2, $Z =$ first digit from right

- Table is too long

$$X \in \{0, 1, 2, \ldots, 9\}$$

$$Y \in \{0, 1\}$$

$$Z \in \{0, 1, 2, \ldots, 9\}$$

$f_{XYZ}(0,0,0) = P($Starts with zero and number is even and number ends in zero$)$

$$= \frac{10}{1000} = \frac{1}{100}$$

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

$f_{XYZ}(8,0,6) = \frac{10}{1000} = \frac{1}{100}$, $f_{XYZ}(8,1,6) = 0$

$f_{XYZ}(t_1,t_2,t_3) =$
0 if $t_2 = 0$ & t_3: odd
0 if $t_2 = 1$ & t_3: even
$\frac{1}{100}$ otherwise

Example: IPL powerplay over

Suppose the over has 6 deliveries. Let X_i denote the number of runs scored in the i-th delivery.

$$X_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$$

$i = 1, 2, \ldots, 6$

Joint PMF

$$\underbrace{f_{X_1 X_2 X_3 X_4 X_5 X_6}}_{\downarrow}(0, 1, 4, 2, 0, 0) = ?$$

8^6 : too many possibilities ....

Multiple discrete random variables: Marginal PMF (individual)

Definition (Marginal PMF (individual))

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

$$f_{X_1}(t) = P(X_1 = t) = \sum_{t'_2 \in T_{X_2}, t'_3 \in T_{X_3}, \ldots, t'_n \in T_{X_n}} f_{X_1 \cdots X_n}(t, t'_2, t'_3, \ldots, t'_n),$$

$$f_{X_2}(t) = P(X_2 = t) = \sum_{t'_1 \in T_{X_1}, t'_3 \in T_{X_3}, \ldots, t'_n \in T_{X_n}} f_{X_1 \cdots X_n}(t'_1, t, t'_3, \ldots, t'_n),$$

$$\vdots$$

$$f_{X_n}(t) = P(X_n = t) = \sum_{t'_1 \in T_{X_1}, \ldots, t'_{n-1} \in T_{X_{n-1}}, t \in T_{X_n}} f_{X_1 \cdots X_n}(t'_1, \ldots, t'_{n-1}, t),$$

where T_{X_i} is the range of X_i.

Example: Toss a fair coin thrice

Let $X_i = 1$ if i-th toss is heads and $X_i = 0$ if i-th toss is tails, $i = 1, 2, 3$.

Joint PMF: $f_{X_1 X_2 X_3}(t_1, t_2, t_3) = 1/8$ for $t_i \in \{0, 1\}$

Example: Toss a fair coin thrice

Let $X_i = 1$ if i-th toss is heads and $X_i = 0$ if i-th toss is tails, $i = 1, 2, 3$.

Joint PMF: $f_{X_1X_2X_3}(t_1, t_2, t_3) = 1/8$ for $t_i \in \{0, 1\}$

$$
\begin{aligned}
f_{X_1}(0) &= f_{X_1X_2X_3}(0, 0, 0) + f_{X_1X_2X_3}(0, 0, 1) + f_{X_1X_2X_3}(0, 1, 0) + f_{X_1X_2X_3}(0, 1, 1) \\
&= 1/8 + 1/8 + 1/8 + 1/8 = 1/2 \\
f_{X_1}(1) &= f_{X_1X_2X_3}(1, 0, 0) + f_{X_1X_2X_3}(1, 0, 1) + f_{X_1X_2X_3}(1, 1, 0) + f_{X_1X_2X_3}(1, 1, 1) \\
&= 1/8 + 1/8 + 1/8 + 1/8 = 1/2
\end{aligned}
$$

Note: $f_{X_1}(0) + f_{X_1}(1) = 1$

Example: Random 3-digit number 000 to 999

X = first digit from left, Y = number modulo 2, Z = first digit from right

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

Example: Random 3-digit number 000 to 999

X = first digit from left, Y = number modulo 2, Z = first digit from right

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

Example: Random 3-digit number 000 to 999

X = first digit from left, Y = number modulo 2, Z = first digit from right

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

Example: IPL powerplay Over 1

Suppose the over has 6 deliveries. Let X_i denote the number of runs scored in the i-th delivery.

Example: IPL powerplay Over 1

Suppose the over has 6 deliveries. Let X_i denote the number of runs scored in the i-th delivery.

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

Example: IPL powerplay Over 1

Suppose the over has 6 deliveries. Let X_i denote the number of runs scored in the i-th delivery.

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

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

Idea: Assign probabilities in the same proportion as data

Example: IPL powerplay Over 1

Suppose the over has 6 deliveries. Let X_i denote the number of runs scored in the i-th delivery.

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

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

Idea: Assign probabilities in the same proportion as data

	0	1	2	3	4	5	6
f_{X_1}	0.5989	0.2685	0.0357	0.0031	0.0864	0.0050	0.0025
f_{X_2}	0.5551	0.2791	0.0438	0.0031	0.1083	0.0044	0.0063
f_{X_3}	0.5338	0.2847	0.0444	0.0044	0.1139	0.0025	0.0163
f_{X_4}	0.5344	0.2516	0.0394	0.0031	0.1489	0.0038	0.0188
f_{X_5}	0.5313	0.2672	0.0407	0.0056	0.1358	0.0025	0.0169
f_{X_6}	0.5056	0.2954	0.0394	0.0050	0.1414	0.0013	0.0119

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 and X_3.

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

Marginalisation

Suppose $X_1, X_2, X_3 \sim f_{X_1 X_2 X_3}$ 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 and X_3.

What about $f_{X_1 X_2}$, the joint PMF of X_1 and X_2? What about $f_{X_1 X_3}$, $f_{X_2 X_3}$?

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

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

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

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 and X_3.

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

$$f_{X_1X_2}(t_1, t_2) = P(X_1 = t_1 \text{ and } X_2 = t_2) = \sum_{t'_3 \in T_{X_3}} f_{X_1X_2X_3}(t_1, t_2, t'_3)$$

$$f_{X_1X_3}(t_1, t_3) = P(X_1 = t_1 \text{ and } X_3 = t_3) = \sum_{t'_2 \in T_{X_2}} f_{X_1X_2X_3}(t_1, t'_2, t_3)$$

$$f_{X_2X_3}(t_2, t_3) = P(X_2 = t_2 \text{ and } X_3 = t_3) = \sum_{t'_1 \in T_{X_1}} f_{X_1X_2X_3}(t'_1, t_2, t_3)$$

Marginalisation: Sum over everything you do not want!

Example: $X_1, X_2, X_3 \sim f_{X_1X_2X_3}$

t_1	t_2	t_3	$f_{X_1X_2X_3}(t_1, t_2, t_3)$
0	0	0	1/9
0	0	1	1/9
0	0	2	1/9
0	1	1	1/9
0	1	2	1/9
1	0	0	1/9
1	0	2	1/9
1	1	0	1/9
1	1	1	1/9

$f_{X_1X_2}$

$t_2 \backslash t_1$	0	1
0	1/3	2/9
1	2/9	2/9

$f_{X_1X_3}$

$t_3 \backslash t_1$	0	1
0	1/9	2/9
1	2/9	1/9
2	2/9	1/9

Working

Marginalisation: More examples

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

$$f_{X_1}(t_1) = P(X_1 = t_1) = \sum_{t_2', t_3', t_4'} f_{X_1X_2X_3X_4}(t_1, t_2', t_3', t_4')$$

$$f_{X_1X_2}(t_1, t_2) = P(X_1 = t_1 \text{ and } X_2 = t_2) = \sum_{t_3', t_4'} f_{X_1X_2X_3X_4}(t_1, t_2, t_3', t_4')$$

$$f_{X_2X_4}(t_2, t_4) = P(X_2 = t_2 \text{ and } X_4 = t_4) = \sum_{t_1', t_3'} f_{X_1X_2X_3X_4}(t_1', t_2, t_3', t_4)$$

$$f_{X_1X_3X_4}(t_1, t_3, t_4) = \sum_{t_2'} f_{X_1X_2X_3X_4}(t_1, t_2', t_3.t_4)$$

Marginalisation: Sum over everything you do not want!

Multiple discrete random variables: Marginal PMF (general)

Definition (Marginal PMF (general))

Suppose $X_1, X_2, \ldots, X_n$ are jointly distributed discrete random variables with joint PMF $f_{X_1 \cdots X_n}$. The joint PMF of the random variables $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$, denoted $f_{X_{i_1} \cdots X_{i_k}}$, is given by

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

Conditioning with multiple discrete random variables

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.

Conditioning with multiple discrete random variables

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.

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

Conditioning with multiple discrete random variables

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.

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

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

Conditioning with multiple discrete random variables

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.

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

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

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

Conditioning with multiple discrete random variables

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.

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

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

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

$$(X_1, X_3|X_2 = t_2, X_4 = t_4) \sim f_{X_1X_3|X_2=t_2,X_4=t_4}(t_1, t_3) = \frac{f_{X_1X_2X_3X_4}(t_1, t_2, t_3, t_4)}{f_{X_2X_4}(t_2, t_4)}$$

Example: $X_1, X_2, X_3, X_4 \sim f_{X_1X_2X_3X_4}$

t_1	t_2	t_3	t_4	$f_{X_1\cdots X_4}(t_1,\ldots,t_4)$
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

$(X_1 | X_2 = 0) \sim \{\overset{\frac{(4/12)}{(7/12)}}{0}, \overset{\frac{(3/12)}{(7/12)}}{1}\}$

$(X_1 | X_3 = 0, X_4 = 1) \sim \{\overset{1/3}{0}, \overset{2/3}{1}\}$

$X_3, X_4 | X_1 = 0$

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

$T_{X_4} = T_{X_2} = T_{X_1} = \{0, 1\}$ $\quad T_{X_3} = \{0, 1, 2\}$

Working

Conditioning and factors of the joint PMF

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

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

Conditioning and factors of the joint PMF

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

$$\begin{aligned} f_{X_1\cdots X_4}(t_1, \ldots, t_4) &= P(X_1 = t_1 \text{ and } X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4) \\ &= P(X_1 = t_1 \text{ and } (X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4)) \end{aligned}$$

Conditioning and factors of the joint PMF

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

$$\begin{aligned}
f_{X_1\cdots X_4}(t_1, \ldots, t_4) &= P(X_1 = t_1 \text{ and } X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4) \\
&= P(X_1 = t_1 \text{ and } (X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4)) \\
&= P(X_1 = t_1 | (X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4)) \\
&\quad\quad P(X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4)
\end{aligned}$$

Conditioning and factors of the joint PMF

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

$$\begin{aligned}
f_{X_1\cdots X_4}(t_1, \ldots, t_4) &= P(X_1 = t_1 \text{ and } X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4) \\
&= P(X_1 = t_1 \text{ and } (X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4)) \\
&= P(X_1 = t_1 | (X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4)) \\
&\qquad P(X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4) \\
&= P(X_1 = t_1 | X_2 = t_2, X_3 = t_3, X_4 = t_4)) \\
&\qquad P(X_2 = t_2 | X_3 = t_3, X_4 = t_4) P(X_3 = t_3, X_4 = t_4)
\end{aligned}$$

Conditioning and factors of the joint PMF

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

$$
\begin{aligned}
f_{X_1\cdots X_4}(t_1, \ldots, t_4) &= P(X_1 = t_1 \text{ and } X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4) \\
&= P(X_1 = t_1 \text{ and } (X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4)) \\
&= P(X_1 = t_1 | (X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4)) \\
&\qquad P(X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4) \\
&= P(X_1 = t_1 | X_2 = t_2, X_3 = t_3, X_4 = t_4)) \\
&\qquad P(X_2 = t_2 | X_3 = t_3, X_4 = t_4) P(X_3 = t_3, X_4 = t_4) \\
&= P(X_1 = t_1 | X_2 = t_2, X_3 = t_3, X_4 = t_4)) \\
&\qquad P(X_2 = t_2 | X_3 = t_3, X_4 = t_4) \\
&\qquad\qquad P(X_3 = t_3 | X_4 = t_4) P(X_4 = t_4)
\end{aligned}
$$

Conditioning and factors of the joint PMF

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

$$\begin{aligned}
f_{X_1\cdots X_4}(t_1,\ldots,t_4) &= P(X_1 = t_1 \text{ and } X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4) \\
&= P(X_1 = t_1 \text{ and } (X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4)) \\
&= P(X_1 = t_1 | (X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4)) \\
&\qquad P(X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4) \\
&= P(X_1 = t_1 | X_2 = t_2, X_3 = t_3, X_4 = t_4)) \\
&\qquad P(X_2 = t_2 | X_3 = t_3, X_4 = t_4) P(X_3 = t_3, X_4 = t_4) \\
&= P(X_1 = t_1 | X_2 = t_2, X_3 = t_3, X_4 = t_4)) \\
&\qquad P(X_2 = t_2 | X_3 = t_3, X_4 = t_4) \\
&\qquad\qquad P(X_3 = t_3 | X_4 = t_4) P(X_4 = t_4) \\
&= f_{X_1|X_2=t_2,X_3=t_3,X_4=t_4}(t_1) f_{X_2|X_3=t_3,X_4=t_4}(t_2) \\
&\qquad\qquad f_{X_3|X_4=t_4}(t_3) f_{X_4}(t_4)
\end{aligned}$$

Example: $X_1, X_2, X_3 \sim f_{X_1X_2X_3}$

t_1	t_2	t_3	$f_{X_1X_2X_3}(t_1, t_2, t_3)$
0	0	0	1/9
0	0	1	1/9
0	0	2	1/9
0	1	1	1/9
0	1	2	1/9
1	0	0	1/9
1	0	2	1/9
1	1	0	1/9
1	1	1	1/9

$$\underbrace{f_{X_1X_2X_3}(0,0,0)}_{1/9} = \underbrace{f_{X_3}(0)}_{3/9} \cdot \underbrace{f_{X_2|X_3=0}(0)}_{2/3} \cdot \underbrace{f_{X_1|X_2=0,X_3=0}(0)}_{1/2}$$

Working

Factoring can be done in any sequence

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

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

Factoring can be done in any sequence

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

$$\begin{aligned} f_{X_1\cdots X_4}(t_1,\ldots,t_4) &= P(X_1 = t_1 \text{ and } X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4) \\ &= P(X_4 = t_4 \text{ and } X_3 = t_3 \text{ and } X_2 = t_2 \text{ and } X_1 = t_1) \end{aligned}$$

Factoring can be done in any sequence

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

$$\begin{aligned}
f_{X_1 \cdots X_4}(t_1, \ldots, t_4) &= P(X_1 = t_1 \text{ and } X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4) \\
&= P(X_4 = t_4 \text{ and } X_3 = t_3 \text{ and } X_2 = t_2 \text{ and } X_1 = t_1) \\
&= f_{X_4 | X_3 = t_3, X_2 = t_2, X_1 = t_1}(t_4) f_{X_3 | X_2 = t_2, X_1 = t_1}(t_3) \\
&\qquad f_{X_2 | X_1 = t_1}(t_2) f_{X_1}(t_1)
\end{aligned}$$

Factoring can be done in any sequence

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

$$\begin{aligned} f_{X_1\cdots X_4}(t_1,\ldots,t_4) &= P(X_1 = t_1 \text{ and } X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4) \\ &= P(X_4 = t_4 \text{ and } X_3 = t_3 \text{ and } X_2 = t_2 \text{ and } X_1 = t_1) \\ &= f_{X_4|X_3=t_3,X_2=t_2,X_1=t_1}(t_4) f_{X_3|X_2=t_2,X_1=t_1}(t_3) \\ &\qquad f_{X_2|X_1=t_1}(t_2) f_{X_1}(t_1) \end{aligned}$$

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

Factoring can be done in any sequence

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

$$
\begin{aligned}
f_{X_1 \cdots X_4}(t_1, \ldots, t_4) &= P(X_1 = t_1 \text{ and } X_2 = t_2 \text{ and } X_3 = t_3 \text{ and } X_4 = t_4) \\
&= P(X_4 = t_4 \text{ and } X_3 = t_3 \text{ and } X_2 = t_2 \text{ and } X_1 = t_1) \\
&= f_{X_4 | X_3 = t_3, X_2 = t_2, X_1 = t_1}(t_4) f_{X_3 | X_2 = t_2, X_1 = t_1}(t_3) \\
&\qquad f_{X_2 | X_1 = t_1}(t_2) f_{X_1}(t_1)
\end{aligned}
$$

$$
\begin{aligned}
f_{X_1 \cdots X_4}(t_1, \ldots, t_4) &= P(X_3 = t_3 \text{ and } X_2 = t_2 \text{ and } X_1 = t_1 \text{ and } X_4 = t_4) \\
&= f_{X_3 | X_2 = t_2, X_1 = t_1, X_4 = t_4}(t_3) f_{X_2 | X_1 = t_1, X_4 = t_4}(t_2) \\
&\qquad f_{X_1 | X_4 = t_4}(t_1) f_{X_4}(t_4)
\end{aligned}
$$

Example: $X_1, X_2, X_3 \sim f_{X_1X_2X_3}$

t_1	t_2	t_3	$f_{X_1X_2X_3}(t_1,t_2,t_3)$
0	0	0	1/9
0	0	1	1/9
0	0	2	1/9
0	1	1	1/9
0	1	2	1/9
1	0	0	1/9
1	0	2	1/9
1	1	0	1/9
1	1	1	1/9

$$\underbrace{f_{X_1X_2X_3}(0,0,0)}_{1/9} = \underbrace{f_{X_1}(0)}_{5/9} \cdot \underbrace{f_{X_2|X_1=0}(0)}_{3/5} \; \underbrace{f_{X_3|X_1=0,X_2=0}(0)}_{1/3}$$

Independence of two random variables

Definition (Independence (two random variables))

Let X and Y be two random variables defined in a probability space with ranges T_X and T_Y, respectively. X and Y are said to be independent if any event defined using X alone is independent of any event defined using Y alone. Equivalently, if the joint PMF of X and Y is f_{XY}, X and Y are independent if

$$f_{XY}(t_1, t_2) = f_X(t_1) f_Y(t_2)$$

for $t_1 \in T_X$, $t_2 \in T_Y$.

- General case vs independent case
  - General: $f_{XY}(t_1, t_2) = f_X(t_1) f_{Y|X=t_1}(t_2)$ (always true)
  - Independent: $f_{Y|X=t_1}(t_2) = f_Y(t_2)$
- If X and Y are independent
  - Joint PMF equals product of marginal PMFs
  - Conditional PMF equals marginal PMF

Simple examples

Independent

$t_2 \setminus t_1$	0	1	f_Y
0	1/4	1/4	1/2
1	1/4	1/4	1/2
f_X	1/2	1/2	

$f_{XY}(t_1, t_2) = f_X(t_1) f_Y(t_2)$ for all t_1, t_2

Simple examples

$t_2 \setminus t_1$	0	1	f_Y
0	1/4	1/4	1/2
1	1/4	1/4	1/2
f_X	1/2	1/2	

indep

NOT INDEPENDENT

$t_2 \setminus t_1$	0	1	f_Y
0	1/2	1/4	3/4
1	1/8	1/8	1/4
f_X	5/8	3/8	

$$f_{XY}(0,0) = 1/2 \neq \underset{5/8}{f_X(0)} \cdot \underset{3/4}{f_Y(0)}$$

Simple examples

$t_2 \setminus t_1$	0	1	f_Y
0	1/4	1/4	
1	1/4	1/4	
f_X	indep		

$t_2 \setminus t_1$	0	1	f_Y
0	1/2	1/4	
1	1/8	1/8	
f_X	not indep		

$t_2 \setminus t_1$	0	1	f_Y
0	x	$1/2 - x$	$1/2$
1	$1/2 - x$	x	$1/2$
f_X	$1/2$	$1/2$	

if $x = 1/4$, indep

$x \neq 1/4$, not indep.

More examples

$t_2 \setminus t_1$	0	1	2	f_Y
0	$1/9$	$1/9$	$1/9$	$1/3$
1	$1/9$	$1/9$	$1/9$	$1/3$
2	$1/9$	$1/9$	$1/9$	$1/3$
f_X	$1/3$	$1/3$	$1/3$	

indep

More examples

$t_2 \setminus t_1$	0	1	2	f_Y
0	1/9	1/9	1/9	
1	1/9	1/9	1/9	
2	1/9	1/9	1/9	
f_X		indep		

$t_2 \setminus t_1$	0	1	2	f_Y
0	1/6	1/12	1/12	1/3
1	1/4	1/8	1/8	1/2
2	1/12	1/24	1/24	1/6
f_X	1/2	1/4	1/4	

indep

More examples

$t_2 \setminus t_1$	0	1	2	f_Y
0	1/9	1/9	1/9	
1	1/9	1/9	1/9	
2	1/9	1/9	1/9	
f_X				

indep

$t_2 \setminus t_1$	0	1	2	f_Y
0	1/6	1/12	1/12	
1	1/4	1/8	1/8	
2	1/12	1/24	1/24	
f_X				

indep

$t_2 \setminus t_1$	0	1	2	f_Y
0	0	1/8	1/8	1/4
1	1/8	1/8	1/8	3/8
2	1/8	1/8	1/8	3/8
f_X	1/4	3/8	3/8	

dependent

More examples

$t_2 \setminus t_1$	0	1	2	f_Y
0	1/9	1/9	1/9	
1	1/9	1/9	1/9	
2	1/9	1/9	1/9	
f_X		indep		

$t_2 \setminus t_1$	0	1	2	f_Y
0	1/6	1/12	1/12	
1	1/4	1/8	1/8	
2	1/12	1/24	1/24	
f_X		indep		

$t_2 \setminus t_1$	0	1	2	f_Y
0	0	1/8	1/8	
1	1/8	1/8	1/8	
2	1/8	1/8	1/8	
f_X		dep		

$t_2 \setminus t_1$	0	1	2	f_Y
0	1/6	1/12	1/8	
1	1/4	1/8	1/8	1/2
2	1/24	1/24	1/24	
f_X			7/24	

dep

Independent vs dependent random variables

- To show X and Y are independent, verify

$$f_{XY}(t_1, t_2) = f_X(t_1) f_Y(t_2)$$

  for **all** $t_1 \in T_X$, $t_2 \in T_Y$.

- To show X and Y are dependent, verify

$$f_{XY}(t_1, t_2) \neq f_X(t_1) f_Y(t_2)$$

  for **some** $t_1 \in T_X$ and $t_2 \in T_Y$.

  - Special case: $f_{XY}(t_1, t_2) = 0$ when $f_X(t_1) \neq 0$, $f_Y(t_2) \neq 0$

Example: Random 2-digit number

A 2-digit number from 00 to 99 is selected at random. Partial information is available about the number as two random variables. Let X be the digit in units place. Let Y be the remainder obtained when the number is divided by 4.

$$X \sim \text{Uniform}\{0, 1, 2, \ldots, 9\} \quad \tfrac{1}{10}$$

$$Y \sim \text{Uniform}\{0, 1, 2, 3\} \quad \tfrac{1}{4}$$

$$f_{XY}(1, 0) = 0 \neq f_X(1) \cdot f_Y(0)$$

$$\Rightarrow X \text{ \& } Y : \text{dependent}$$

Example: IPL powerplay over

Let X = number of runs in the over. Let Y = number of wickets in the over.

Are X & Y independent?

Independence of multiple random variables

Definition (Independence of random variables)

Let $X_1, \ldots, X_n$ be random variables defined in a probability space with range of X_i denoted T_{X_i}. $X_1, \ldots, X_n$ are said to be independent if events defined using different X_i are mutually independent. Equivalently, $X_1, \ldots, X_n$ are independent iff

$$f_{X_1 \cdots X_n}(t_1, \ldots, t_2) = f_{X_1}(t_1) f_{X_2}(t_2) \cdots f_{X_n}(t_n)$$

for all $t_i \in T_{X_i}$.

- If X and Y are independent
  - Joint PMF equals product of marginal PMFs
  - Conditional PMF equals unconditioned PMF
- All subsets of independent random variables are independent

Example: Toss a fair coin thrice

Let $X_i = 1$ if i-th toss is heads and $X_i = 0$ if i-th toss is tails, $i = 1, 2, 3$.

Joint PMF: $f_{X_1 X_2 X_3}(t_1, t_2, t_3) = 1/8$ for $t_i \in \{0, 1\}$

Example: Toss a fair coin thrice

Let $X_i = 1$ if i-th toss is heads and $X_i = 0$ if i-th toss is tails, $i = 1, 2, 3$.

Joint PMF: $f_{X_1 X_2 X_3}(t_1, t_2, t_3) = 1/8$ for $t_i \in \{0, 1\}$

- $X_i \sim \text{Uniform}\{0, 1\}$

$$f_{X_1 X_2 X_3}(t_1, t_2, t_3) = f_{X_1}(t_1) f_{X_2}(t_2) f_{X_3}(t_3)$$

Example: Toss a fair coin thrice

Let $X_i = 1$ if i-th toss is heads and $X_i = 0$ if i-th toss is tails, $i = 1, 2, 3$.

Joint PMF: $f_{X_1X_2X_3}(t_1, t_2, t_3) = 1/8$ for $t_i \in \{0, 1\}$

- $X_i \sim \text{Uniform}\{0, 1\}$

$$f_{X_1X_2X_3}(t_1, t_2, t_3) = f_{X_1}(t_1)f_{X_2}(t_2)f_{X_3}(t_3)$$

X_1, X_2, X_3: Independent random variables

Example: Random 3-digit number 000 to 999

X = first digit from left, Y = number modulo 2, Z = first digit from right

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

Example: Random 3-digit number 000 to 999

X = first digit from left, Y = number modulo 2, Z = first digit from right

- $X \sim \text{Uniform}\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
- $Y \sim \text{Uniform}\{0, 1\}$
- $Z \sim \text{Uniform}\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
- X, Z
  - $f_{XZ}(t_1, t_3) = 1/100 = f_X(t_1)f_Z(t_3)$
  - Independent
- X, Y
  - $f_{XY}(t_1, t_2) = 1/20 = f_X(t_1)f_Y(t_2)$
  - Independent
- Y, Z
  - $f_{YZ}(1, 2) = 0 \neq f_Y(1)f_Z(2)$
  - Not independent

Example: Random 3-digit number 000 to 999

X = first digit from left, Y = number modulo 2, Z = first digit from right

- $X \sim \text{Uniform}\{0,1,2,3,4,5,6,7,8,9\}$
- $Y \sim \text{Uniform}\{0,1\}$
- $Z \sim \text{Uniform}\{0,1,2,3,4,5,6,7,8,9\}$
- X, Z
  - $f_{XZ}(t_1, t_3) = 1/100 = f_X(t_1) f_Z(t_3)$
  - Independent
- X, Y
  - $f_{XY}(t_1, t_2) = 1/20 = f_X(t_1) f_Y(t_2)$
  - Independent
- Y, Z
  - $f_{YZ}(1,2) = 0 \neq f_Y(1) f_Z(2)$
  - Not independent

X, Y, Z: Dependent random variables

Example: Even parity

t_1	t_2	t_3	$f_{X_1X_2X_3}(t_1, t_2, t_3)$
0	0	0	1/4
0	1	1	1/4
1	0	1	1/4
1	1	0	1/4

- Number of 1s in (X_1, X_2, X_3) is even (hence, the name)

Example: Even parity

t_1	t_2	t_3	$f_{X_1X_2X_3}(t_1, t_2, t_3)$
0	0	0	1/4
0	1	1	1/4
1	0	1	1/4
1	1	~~1~~ 0	1/4

- Number of 1s in (X_1, X_2, X_3) is even (hence, the name)
- $X_i \sim \text{Uniform}\{0, 1\}$
- All pairs are independent
- $f_{X_1X_2X_3}(0, 0, 1) = 0 \neq f_{X_1}(0) f_{X_2}(0) f_{X_3}(1)$

Example: Even parity

t_1	t_2	t_3	$f_{X_1X_2X_3}(t_1, t_2, t_3)$
0	0	0	1/4
0	1	1	1/4
1	0	1	1/4
1	1	1	1/4

- Number of 1s in (X_1, X_2, X_3) is even (hence, the name)
- $X_i \sim \text{Uniform}\{0, 1\}$
- All pairs are independent
- $f_{X_1X_2X_3}(0, 0, 1) = 0 \neq f_{X_1}(0)f_{X_2}(0)f_{X_3}(1)$

X_1, X_2, X_3: Dependent random variables

Independent and Identically Distributed (i.i.d.)

Definition (i.i.d.)

Random variables $X_1, \ldots, X_n$ are said to be independent and identically distributed (i.i.d.), if

1. they are independent,
2. the marginal PMFs f_{X_i} are identical.

Independent and Identically Distributed (i.i.d.)

Definition (i.i.d.)

Random variables $X_1, \ldots, X_n$ are said to be independent and identically distributed (i.i.d.), if

1. they are independent,
2. the marginal PMFs f_{X_i} are identical.

- Repeated trials of an experiment creates i.i.d. sequence of random variables
  - Toss a coin multiple times
  - Throw a die multiple times

Independent and Identically Distributed (i.i.d.)

Definition (i.i.d.)

Random variables $X_1, \ldots, X_n$ are said to be independent and identically distributed (i.i.d.), if

1. they are independent,
2. the marginal PMFs f_{X_i} are identical.

- Repeated trials of an experiment creates i.i.d. sequence of random variables
  - Toss a coin multiple times
  - Throw a die multiple times

Notation: $X_1, X_2, \ldots, X_n$ are i.i.d. with distribution $\underline{X}$ means that f_{X_i} is the same as f_X

PMF f_X

$$\boxed{X_1, X_2, \ldots, X_n \sim \text{iid } X}$$

(or)

$$X_1, X_2, \ldots, X_n \sim \text{iid } f_X$$

Problem: i.i.d. Geometric

Let $X_1, \ldots, X_n$ be i.i.d. with a Geometric(p) distribution. What is the probabilty that all of these random variables are larger than some positive integer j?

$X \sim \text{Geometric}(p) : X \in \{ \overset{p}{1}, \overset{(1-p)p}{2}, \overset{(1-p)^2 p}{3}, \ldots \}$ $\quad P_r(X=k) = (1-p)^{k-1} p$

$$P(X_1 > j, X_2 > j, \ldots, X_n > j) = P(X_1 > j) P(X_2 > j) \cdots P(X_n > j)$$
$$= \left(P(X > j) \right)^n = (1-p)^{jn}$$

$$P(X > j) = \sum_{k=j+1}^{\infty} (1-p)^{k-1} p = \frac{(1-p)^j p}{1-(1-p)} = (1-p)^j$$

Problem: i.i.d. samples

Let $X \sim \left\{ \begin{matrix} 1/2 & 1/4 & 1/8 & 1/16 & 1/16 \\ 0, & 1, & 2, & 3, & 4 \end{matrix} \right\}$, and let $X_1, \ldots, X_n$ be i.i.d. samples with distribution X. $\sim$ i.i.d. X

1. What is the probability that 4 is missing in the samples?
2. What is the probability that 4 appears exactly once in the samples?
3. What is the probability that 3 and 4 appear at least once in the samples?

① $P(X_1 \neq 4, X_2 \neq 4, \ldots, X_n \neq 4) = (P(X \neq 4))^n = (15/16)^n$

② $P(4 \text{ appears exactly once}) = P(X_1 = 4, X_2 \neq 4, \ldots, X_n \neq 4) + P(X_1 \neq 4, X_2 = 4, X_3 \neq 4, \ldots, X_n \neq 4) + \cdots + P(X_1 \neq 4, \ldots, X_{n-1} \neq 4, X_n = 4) = n\,P(X=4)(P(X \neq 4))^{n-1} = n\left(\frac{1}{16}\right)\left(\frac{15}{16}\right)^{n-1}$

③ $P(\underbrace{3 \text{ at least once}}_{A} \cap \underbrace{4 \text{ at least once}}_{B})$ $\quad A \cap B = (A^c \cup B^c)^c$

$P(A^c) = \left(\frac{15}{16}\right)^n$, $P(B^c) = \left(\frac{15}{16}\right)^n$, $P(A^c \cap B^c) = \left(\frac{14}{16}\right)^n \Rightarrow P(A^c \cup B^c) = 2\left(\frac{15}{16}\right)^n - \left(\frac{14}{16}\right)^n$

$P(A \cap B) = 1 - \left(2\left(\frac{15}{16}\right)^n - \left(\frac{14}{16}\right)^n\right)$

Problem: Memoryless property of Geometric

Let $X \sim \text{Geometric}(p)$. Find the following.

1. $P(X > n)$
2. $P(X > m + n | X > m)$

① $P(X>n) = \sum_{k=n+1}^{\infty} (1-p)^{k-1} p = (1-p)^n p + (1-p)^{n+1} p + \cdots$

$= \frac{(1-p)^n p}{1-(1-p)} = (1-p)^n$

② $P(X>m+n | X>m) = \frac{P(X>m+n \cap X>m)}{P(X>m)} = \frac{P(X>m+n)}{P(X>m)}$

$= \frac{(1-p)^{m+n}}{(1-p)^m} = (1-p)^n$

$$\boxed{P(X>m+n | X>m) = P(X>n)}$$

Section 4

Functions of random variables

Example: Throw a die twice

A fair die is thrown twice. What is the probability that the sum of the two numbers seen is 6? What is the PMF of the sum?

$$\underset{\text{first number}}{X_1}, \ \underset{\text{second number}}{X_2} \qquad \text{Let } S = X_1 + X_2$$

$$S \in \{2, 3, 4, \ldots, 12\}$$

$$\text{Sum} = \underbrace{X_1 + X_2}_{\text{function of } X_1 \text{ and } X_2}$$

$$P(S=2) = \frac{1}{36}, \ P(S=3) = \frac{2}{36}, \ P(S=4) = \frac{3}{36}, \ \ldots$$

Example: Area of a random rectangle

The length of a rectangle $L \sim$ Uniform$\{5, 7, 9, 11\}$. Given $L = l$, the breadth $B \sim$ Uniform$\{l-1, l-2, l-3\}$.

f_{LB}

t_1	t_2	$f_{LB}(t_1,t_2)$	Area $= \underline{\underline{LB}}$
5	4	1/12	20
	3	1/12	15
	2	1/12	10
7	6	1/12	42
	5	1/12	35
	4	1/12	28
9	8	1/12	72
	7	1/12	63
	6	1/12	54
11	10	1/12	110
	9	1/12	99
	8	1/12	88

PMF of $g(X_1, \ldots, X_n)$

Definition (PMF of a $g(X_1, \ldots, X_n)$)

Suppose $X_1, \ldots, X_n$ have joint PMF $f_{X_1 \cdots X_n}$ with T_{X_i} denoting the range of X_i. Let $g : T_{X_1} \times \cdots \times T_{X_n} \to \mathbb{R}$ be a function with range T_g. The PMF of $X = g(X_1, \ldots, X_n)$ is given by

$$f_X(t) = P(g(X_1, \ldots, X_n) = t) = \sum_{\underbrace{(t_1,\ldots,t_n):g(t_1,\ldots,t_n)=t}_{g^{-1}(t)}} f_{X_1 \cdots X_n}(t_1, \ldots, t_n).$$

- Proof: write the event and use definition of joint PMF
- Directly useful for small problems
- Can be extended for joint PMF of two functions g and h

Example: Given a small table for joint PMF

$t_2 t_3$	$g(t_1,t_2,t_3)$	$t_1+t_2+t_3$	t_1	t_2	t_3	$f_{X_1X_2X_3}(t_1,t_2,t_3)$
0		0	0	0	0	1/9
0		1	0	0	1	1/9
0		2	0	0	2	1/9
1		2	0	1	1	1/9
2		3	0	1	2	1/9
0		1	1	0	0	1/9
0		3	1	0	2	1/9
0		2	1	1	0	1/9
1		3	1	1	1	1/9

① $g(X_1,X_2,X_3) = X_1+X_2+X_3 \in \{\overset{1/9}{0}, \overset{2/9}{1}, \overset{3/9}{2}, \overset{3/9}{3}\}$

→ Any other function is similar

② $X = g(X_1,X_2,X_3) = X_1+X_2+X_3$
$Y = h(X_1,X_2,X_3) = X_2 X_3$

f_{XY}

$t_2 t_3$ \ $t_1+t_2+t_3$	0	1	2	3
0	1/9	2/9	2/9	1/9
1	0	0	1/9	1/9
2	0	0	0	1/9

Working

Example: Binomial from Bernoulli trials

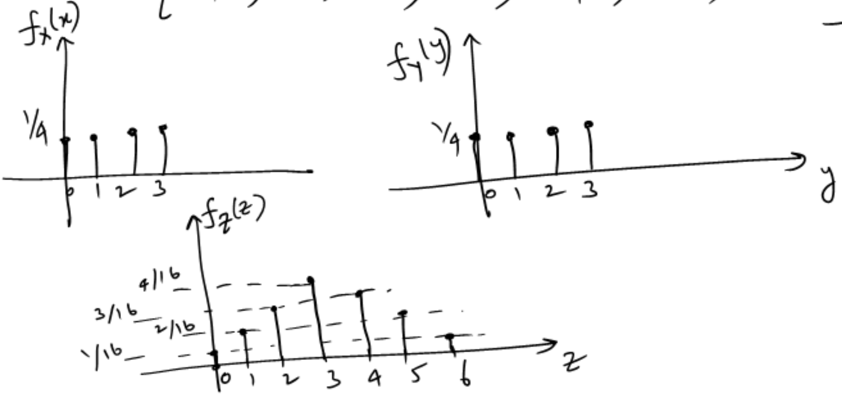
Let $X_1, \ldots, X_n$ be the results of n i.i.d. Bernoulli(p) trials. The sum of the n random variables $\underbrace{X_1 + \cdots + X_n}_{\text{\# of successes in } n \text{ indep Bernoulli trials}}$ is Binomial(n, p).

$\longrightarrow$ Sum of n indep Bernoulli(p) = Binomial(n,p)

Example: Sum of two uniforms

Let $X \sim \text{Uniform}\{0,1,2,3\}$ and $Y \sim \text{Uniform}\{0,1,2,3\}$ be independent. Find the PMF of $Z = X + Y$.

$$Z \in \{ \overset{1/16}{0}, \overset{2/16}{1}, \overset{3/16}{2}, \overset{4/16}{3}, \overset{3/16}{4}, \overset{2/16}{5}, \overset{1/16}{6} \}$$



Sums of two random variables taking integer values

Suppose X and Y take integer values and let their joint PMF be f_{XY}. Let $Z = X + Y$.

Sums of two random variables taking integer values

Suppose X and Y take integer values and let their joint PMF be f_{XY}. Let $Z = X + Y$.

Let z be some integer.

$$\begin{aligned}
P(Z = z) &= P(X + Y = z) \\
&= \sum_{x=-\infty}^{\infty} P(X = x, Y = z - x) \\
&= \sum_{x=-\infty}^{\infty} f_{XY}(x, z - x) \\
&= \sum_{y=-\infty}^{\infty} f_{XY}(z - y, y)
\end{aligned}$$

any integer

Sums of two random variables taking integer values

Suppose X and Y take integer values and let their joint PMF be f_{XY}. Let $Z = X + Y$.

Let z be some integer.

$$\begin{aligned} P(Z = z) &= P(X + Y = z) \\ &= \sum_{x=-\infty}^{\infty} P(X = x, Y = z - x) \\ &= \sum_{x=-\infty}^{\infty} f_{XY}(x, z - x) \\ &= \sum_{y=-\infty}^{\infty} f_{XY}(z - y, y) \end{aligned}$$

Convolution: If X and Y are independent,

$$f_{X+Y}(z) = \sum_{x=-\infty, \text{ integer}}^{\infty} f_X(x) f_Y(z - x)$$

Sum of two independent Poissons

Let $X \sim \text{Poisson}(\lambda_1)$ and $Y \sim \text{Poisson}(\lambda_2)$ be independent.

1. Find the PMF of $Z = X + Y$.
2. Find the conditional distribution of $X|Z$.

$z \in \{0, 1, 2, \ldots\}$

① $f_Z(z) = \sum_{x=0}^{\infty\, z} \underbrace{f_X(x)}_{=0 \text{ if } x<0} \underbrace{f_Y(z-x)}_{=0 \text{ if } x>z} = \sum_{x=0}^{z} \frac{e^{-\lambda_1} \lambda_1^x}{x!} \cdot \frac{e^{-\lambda_2} \lambda_2^{z-x}}{(z-x)!}$

$= \frac{e^{-\lambda_1} \cdot e^{-\lambda_2}}{z!} \sum_{x=0}^{z} \frac{z!}{x!(z-x)!} \lambda_1^x \lambda_2^{z-x}$

$\sum_{x=0}^{z} \frac{z!}{x!(z-x)!} \lambda_1^x \lambda_2^{z-x} = (\lambda_1+\lambda_2)^z$

$f_Z(z) = \frac{e^{-(\lambda_1+\lambda_2)} \cdot (\lambda_1+\lambda_2)^z}{z!}$

$Z \sim \text{Poisson}(\lambda_1+\lambda_2)$

② $P(X=k \mid Z=n) = \frac{P(X=k, Z=n)}{P(Z=n)} = \frac{P(X=k) \cdot P(Z=n \mid X=k)}{P(Z=n)} = \frac{P(X=k) \cdot P(Y=n-k)}{P(Z=n)}$

$= \frac{\cancel{e^{-\lambda_1}} \lambda_1^k / k! \cdot \cancel{e^{-\lambda_2}} \lambda_2^{n-k} / (n-k)!}{\cancel{e^{-(\lambda_1+\lambda_2)}} (\lambda_1+\lambda_2)^n / n!} = \frac{n!}{k!(n-k)!} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$

$\boxed{X|Z \sim \text{Binomial}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right) \quad Y|Z \sim \text{Binomial}\left(n, \frac{\lambda_2}{\lambda_1+\lambda_2}\right)}$

Functions and independence

- If X and Y are independent, $g(X)$ and $h(Y)$ are independent for any two functions g and h

Functions and independence

- If X and Y are independent, $g(X)$ and $h(Y)$ are independent for any two functions g and h
- If X_1, X_2, X_3, X_4 are mutually independent,

Functions and independence

- If X and Y are independent, $g(X)$ and $h(Y)$ are independent for any two functions g and h
- If X_1, X_2, X_3, X_4 are mutually independent,
  - $g(X_1, X_2)$ is independent of $h(X_3, X_4)$

Functions and independence

- If X and Y are independent, $g(X)$ and $h(Y)$ are independent for any two functions g and h
- If X_1, X_2, X_3, X_4 are mutually independent,
  - $g(X_1, X_2)$ is independent of $h(X_3, X_4)$
  - $g(X_1, X_2, X_3)$ is independent of $h(X_4)$

Functions and independence

- If X and Y are independent, $g(X)$ and $h(Y)$ are independent for any two functions g and h
- If X_1, X_2, X_3, X_4 are mutually independent,
  - $g(X_1, X_2)$ is independent of $h(X_3, X_4)$
  - $g(X_1, X_2, X_3)$ is independent of $h(X_4)$
- Functions of non-overlapping sets of independent random variables are also independent.

Exercises

- Sum of independent Binomial(m, p) and Binomial(n, p)
- Sum of independent Geometric(p) and Geometric(q)
- Sum of r i.i.d. Geometric(p)
- Sum of independent Negative-Binomial(r, p) and Negative-Binomial(s, p)

Minimum of two random variables

$$X, Y \sim f_{XY}$$

$$Z = \min(X, Y)\text{: function of } X \text{ and } Y$$

Ex: Throw a die twice. $Z =$ minimum of the two numbers seen.

Minimum of two random variables

$$X, Y \sim f_{XY}$$

$$Z = \min(X, Y)\text{: function of } X \text{ and } Y$$

Ex: Throw a die twice. $Z =$ minimum of the two numbers seen.

$$\begin{aligned} f_Z(z) &= P(\min(X, Y) = z) \\ &= P((X = z \text{ and } Y = z) \text{ or } (X = z \text{ and } Y > z) \\ &\qquad \text{or } (X > z \text{ and } Y = z)) \\ &= f_{XY}(z, z) + \sum_{t_2 > z} f_{XY}(z, t_2) + \sum_{t_1 > z} f_{XY}(t_1, z) \end{aligned}$$

Minimum of two random variables

$$X, Y \sim f_{XY}$$

$$Z = \min(X, Y)\text{: function of } X \text{ and } Y$$

Ex: Throw a die twice. $Z =$ minimum of the two numbers seen.

$$\begin{aligned} f_Z(z) &= P(\min(X, Y) = z) \\ &= P((X = z \text{ and } Y = z) \text{ or } (X = z \text{ and } Y > z) \\ &\qquad\qquad \text{or } (X > z \text{ and } Y = z)) \\ &= f_{XY}(z, z) + \sum_{t_2 > z} f_{XY}(z, t_2) + \sum_{t_1 > z} f_{XY}(t_1, z) \end{aligned}$$

What about maximum?

Example: Throw a die twice

Throw a die twice. Z = minimum of the two numbers seen.

	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

$P(Z=1)=\frac{11}{36}$ $P(Z=2)=\frac{9}{36}$ $P(Z=3)=\frac{7}{36}$

$P(Z=4)=\frac{5}{36}$ $P(Z=5)=\frac{3}{36}$ $P(Z=6)=\frac{1}{36}$

Independent case: cumulative distribution function (CDF) of maximum

Definition (CDF of a random variable)

Cumulative distribution function of a random variable X is a function $F_X : \mathbb{R} \to [0, 1]$ defined as

$$F_X(x) = P(X \leq x).$$

Independent case: cumulative distribution function (CDF) of maximum

Definition (CDF of a random variable)

Cumulative distribution function of a random variable X is a function $F_X : \mathbb{R} \to [0,1]$ defined as

$$F_X(x) = P(X \leq x).$$

Suppose X and Y are **independent** and $Z = \max(X, Y)$.

$$\begin{aligned} F_Z(z) &= P(\max(X, Y) \leq z) \\ &= P((X \leq z) \text{ and } (Y \leq z)) \\ &= P(X \leq z)P(Y \leq y) \\ &= F_X(z)F_Y(z) \end{aligned}$$

by independence

Independent case: cumulative distribution function (CDF) of maximum

Definition (CDF of a random variable)

Cumulative distribution function of a random variable X is a function $F_X : \mathbb{R} \to [0,1]$ defined as

$$F_X(x) = P(X \leq x).$$

Suppose X and Y are **independent** and $Z = \max(X, Y)$.

$$\begin{aligned} F_Z(z) &= P(\max(X, Y) \leq z) \\ &= P((X \leq z) \text{ and } (Y \leq z)) \\ &= P(X \leq z)P(Y \leq y) \\ &= F_X(z)F_Y(z) \end{aligned}$$

Independent: CDF of maximum is product of CDFs. What about min?

Problem: Min and max of i.i.d. sequences

Let $X_1, \dots, X_n \sim$ i.i.d. X. Find the distribution of the following:

1. $\min(X_1, \dots, X_n)$
2. $\max(X_1, \dots, X_n)$

(2) $P(\max(X_1, \dots, X_n) \le z) = P(X_1 \le z, X_2 \le z, \dots, X_n \le z)$

$= (P(X \le z))^n = (F_X(z))^n$

(1) $P(\min(X_1, \dots, X_n) \geq z) = P(X_1 \geq z, X_2 \geq z, \dots, X_n \geq z)$

$= (P(X \geq z))^n$

Problem: Min of two independent Geometrics

Let $X \sim \text{Geometric}(p)$ and $Y \sim \text{Geometric}(p)$ be independent. Find the distribution of $\min(X, Y)$.

$P(X=k) = (1-p)^{k-1} p$

$$P(\underline{\underline{\min(X,Y)}} \geq k) = P(X \geq k) \cdot P(Y \geq k)$$

$$= (1-p)^{k-1}(1-p)^{k-1} = \left((1-p)^2\right)^{k-1}$$

$$P(\min(X,Y) \geq k+1) = \left((1-p)^2\right)^k \qquad q = (1-p)^2$$

$$P(\underline{\underline{\min(X,Y)}} = k) = P(\min(X,Y) \geq k) - P(\min(X,Y) \geq k+1)$$

$$= q^{k-1} - q^k = q^{k-1}(1-q)$$

$$\min(X,Y) \sim \text{Geometric}(1-q)$$

$X_1 \sim \text{Geometric}(p_1)$, $X_2 \sim \text{Geometric}(p_2)$ indep

$$\min(X_1, X_2) \sim \text{Geometric}(1 - (1-p_1)(1-p_2))$$

$\hookrightarrow p_1 + p_2 - p_1 p_2$