Independence of Random Variables

- Recall that events A and B are independent if
$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B].$$

- A pair of random variables X and Y are independent if, for any valid events $\{X \in A\}$ and $\{Y \in B\}$, we have that $\mathbb{P}[\{X \in A\} \cap \{Y \in B\}] = \mathbb{P}[\{X \in A\}]\, \mathbb{P}[\{Y \in B\}]$.

→ This is a stronger notion of independence, but seems hard to check. Good news: there is an easier way to check.

- A pair of random variables X and Y is independent if and only if
$$F_{X,Y}(x,y) = F_X(x)\, F_Y(y) \quad \text{for all } x, y$$

→ For discrete X and Y can just check $P_{X,Y}(x,y) = P_X(x)\, P_Y(y)$.

→ For continuous X and Y can just check $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$.

- Intuition: If X and Y are independent, then we cannot predict X from observing Y any better than predicting X from no observation (and vice versa).

- This intuition can be formalized using conditional distributions:

→ Discrete random variables X and Y are independent if and only if $\underbrace{P_{X|Y}(x|y) = P_X(x)}$ and $\underbrace{P_{Y|X}(y|x) = P_Y(y)}$.

Suffices to check one of these conditions since $P_{X,Y}(x,y) = P_{X|Y}(x|y)\,P_Y(y)$
$= P_{Y|X}(y|x)\,P_X(x)$.

→ Continuous random variables X and Y are independent if and only if $\underbrace{f_{X|Y}(x|y) = f_X(x)}$ and $\underbrace{f_{Y|X}(y|x) = f_Y(y)}$.

Suffices to check one of these conditions since $f_{X,Y}(x,y) = f_{X|Y}(x|y)\,f_Y(y)$
$= f_{Y|X}(y|x)\,f_X(x)$.

- **Example:** X and Y are discrete with joint PMF

$P_{X,Y}(x,y)$		$x=0$	$x=1$
y	0	$\frac{1}{6}$	$\frac{1}{2}$
	1	$\frac{1}{12}$	$\frac{1}{4}$

Are X and Y independent?

① Compute $P_X(x)$ and $P_Y(y)$.

② Check if $P_{X,Y}(x,y) = P_X(x)\,P_Y(y)$.

① $$P_X(x) = \begin{cases} \frac{1}{6} + \frac{1}{12} & x = 0 \\ \frac{1}{2} + \frac{1}{4} & x = 1 \end{cases} = \begin{cases} \frac{1}{4} & x = 0 \\ \frac{3}{4} & x = 1 \end{cases}$$

add up each column

$$P_Y(y) = \begin{cases} \frac{1}{6} + \frac{1}{2} & y = 0 \\ \frac{1}{12} + \frac{1}{4} & y = 1 \end{cases} = \begin{cases} \frac{2}{3} & y = 0 \\ \frac{1}{3} & y = 1 \end{cases}$$

add up each row

②

$P_X(x)P_Y(y)$		$x=0$	$x=1$
y	0	$\frac{1}{4}\cdot\frac{2}{3} = \frac{1}{6}$ ✓	$\frac{3}{4}\cdot\frac{2}{3} = \frac{1}{2}$ ✓
	1	$\frac{1}{4}\cdot\frac{1}{3} = \frac{1}{12}$ ✓	$\frac{3}{4}\cdot\frac{1}{3} = \frac{1}{4}$ ✓

Since $P_{X,Y}(x,y) = P_X(x)\,P_Y(y)$, X and Y are independent.

- **Example**: X and Y are continuous with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{3}(x+y) & 0 \le x \le 2,\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Same example as last lecture.

Are X and Y independent?

① Determine marginal PDFs.

$$f_X(x) = \int_0^1 \frac{1}{3}(x+y)\,dy = \left(\frac{1}{3}xy + \frac{1}{6}y^2\right)\Big|_0^1 = \begin{cases} \frac{x}{3} + \frac{1}{6} & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \int_0^2 \frac{1}{3}(x+y)\,dx = \left(\frac{1}{6}x^2 + \frac{1}{3}xy\right)\Big|_0^2 = \begin{cases} \frac{2y}{3} + \frac{2}{3} & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

② Check if $f_X(x)f_Y(y) = f_{X,Y}(x,y)$.

$$f_X(x)f_Y(y) = \underbrace{\left(\frac{x}{3}+\frac{1}{6}\right)\left(\frac{2y}{3}+\frac{2}{3}\right)}_{\text{Range deliberately omitted to save space.}} = \begin{cases} \frac{1}{9}(2x+1)(y+1) & \underbrace{0 \le x \le 2,\ 0 \le y \le 1}_{\text{Don't forget the range in the end.}} \\ 0 & \text{otherwise} \end{cases}$$

$\neq f_{X,Y}(x,y) \Rightarrow$ X and Y dependent

- Is there an easier way to check independence? Sometimes.
- If the range does not factor as $R_{X,Y} = R_X \times R_Y$, then the joint PMF (or PDF) will not factor either.

→ Discrete: If there is a pair (x,y) for which $P_{X,Y}(x,y) = 0$ but neither $P_X(x) = 0$ nor $P_Y(y) = 0$, then X and Y are dependent.

Joint PMF Table: There is a zero entry for which neither the entire column is zero nor the entire row is zero.

→ Continuous: If there is a pair (x,y) for which $f_{X,Y}(x,y) = 0$ but neither $f_X(x) = 0$ nor $f_Y(y) = 0$, then X and Y are dependent.

Joint Range Sketch: The range is not a collection of rectangles parallel to the axes.

- Even if the range factors, X and Y may be dependent. (Previous example)

- Examples:

① $P_{X,Y}(x,y)$

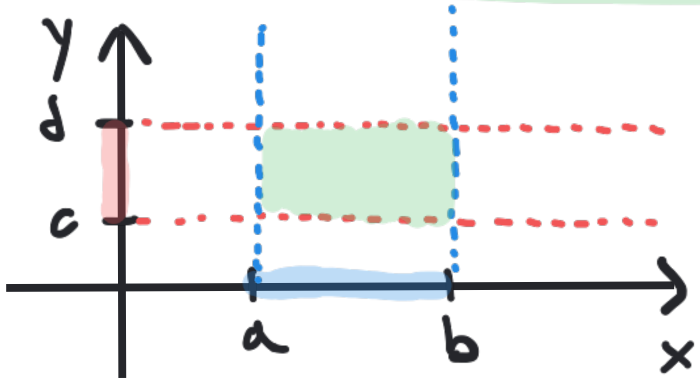
y \ x	1	2	3
1	$\frac{1}{5}$	0	$\frac{1}{5}$
2	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

Zero entry for which neither the row nor the column are all zero.

X, Y dependent.

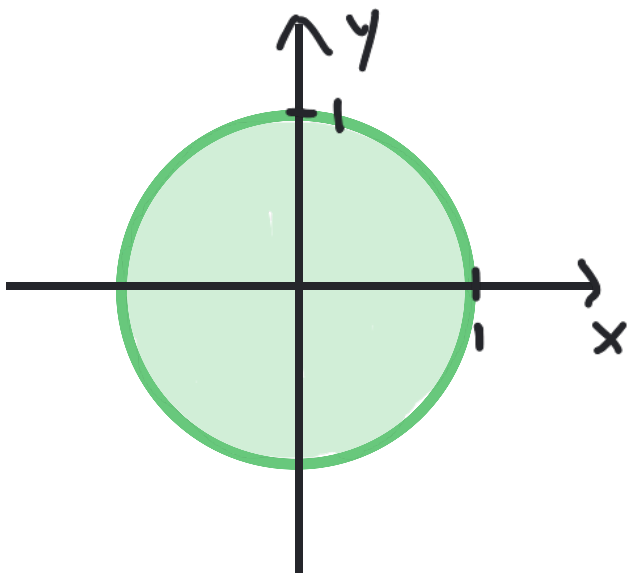
② $R_{X,Y} = \boxed{\{(x,y) : a \le x \le b,\ c \le y \le d\}} = \boxed{\{x : a \le x \le b\}} \times \boxed{\{y : c \le y \le d\}}$

$= R_X \times R_Y$



Range factors so X and Y may be independent. Check if $f_{X,Y}(x,y) = f_X(x) f_Y(y)$.

③ $R_{X,Y} = \boxed{\{(x,y) : x^2 + y^2 \le 1\}}$



Range does not factor so X and Y are dependent.