CSE 312

# Foundations of Computing II

## Lecture 14: Joint Distributions

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

**Aleks Jovcic**

Slide Credit: Based on Stefano Tessaro's slides for 312 19au
incorporating ideas from Alex Tsun, Rachel Lin, Hunter Schafer & myself ☺

# Agenda

- Cartesian Products
- Joint PMFs/PDFs/CDFs and Joint Range
- Independence
- Marginal Distributions
- Expectation
- Joint Continuous Random Variables

# Why joint distributions?

- Given all of its user's ratings for different movies, and any preferences you have expressed, Netflix wants to recommend a new movie for you.

- Given a bunch of medical data correlating symptoms and personal history with diseases, predict what is ailing a person with a particular medical history and set of symptoms.

- Given current traffic, pedestrian locations, weather, lights, etc. decide whether a self-driving car should slow down or come to a stop.

# Review Cartesian Product

**Definition.** Let $A$ and $B$ be sets. The **Cartesian product** of $A$ and $B$ is denoted

$$A \times B = \{(a, b) : a \in A, b \in B\}$$

**Example.**

$$\{1,2,3\} \times \{4, 5\} = \{(1, 4), (1,5), (2,4), (2,5), (3, 4), (3,5)\}$$

If $A$ and $B$ are finite sets, then $|A \times B| = |A| \cdot |B|$.

The sets don't need to be finite! You can have $\mathbb{R} \times \mathbb{R}$ (often denoted $\mathbb{R}^2$)

# Agenda

- Cartesian Products
- Joint PMFs/PDFs/CDFs and Joint Range
- Independence
- Marginal Distributions
- Expectation
- Joint Continuous Random Variables

# Joint PMFs and Joint Range

**Definition.** Let $X$ and $Y$ be discrete random variables. The **Joint PMF** of $X$ and $Y$ is

$$p_{X,Y}(a,b) = \Pr(X = a, Y = b)$$

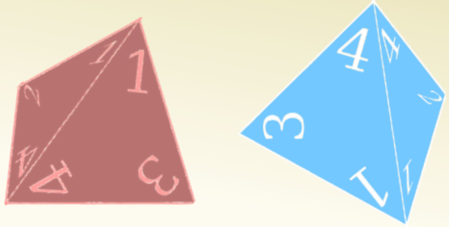
**Definition.** The **joint range** of $p_{X,Y}$ is

$$\Omega(X,Y) = \{(c,d) : p_{X,Y}(c,d) > 0\} \subseteq \Omega(X) \times \Omega(Y)$$

Note that

$$\sum_{(s,t)\in\Omega(X,Y)} p_{X,Y}(s,t) = 1$$

# Example: Weird Dice



Suppose I roll two fair 4-sided die independently. Let $X$ be the value of the first die, and $Y$ be the value of the second die.

$\Omega(X) = \{1,2,3,4\}$ and $\Omega(Y) = \{1,2,3,4\}$

| X\Y | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1/16 | 1/16 | 1/16 | 1/16 |
| 2 | 1/16 | 1/16 | 1/16 | 1/16 |
| 3 | 1/16 | 1/16 | 1/16 | 1/16 |
| 4 | 1/16 | 1/16 | 1/16 | 1/16 |

In this problem, the joint PMF is

$$p_{X,Y}(x,y) = \begin{cases} 1/16, & x,y \in \Omega(X,Y) \\ 0, & \text{otherwise} \end{cases}$$

and the joint range is (since all combinations have non-zero probability)

$\Omega(X,Y) = \Omega(X) \times \Omega(Y)$

# Agenda

- Cartesian Products
- Joint PMFs/PDFs/CDFs and Joint Range
- Independence
- Marginal Distributions
- Expectation
- Joint Continuous Random Variables

# Independence

**Definition.** Let $X$ and $Y$ be discrete random variables. The **Joint PMF** of $X$ and $Y$ is

$$p_{X,Y}(a,b) = \Pr(X = a, Y = b)$$

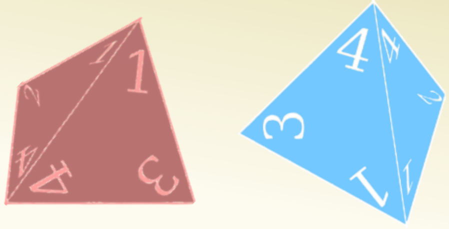
**Definition.** The **joint range** of $p_{X,Y}$ is

$$\Omega(X,Y) = \{(c,d) : p_{X,Y}(c,d) > 0\} \subseteq \Omega(X) \times \Omega(Y)$$

**Definition.** $X$ and $Y$ are **independent** iff for all $a, b$

$$\Pr(X = a, Y = b) = \Pr(X = a) \cdot \Pr(Y = b)$$

# Example: Weirder Dice



Suppose I roll two fair 4-sided die independently. Let $X$ be the value of the first die, and $Y$ be the value of the second die. Let $U = \min(X, Y)$ and $W = \max(X, Y)$

$\Omega(U) = \{1,2,3,4\}$ and $\Omega(W) = \{1,2,3,4\}$

$$\Omega(U, W) = \{(u, w) \in \Omega(U) \times \Omega(W) : u \le w\} \neq \Omega(U) \times \Omega(W)$$

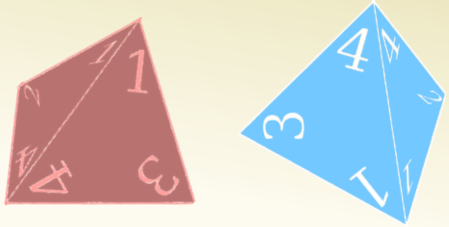
**Poll**:

What is $p_{U,W}(1, 3) = \Pr(U = 1, W = 3)$ ?

a. $1/16$
b. $2/16$
c. $1/2$
d. Not sure

| U\W | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |

# Example: Weirder Dice



Suppose I roll two fair 4-sided die independently. Let $X$ be the value of the first die, and $Y$ be the value of the second die. Let $U = \min(X, Y)$ and $W = \max(X, Y)$

$\Omega(U) = \{1,2,3,4\}$ and $\Omega(W) = \{1,2,3,4\}$

$$\Omega(U, W) = \{(u, w) \in \Omega(U) \times \Omega(W) : u \leq w\} \neq \Omega(U) \times \Omega(W)$$

The joint PMF $p_{U,W}(u, w) = \Pr(U = u, W = w)$ is

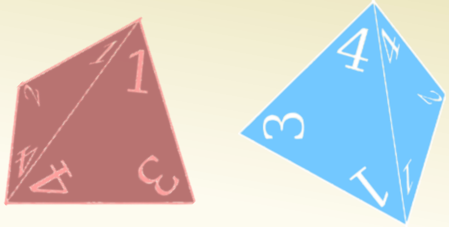
$$p_{U,W}(u,w) = \begin{cases} 2/16, & (u,w) \in \Omega(U) \times \Omega(W) \ \text{ where } w > u \\ 1/16, & (u,w) \in \Omega(U) \times \Omega(W) \ \text{ where } w = u \\ 0, & \text{otherwise} \end{cases}$$

| U\W | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1/16 | 2/16 | 2/16 | 2/16 |
| 2 | 0 | 1/16 | 2/16 | 2/16 |
| 3 | 0 | 0 | 1/16 | 2/16 |
| 4 | 0 | 0 | 0 | 1/16 |

# Agenda

- Cartesian Products
- Joint PMFs/PDFs/CDFs and Joint Range
- Independence
- Marginal Distributions
- Expectation
- Joint Continuous Random Variables

# Example: Weirder Dice



Suppose I roll two fair 4-sided die independently. Let $X$ be the value of the first die, and $Y$ be the value of the second die. Let $U = \min(X, Y)$ and $W = \max(X, Y)$

Suppose we didn't know how to compute $\Pr(U = u)$ directly. Can we figure it out if we know $p_{U,W}(u, w)$?

$$p_U(u) = \begin{cases} & u = 1 \\ & u = 2 \\ & u = 3 \\ & u = 4 \end{cases}$$

| U\W | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1/16 | 2/16 | 2/16 | 2/16 |
| 2 | 0 | 1/16 | 2/16 | 2/16 |
| 3 | 0 | 0 | 1/16 | 2/16 |
| 4 | 0 | 0 | 0 | 1/16 |

# Example: Weirder Dice

Suppose I roll two fair 4-sided die independently. Let $X$ be the value of the first die, and $Y$ be the value of the second die. Let $U = \min(X, Y)$ and $W = \max(X, Y)$

Suppose we didn't know how to compute $\Pr(U = u)$ directly. Can we figure it out if we know $p_{U,W}(u, w)$?

$$p_U(u) = \begin{cases} 7/16, & u = 1 \\ 5/16, & u = 2 \\ 3/16, & u = 3 \\ 1/16, & u = 4 \end{cases}$$

| U\W | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1/16 | 2/16 | 2/16 | 2/16 |
| 2 | 0 | 1/16 | 2/16 | 2/16 |
| 3 | 0 | 0 | 1/16 | 2/16 |
| 4 | 0 | 0 | 0 | 1/16 |

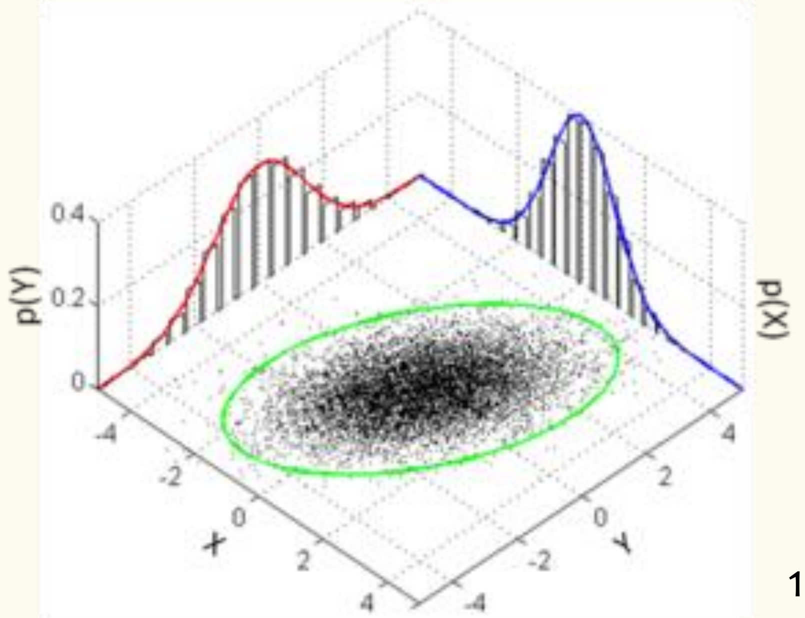
# Marginal PMF

**Definition.** Let $X$ and $Y$ be discrete random variables and $p_{X,Y}(a,b)$ their joint PMF. The **marginal PMF** of $X$

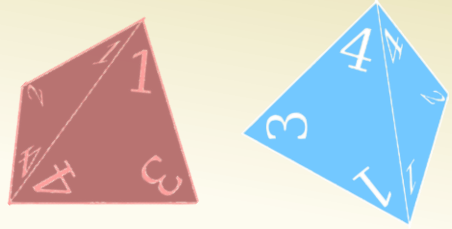
$$p_X(a) = \sum_{b \in \Omega(Y)} p_{X,Y}(a,b)$$

Similarly, $p_Y(b) = \sum_{a \in \Omega(X)} p_{X,Y}(a,b)$

Visual (for continuous $X$ and $Y$ )



# Marginal PMF Example



Suppose the table below gives us the joint pmf of X and Y.

What is the marginal pmf of X? What is the marginal pmf of Y?
Are X and Y independent?

| X\Y | 1 | 2 |
|---|---|---|
| 1 | 0.4 | 0.1 |
| 2 | 0.1 | 0.4 |

# Agenda

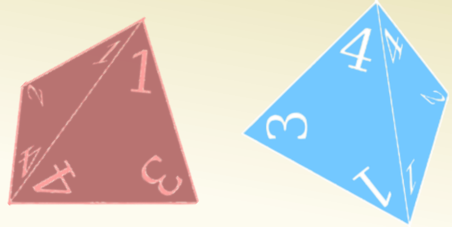
- Cartesian Products
- Joint PMFs/PDFs/CDFs and Joint Range
- Independence
- Marginal Distributions
- Expectation
- Joint Continuous Random Variables

# Joint Expectation

**Definition.** Let $X$ and $Y$ be discrete random variables and $p_{X,Y}(a,b)$ their joint PMF. The **expectation** of some function $g(x,y)$ with inputs $X$ and $Y$

$$E[g(X,Y)] = \sum_{a \in \Omega(X)} \sum_{b \in \Omega(Y)} g(a,b) p_{X,Y}(a,b)$$

# Expectation Example



Suppose the table below gives us the joint pmf of X and Y.

What is E(XY)?

| X\Y | 1 | 2 |
|---|---|---|
| 1 | 0.4 | 0.1 |
| 2 | 0.1 | 0.4 |

- Suppose the number of requests Z to a particular web server per hour is Poisson($\lambda$). And that the request comes from within the US with probability $p$.
- Let X be the number of requests per hour from the US and let Y be the number of requests per hour from outside the US. What is the joint pmf of X and Y? Are they independent?

# Agenda

- Cartesian Products
- Joint PMFs/PDFs/CDFs and Joint Range
- Independence
- Marginal Distributions
- Expectation
- Joint Continuous Random Variables

| | Discrete | Continuous |
|---|---|---|
| **Joint PMF/PDF** | $p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)$ | $f_{X,Y}(x,y) \neq \mathbb{P}(X = x, Y = y)$ |
| **Joint range/support** $\Omega_{X,Y}$ | $\{(x,y) \in \Omega_X \times \Omega_Y : p_{X,Y}(x,y) > 0\}$ | $\{(x,y) \in \Omega_X \times \Omega_Y : f_{X,Y}(x,y) > 0\}$ |
| **Joint CDF** | $F_{X,Y}(x,y) = \sum_{t \le x, s \le y} p_{X,Y}(t,s)$ | $F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\, ds dt$ |
| **Normalization** | $\sum_{x,y} p_{X,Y}(x,y) = 1$ | $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx dy = 1$ |
| **Marginal PMF/PDF** | $p_X(x) = \sum_y p_{X,Y}(x,y)$ | $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$ |
| **Expectation** | $\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y) p_{X,Y}(x,y)$ | $\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$ |

# Independence (continuous random variables)

**Definition.** Let $X$ and $Y$ be continuous random variables. The **joint pdf** of $X$ and $Y$ is

$$f_{X,Y}(a,b) \neq \Pr(X = a, Y = b)$$

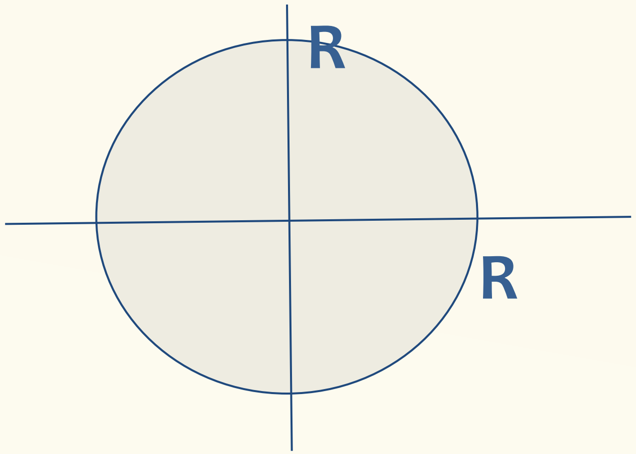
**Definition.** The **joint range** of $p_{X,Y}$ is

$$\Omega(X,Y) = \{(c,d) : p_{X,Y}(c,d) > 0\} \subseteq \Omega(X) \times \Omega(Y)$$

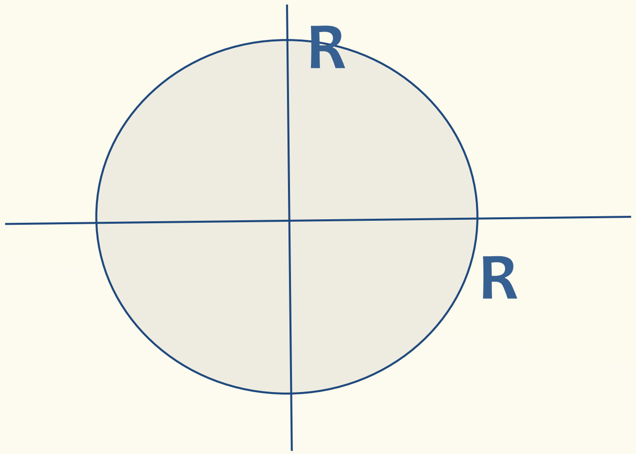
**Definition.** $X$ and $Y$ are **independent** iff for all $a, b$

$f_{X,Y}(a,b) = f_X(a) \cdot f_Y(b)$

- Suppose that the surface of a disk is a circle with area R centered at the origin and that there is a single point imperfection at a location with is uniformly distributed across the surface of the disk. Let X and Y be the x and y coordinates of the imperfection (random variables) and let Z be the distance of the imperfection from the origin.
  - What is their joint density f(x,y)?



- Suppose that the surface of a disk is a circle with area R centered at the origin and that there is a single point imperfection at a location with is uniformly distributed across the surface of the disk. Let X and Y be the x and y coordinates of the imperfection (random variables) and let Z be the distance of the imperfection from the origin.
  – What is the range of X & Y and the marginal density of X and of Y?




Poll:

What is $\Omega_X$?

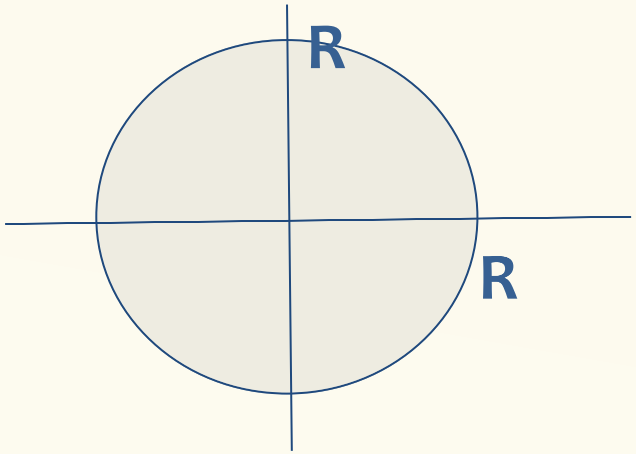
a. $[-\sqrt{R^2 - x^2}, \sqrt{R^2 - x^2}]$

b. $[-R, R]$

c. $[-\sqrt{R^2 - y^2}, \sqrt{R^2 - y^2}]$

d. Not sure

- Suppose that the surface of a disk is a circle with area R centered at the origin and that there is a single point imperfection at a location with is uniformly distributed across the surface of the disk. Let X and Y be the x and y coordinates of the imperfection (random variables) and let Z be the distance of the imperfection from the origin.
  - Are X and Y independent?

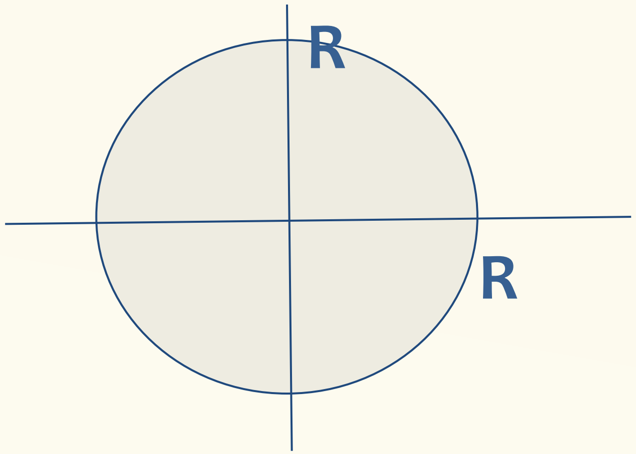



Poll:

Are X and Y independent?

a. yes
b. no

- Suppose that the surface of a disk is a circle with area R centered at the origin and that there is a single point imperfection at a location with is uniformly distributed across the surface of the disk. Let X and Y be the x and y coordinates of the imperfection (random variables) and let Z be the distance of the imperfection from the origin.
  - What is E(Z)?



# All of this generalizes to more than 2 random variables

| | Discrete | Continuous |
|---|---|---|
| **Joint PMF/PDF** | $p_{X,Y}(x,y) = \mathbb{P}(X=x, Y=y)$ | $f_{X,Y}(x,y) \neq \mathbb{P}(X=x, Y=y)$ |
| **Joint range/support** $\Omega_{X,Y}$ | $\{(x,y) \in \Omega_X \times \Omega_Y : p_{X,Y}(x,y) > 0\}$ | $\{(x,y) \in \Omega_X \times \Omega_Y : f_{X,Y}(x,y) > 0\}$ |
| **Joint CDF** | $F_{X,Y}(x,y) = \sum_{t \leq x, s \leq y} p_{X,Y}(t,s)$ | $F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\, ds dt$ |
| **Normalization** | $\sum_{x,y} p_{X,Y}(x,y) = 1$ | $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx dy = 1$ |
| **Marginal PMF/PDF** | $p_X(x) = \sum_y p_{X,Y}(x,y)$ | $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$ |
| **Expectation** | $\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y) p_{X,Y}(x,y)$ | $\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$ |