# ECE 302: Lecture 5.2 Joint Expectation

Prof Stanley Chan

School of Electrical and Computer Engineering
Purdue University

PURDUE
UNIVERSITY

# Joint Expectation

Recall:

$$\mathbb{E}[X] = \int_{\Omega} x f_X(x) dx.$$

How about the expectation for two variables?

Definition

Let $X$ and $Y$ be two random variables. The **joint expectation** is

$$\mathbb{E}[XY] = \sum_{y \in \Omega_Y} \sum_{x \in \Omega_X} xy \cdot p_{X,Y}(x, y) \tag{1}$$

if $X$ and $Y$ are discrete, or

$$\mathbb{E}[XY] = \int_{y \in \Omega_Y} \int_{x \in \Omega_X} xy \cdot f_{X,Y}(x, y) dx dy \tag{2}$$

if $X$ and $Y$ are continuous. Joint expectation is also called **correlation**.

# Example

**Example 1**. Let $X$ be a coin flip, $Y$ be a dice. Find $\mathbb{E}[XY]$.

**Solution**.

| | Y | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| X = 0 | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{12}$ |
| X = 1 | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{12}$ |

$$\mathbb{E}[XY] = \sum_{x,y} p_{X,Y}(x,y) =$$

$$= \frac{21}{12}.$$

# Outline

- Joint PDF and CDF
- Joint Expectation
- Conditional Distribution
- Conditional Expectation
- Sum of Two Random Variables
- Random Vectors
- High-dimensional Gaussians and Transformation
- Principal Component Analysis

Today's lecture

- What is joint expectation?
- Why define joint expectation as $\mathbb{E}[XY]$?
- Independence
- Covariance
- Correlation

# Understanding $\mathbb{E}[XY]$

So what is $\mathbb{E}[XY]$?

$$\mathbb{E}[XY] = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i y_j \cdot p_{X,Y}(x_i, y_j),$$

This can actually be written in the matrix form

$$\text{PMF} = \begin{bmatrix} p_{X,Y}(x_1, y_1) & p_{X,Y}(x_1, y_2) & \dots & p_{X,Y}(x_1, y_N) \\ p_{X,Y}(x_2, y_1) & p_{X,Y}(x_2, y_2) & \dots & p_{X,Y}(x_2, y_N) \\ \vdots & \vdots & \ddots & \vdots \\ p_{X,Y}(x_N, y_1) & p_{X,Y}(x_N, y_2) & \dots & p_{X,Y}(x_N, y_N) \end{bmatrix}. \tag{3}$$

# So, $\mathbb{E}[XY]$ is actually ...

$$\mathbb{E}[XY] = \underbrace{\begin{bmatrix} x_1 & \dots & x_N \end{bmatrix}}_{\boldsymbol{x}^T} \underbrace{\begin{bmatrix} p_{X,Y}(x_1,y_1) & \dots & p_{X,Y}(x_1,y_N) \\ \vdots & \ddots & \vdots \\ p_{X,Y}(x_N,y_1) & \dots & p_{X,Y}(x_N,y_N) \end{bmatrix}}_{\boldsymbol{P}} \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}_{\boldsymbol{y}}$$

$$= \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{y}. \tag{4}$$

**Why correlation is defined as $\mathbb{E}[XY]$?**

- $\mathbb{E}[XY]$ is a weighted inner product between the states:

$$\mathbb{E}[XY] = \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{y}. \tag{5}$$

- $\boldsymbol{x}$ and $\boldsymbol{y}$ are the states of the random variables $X$ and $Y$.
- Inner product measures the similarity between two vectors.

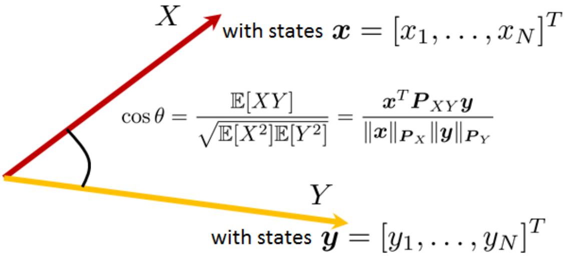
# Then, what is $\boldsymbol{x}^T \boldsymbol{P} \boldsymbol{y}$?

**Definition**

Let $\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{y} \in \mathbb{R}^N$ be two vectors. Define the **cosine angle** $\cos\theta$

$$\cos\theta = \frac{\boldsymbol{x}^T \boldsymbol{y}}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|}, \tag{6}$$

where $\|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^N x_i^2}$ is the **norm** of the vector $\boldsymbol{x}$, $\|\boldsymbol{y}\| = \sqrt{\sum_{i=1}^N y_i^2}$ is the norm of the vector $\boldsymbol{y}$.




**Distributions**

$p_X(x)$ or $\boldsymbol{P}_X$

$p_{X,Y}(x, y)$ or $\boldsymbol{P}_{XY}$

$p_Y(y)$ or $\boldsymbol{P}_Y$

# The denominators in cosine angle

$$\mathbb{E}[X^2] = \sum_{i=1}^{N} x_i x_i \cdot p_X(x_i)$$

$$= \underbrace{\begin{bmatrix} x_1 & \dots & x_N \end{bmatrix}}_{\boldsymbol{x}^T} \underbrace{\begin{bmatrix} p_X(x_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & p_X(x_N) \end{bmatrix}}_{\boldsymbol{P}_X} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}}_{\boldsymbol{y}} = \boldsymbol{x}^T \boldsymbol{P}_X \boldsymbol{x} = \|\boldsymbol{x}\|^2_{\boldsymbol{P}_X},$$

$$\mathbb{E}[Y^2] = \sum_{j=1}^{N} y_j y_j \cdot p_Y(y_j)$$

$$= \underbrace{\begin{bmatrix} y_1 & \dots & y_N \end{bmatrix}}_{\boldsymbol{y}^T} \underbrace{\begin{bmatrix} p_Y(y_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & p_Y(y_N) \end{bmatrix}}_{\boldsymbol{P}_Y} \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}_{\boldsymbol{y}} = \boldsymbol{y}^T \boldsymbol{P}_Y \boldsymbol{y} = \|\boldsymbol{y}\|^2_{\boldsymbol{P}_Y}.$$

# So, the cosine angle is

$$\cos\theta = \frac{\boldsymbol{x}^T \boldsymbol{P}_{XY}\boldsymbol{y}}{\|\boldsymbol{x}\|_{\boldsymbol{P}_X}\|\boldsymbol{y}\|_{\boldsymbol{P}_Y}}, \tag{7}$$

which is the same as

$$\cos\theta = \frac{\boldsymbol{x}^T \boldsymbol{P}_{XY}\boldsymbol{y}}{\|\boldsymbol{x}\|_{\boldsymbol{P}_X}\|\boldsymbol{y}\|_{\boldsymbol{P}_Y}} = \frac{\mathbb{E}[XY]}{\sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}}. \tag{8}$$

**What is $\mathbb{E}[XY]$ then?**

- $\mathbb{E}[XY]$ is the cosine angle between $\boldsymbol{x}$ and $\boldsymbol{y}$
- Measures how similar they are
- Weight by the PMF/PDF

# Therefore, ...

Since the cosine angle is between -1 and 1, we have that

$$-1 \leq \frac{\mathbb{E}[XY]}{\sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}} \leq 1. \tag{9}$$

If you like to prove this formally, ...

**Theorem (Cauchy-Schwarz Inequality)**

*For any random variables $X$ and $Y$, it holds that*

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]. \tag{10}$$

Proof: See ebook.

**Questions?**