ECE 302: Lecture 5.1 Joint PDF and CDF

Prof Stanley Chan

School of Electrical and Computer Engineering
Purdue University

PURDUE UNIVERSITY

©Stanley Chan 2022. All Rights Reserved.

What are joint distributions?

Joint distributions are **high-dimensional** PDF (or PMF or CDF).

$$\underbrace{f_X(x)}_{\text{one variable}} \implies \underbrace{f_{X_1,X_2}(x_1,x_2)}_{\text{two variables}} \implies \underbrace{f_{X_1,X_2,X_3}(x_1,x_2,x_3)}_{\text{three variables}}$$

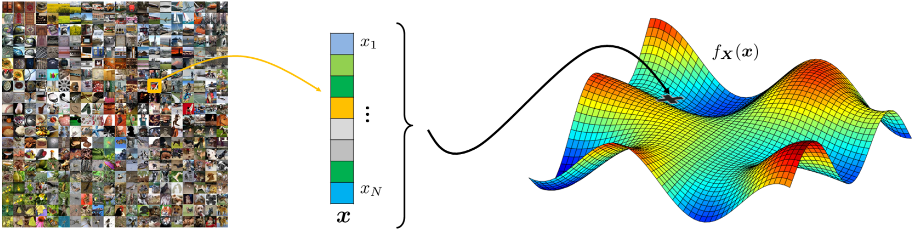
$$\implies \ldots \implies \underbrace{f_{X_1,\ldots,X_N}(x_1,\ldots,x_N)}_{N \text{ variables}}.$$

Notation:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1,\ldots,X_N}(x_1,\ldots,x_N).$$

©Stanley Chan 2022. All Rights Reserved.

Why study joint distributions?



- Joint distributions are ubiquitous in modern data analysis.
- For example, an image from a dataset can be represented by a high-dimensional vector $\boldsymbol{x}$.
- Each vector has certain probability to be present.
- Such probability is described by the high-dimensional joint PDF $f_{\boldsymbol{X}}(\boldsymbol{x})$.

©Stanley Chan 2022. All Rights Reserved.

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

- Joint PMF, PDF
- Joint CDF
- Marginal PDF
- Independence

©Stanley Chan 2022. All Rights Reserved.

Joint PMF

Definition

Let X and Y be two discrete random variables. The **joint PMF** of X and Y is defined as

$$p_{X,Y}(x,y) = \mathbb{P}[X = x \text{ and } Y = y]. \quad (1)$$

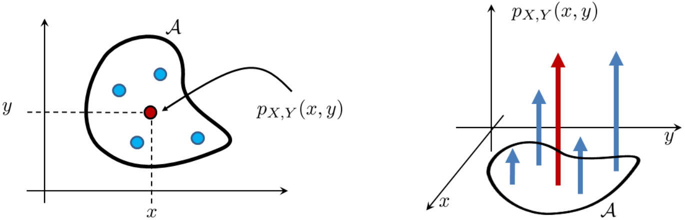


Figure: A joint PMF for a pair of discrete random variables consists of an array of impulses. To measure the size of the event $\mathcal{A}$, we sum all the impulses inside $\mathcal{A}$.

©Stanley Chan 2022. All Rights Reserved.

Example

Example 1. Let X be a coin flip, Y be a dice. Find the joint PMF.

Solution. The sample space of X is $\{0, 1\}$. The sample space of Y is $\{1, 2, 3, 4, 5, 6\}$. The joint PMF is

	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	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

Or written in equation:

$$p_{X,Y}(x, y) = \frac{1}{12}, \quad x = 0, 1, \quad y = 1, 2, 3, 4, 5, 6.$$

©Stanley Chan 2022. All Rights Reserved.

Example

Example 2. In the previous example, define $\mathcal{A} = \{X + Y = 3\}$ and $\mathcal{B} = \{\min(X, Y) = 1\}$. Find $\mathbb{P}[\mathcal{A}]$ and $\mathbb{P}[\mathcal{B}]$.

Solution:

$$\begin{aligned}
\mathbb{P}[\mathcal{A}] &= \sum_{(x,y)\in\mathcal{A}} p_{X,Y}(x,y) = p_{X,Y}(0,3) + p_{X,Y}(1,2) \\
&= \frac{2}{12} \\
\mathbb{P}[\mathcal{B}] &= \sum_{(x,y)\in\mathcal{B}} p_{X,Y}(x,y) \\
&= p_{X,Y}(1,1) + p_{X,Y}(1,2) + \ldots + p_{X,Y}(1,5) + p_{X,Y}(1,6) \\
&= \frac{6}{12}.
\end{aligned}$$

©Stanley Chan 2022. All Rights Reserved.

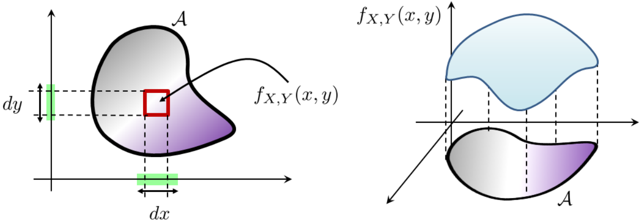
Joint PDF

Definition

Let X and Y be two continuous random variables. The **joint PDF** of X and Y is a function $f_{X,Y}(x,y)$ that can be integrated to yield a probability:

$$\mathbb{P}[\mathcal{A}] = \int_{\mathcal{A}} f_{X,Y}(x,y)dxdy, \tag{2}$$

for any event $\mathcal{A} \subseteq \Omega_X \times \Omega_Y$.



©Stanley Chan 2022. All Rights Reserved.

Example

Example 1. Consider a uniform joint PDF $f_{X,Y}(x,y)$ defined on $[0,2]^2$ with $f_{X,Y}(x,y) = \frac{1}{4}$. Let $\mathcal{A} = [a,b] \times [c,d]$. Find $\mathbb{P}[\mathcal{A}]$.

Solution:

$$
\begin{aligned}
\mathbb{P}[\mathcal{A}] &= \mathbb{P}[a \le X \le b, \;\; c \le X \le d] \\
&= \int_c^d \int_a^b f_{X,Y}(x,y) dx dy \\
&= \int_c^d \int_a^b \frac{1}{4} dx dy = \frac{(d-c)(b-a)}{4}.
\end{aligned}
$$

©Stanley Chan 2022. All Rights Reserved.

Example

Example 2. In the previous example, let $\mathcal{B} = \{X + Y \leq 2\}$. Find $\mathbb{P}[\mathcal{B}]$.

Solution:

$$\begin{aligned}\mathbb{P}[\mathcal{B}] = \int_{\mathcal{B}} f_{X,Y}(x,y)dxdy &= \int_0^2 \int_0^{2-y} f_{X,Y}(x,y)dxdy \\ &= \int_0^2 \int_0^{2-y} \frac{1}{4}dxdy \\ &= \int_0^2 \frac{2-y}{4}dy = \frac{1}{2}.\end{aligned}$$

©Stanley Chan 2022. All Rights Reserved.

Review of 2D Integration

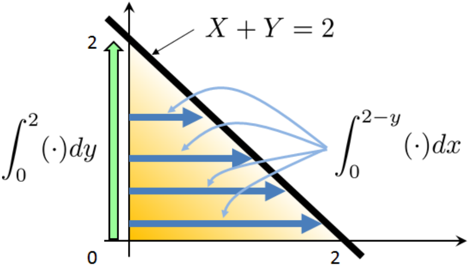


Figure: To integrate the probability $\mathbb{P}[X + Y \leq 2]$, we perform a 2D integration over a triangle.

©Stanley Chan 2022. All Rights Reserved.

Example

Example 3. Consider a joint PDF

$$f_{X,Y}(x,y) = \begin{cases} ce^{-x}e^{-y}, & 0 \le y \le x < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

Find the constant c.

Solution. There are two ways to take the integration. We choose the inner integration w.r.t. y first.

$$\begin{aligned} \int_{\Omega} f_{X,Y}(x,y)dxdy &= \int_0^{\infty} \int_0^{x} ce^{-x}e^{-y} dydx \\ &= \int_0^{\infty} ce^{-x}(1-e^{-x}) = \frac{c}{2}. \end{aligned}$$

Therefore, $c = 2$.

©Stanley Chan 2022. All Rights Reserved.

Review of 2D Integration

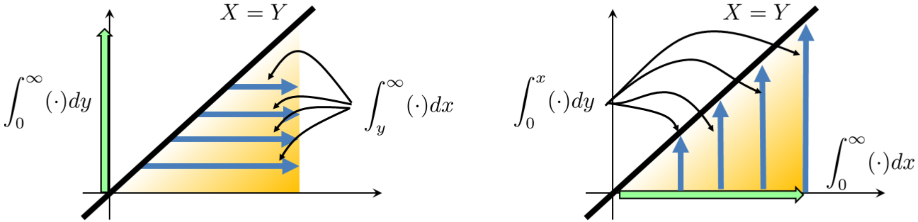


Figure: To integrate the probability $\mathbb{P}[0 \leq Y \leq X]$, we perform a 2D integration over a triangle.

©Stanley Chan 2022. All Rights Reserved.

Marginal PDF

Definition

The **marginal PMF** is defined as

$$p_X(x) = \sum_{y \in \Omega_Y} p_{X,Y}(x,y) \quad \text{and} \quad p_Y(y) = \sum_{x \in \Omega_X} p_{X,Y}(x,y), \tag{3}$$

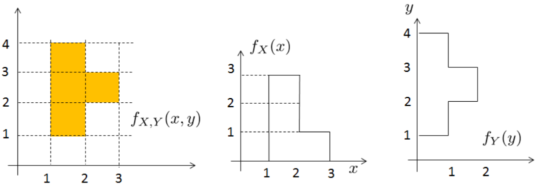
and the **marginal PDF** is defined as

$$f_X(x) = \int_{\Omega_Y} f_{X,Y}(x,y)dy \quad \text{and} \quad f_Y(y) = \int_{\Omega_X} f_{X,Y}(x,y)dx \tag{4}$$

©Stanley Chan 2022. All Rights Reserved.

Example

Example 1. Consider the joint PDF $f_{X,Y}(x, y) = \frac{1}{4}$ shown below. Find the marginal PDFs.



Solution. If we integrate over x and y, then we have

$$f_X(x) = \begin{cases} 3, & \text{if } 1 < x \leq 2, \\ 1, & \text{if } 2 < x \leq 3, \\ 0, & \text{otherwise.} \end{cases} \quad \text{and} \quad f_Y(y) = \begin{cases} 1, & \text{if } 1 < x \leq 2, \\ 2, & \text{if } 2 < x \leq 3, \\ 1, & \text{if } 3 < x \leq 4, \\ 0, & \text{otherwise.} \end{cases}$$

©Stanley Chan 2022. All Rights Reserved.

Example

Example 2. A joint Gaussian random variable (X, Y) has joint PDF

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{((x-\mu_X)^2 + (y-\mu_Y)^2)}{2\sigma^2}\right\}.$$

Find the marginal PDF $f_X(x)$.

Solution.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{((x-\mu_X)^2 + (y-\mu_Y)^2)}{2\sigma^2}\right\} dy \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu_X)^2}{2\sigma^2}\right\} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\mu_Y)^2}{2\sigma^2}\right\} dy \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu_X)^2}{2\sigma^2}\right\}. \end{aligned}$$

©Stanley Chan 2022. All Rights Reserved.

Independence

Definition

If two random variables X and Y are **independent**, then

$$p_{X,Y}(x, y) = p_X(x)p_Y(y), \quad \text{and} \quad f_{X,Y}(x, y) = f_X(x)f_Y(y).$$

Definition

If a sequence of random variables $X_1, \ldots, X_N$ are independent, then their joint PDF (or joint PMF) can be factorized.

$$f_{X_1,\ldots,X_N}(x_1, \ldots, x_N) = \prod_{n=1}^{N} f_{X_n}(x_n). \tag{5}$$

©Stanley Chan 2022. All Rights Reserved.

Example

Example 1. Consider two random variables with a joint PDF given by

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{(x-\mu_X)^2 + (y-\mu_Y)^2}{2\sigma^2}\right\}.$$

Then, from the previous example, we know that

$$f_{X,Y}(x,y) = \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu_X)^2}{2\sigma^2}\right\}}_{f_X(x)} \times \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-\mu_Y)^2}{2\sigma^2}\right\}}_{f_Y(y)}$$

Therefore, the random variables X and Y are independent.

©Stanley Chan 2022. All Rights Reserved.

Example

Example 2. Consider two random variables X and Y with a joint PDF given by

$$
\begin{aligned}
f_{X,Y}(x,y) &\propto \exp\left\{-(x-y)^2\right\} \\
&= \exp\left\{-x^2 + 2xy - y^2\right\} \\
&= \underbrace{\exp\left\{-x^2\right\}}_{f_X(x)} \; \underbrace{\exp\left\{2xy\right\}}_{\text{extra term}} \; \underbrace{\exp\left\{-y^2\right\}}_{f_Y(y)}
\end{aligned}
$$

This PDF cannot be factorized into a product of two marginal PDFs. Therefore, the random variables are dependent.

©Stanley Chan 2022. All Rights Reserved.

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

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

A collection of random variables $X_1, \ldots, X_N$ are called independent and identically distributed (i.i.d.) if

- All $X_1, \ldots, X_N$ are independent;
- All $X_1, \ldots, X_N$ have the same distribution, i.e., $f_{X_1}(x) = \ldots = f_{X_N}(x)$.

Why is i.i.d. so important?

- If a set of random variables are i.i.d., then the joint PDF can be written as a product of PDFs.
- Integrating a joint PDF is not fun. Integrating a product of PDFs is a lot easier.

©Stanley Chan 2022. All Rights Reserved.

Example

Example 1. Let $X_1, X_2, \ldots, X_N$ be a sequence of i.i.d. Gaussian random variables where each X_i has a PDF

$$f_{X_i}(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}.$$

The joint PDF of $X_1, X_2, \ldots, X_N$ is

$$\begin{aligned} f_{X_1,\ldots,X_N}(x_1, \ldots, x_N) &= \prod_{i=1}^{N} \left\{ \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x_i^2}{2}\right\}\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^N \exp\left\{-\sum_{i=1}^{N} \frac{x_i^2}{2}\right\} \end{aligned}$$

©Stanley Chan 2022. All Rights Reserved.

Joint CDF

Definition

Let X and Y be two random variables. The **joint CDF** of X and Y is the function $F_{X,Y}(x,y)$ such that

$$F_{X,Y}(x,y) = \mathbb{P}[X \leq x \;\cap\; Y \leq y]. \tag{6}$$

Definition

If X and Y are discrete, then

$$F_{X,Y}(x,y) = \sum_{y' \leq y} \sum_{x' \leq x} p_{X,Y}(x',y'). \tag{7}$$

If X and Y are continuous, then

$$F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(x',y') dx' dy'. \tag{8}$$

©Stanley Chan 2022. All Rights Reserved.

Example

Example 1. Let X and Y be two independent uniform random variables Uniform$(0, 1)$. Then, the joint CDF is

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) = \int_0^x f_X(x')dx' \int_0^y f_Y(y')dy'$$
$$= \int_0^x 1dx' \int_0^y 1dy' = xy.$$

Example 2. Let X and Y be two independent uniform random variables Gaussian(μ, σ^2). Then, the joint CDF is

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) = \int_{-\infty}^x f_X(x')dx' \int_{-\infty}^y f_Y(y')dy'$$
$$= \Phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\frac{y-\mu}{\sigma}\right),$$

where $\Phi(\cdot)$ is the CDF of the standard Gaussian.

©Stanley Chan 2022. All Rights Reserved.

Properties

Proposition

Let X and Y be two random variables. The **marginal CDF** *is*

$$F_X(x) = F_{X,Y}(x, \infty)$$
$$F_Y(y) = F_{X,Y}(\infty, y).$$

Definition

Let $F_{X,Y}(x, y)$ be the joint CDF of X and Y. Then, the joint PDF can be obtained through

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial y \partial x} F_{X,Y}(x, y).$$

©Stanley Chan 2022. All Rights Reserved.

Summary

- Joint PMF and PDF:

Joint distributions are **high-dimensional** PDF (or PMF or CDF).

- From 1D to 2D:
  - Replace 1D integration to 2D integration
- Independence:

$$f_{X_1,\ldots,X_N}(x_1,\ldots,x_N) = \prod_{n=1}^{N} f_{X_n}(x_n). \tag{9}$$

A collection of random variables $X_1,\ldots,X_N$ are called independent and identically distributed (i.i.d.) if

- All $X_1,\ldots,X_N$ are independent;
- All $X_1,\ldots,X_N$ have the same distribution, i.e., $f_{X_1}(x) = \ldots = f_{X_N}(x)$.

©Stanley Chan 2022. All Rights Reserved.

Questions?

©Stanley Chan 2022. All Rights Reserved.