Probability & Statistics for DS & AI

Joint distributions

Michele Guindani

Summer

What are joint distributions?

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

What are joint distributions?

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

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

What are joint distributions?

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

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

What are joint distributions?

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

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

Notation:

What are joint distributions?

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

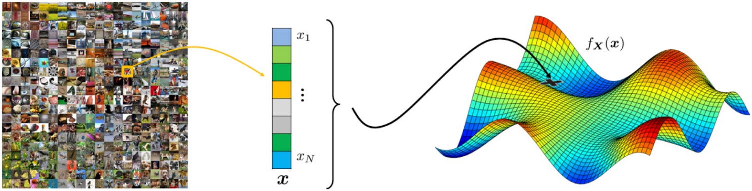
$$\underbrace{f_X(x)}_{\text{one variable}} \Longrightarrow \underbrace{f_{X_1,X_2}(x_1,x_2)}_{\text{two variables}} \Longrightarrow \underbrace{f_{X_1,X_2,X_3}(x_1,x_2,x_3)}_{\text{three variables}}$$
$$\Longrightarrow \dots \Longrightarrow \underbrace{f_{X_1,\dots,X_N}(x_1,\dots,x_N)}_{N \text{ variables}}$$

Notation:

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1,\dots,X_N}(x_1,\dots,x_N)$$

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 $f_X(\boldsymbol{x})$.



Joint PMF

Definition

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

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

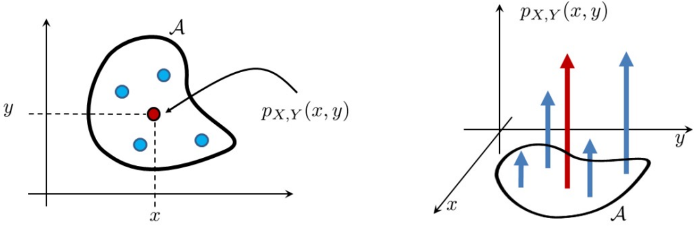


Figure: A joint PMF for a pair of discrete random variables consists of an array of impulses. To measure the size of the event $\mathcal{A}$, we sum all the impulses inside $\mathcal{A}$.

Example (Coin and die)

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

Example (Coin and die)

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					
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:

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

Example (Coin and die - contd.)

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:

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

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

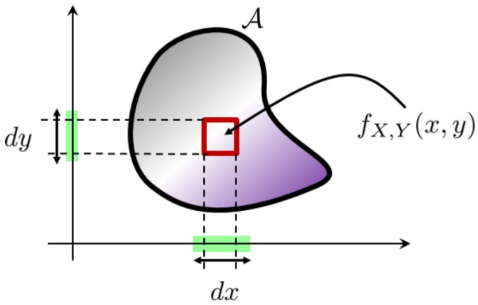
Joint PDF

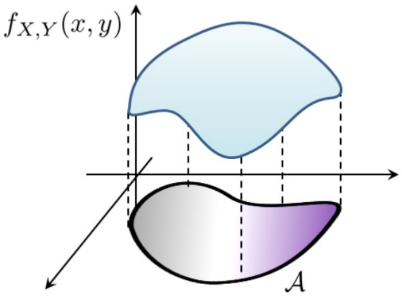
Definition

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:

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

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





Example

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:

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

Suppose $[a, b] \equiv [0, 1]$, $[c, d] \equiv [0.5, 1.5]$, then

$$\mathbb{P}[\mathcal{A}] = \mathbb{P}[0 \leq X \leq 1, \quad 0.5 \leq Y \leq 1.5] = \frac{1}{4}$$

Marginal PMF and PDF

The marginal PMF is defined as

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

and the marginal PDF is defined as

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

Independence of random variables

Definition

If two random variables X and Y are independent, then

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

Independence of random variables

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:

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

Example

- 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 $\mathbb{P}[\mathcal{A}]$

  In this case,

$$f_{X,Y}(x,y) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = f_X(x)\, f_Y(y) \quad 0 \leq x, y \leq 2$$

Example

- 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 $\mathbb{P}[\mathcal{A}]$

  In this case,

$$f_{X,Y}(x,y) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = f_X(x)\, f_Y(y) \quad 0 \leq x, y \leq 2$$

🤔 Why is this important?

Example

- 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 $\mathbb{P}[\mathcal{A}]$

  In this case,

$$f_{X,Y}(x,y) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = f_X(x)\, f_Y(y) \quad 0 \leq x, y \leq 2$$

🤔 Why is this important?

👉 It is easy to compute in Python the probability of events of vectors of random variables (multivariate distributions) if we know the independence structure between the random variables.💞

Example (Python example)

- Let $[a, b] \equiv [0, 1], [c, d] \equiv [0.5, 1.5]$. Compute $\mathbb{P}[\mathcal{A}] = \mathbb{P}[0 \leq X \leq 1, \quad 0.5 \leq Y \leq 1.5]$.

Example (Python example)

- Let $[a, b] \equiv [0, 1], [c, d] \equiv [0.5, 1.5]$. Compute $\mathbb{P}[\mathcal{A}] = \mathbb{P}[0 \leq X \leq 1, \quad 0.5 \leq Y \leq 1.5]$.
- We can use Monte Carlo approximation.

```
import numpy as np
import random
np.random.seed(12345)
nreps=1000000
x= np.random.uniform(low=0, high=2, size=nreps)
y= np.random.uniform(low=0, high=2, size=nreps)

# Probability of A
condition=np.zeros(nreps)
for rep in range(nreps):
condition[rep]=(((x[rep]>0) and (x[rep]<1)) and
                              ((y[rep]>0.5) and (y[rep]<1.5)))  #same line

count=sum(condition)
freq=count/nreps
print(freq)  #0.250293
```

Example

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

Example

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

```
import numpy as np
import random
np.random.seed(12345)
nreps=1000000
x= np.random.uniform(low=0, high=2, size=nreps)
y= np.random.uniform(low=0, high=2, size=nreps)

# Probability of B
sumxy=np.zeros(nreps)
for rep in range(nreps):
sumxy[rep]=x[rep]+y[rep]

freq=sum(sumxy<=2)/nreps
print(freq) #0.4998
```

Not all r.v. are independent!!

- Consider two random variables X and Y with a joint PDF given by

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

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

An interesting case

Example

- A joint Gaussian random variable (X, Y) has a joint PDF given by:

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

  Find the marginal PDFs $f_X(x)$ and $f_Y(y)$

- **Solution:**

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

An interesting case

Example

- A joint Gaussian random variable (X, Y) has a joint PDF given by:

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

Find the marginal PDFs $f_X(x)$ and $f_Y(y)$

- **Solution:**

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

An interesting case

Example

- A joint Gaussian random variable (X, Y) has a joint PDF given by:

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

  Find the marginal PDFs $f_X(x)$ and $f_Y(y)$

- **Solution:**

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

Example

- Recognizing that the last integral is equal to unity because it integrates a Gaussian PDF over the real line, it follows that

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu_X)^2}{2\sigma^2}\right\}$$

- Similarly, we have

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\mu_Y)^2}{2\sigma^2}\right\}$$

- It's immediate to see that $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, hence X, Y are independent.

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.

Example

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

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

Joint CDF

- We can extend the definition of CDF also to vectors of r.v.'s.

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

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

- We won't say anything more about joint CDFs (see textbook)

Joint CDF

- We can extend the definition of CDF also to vectors of r.v.'s.

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

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

- We won't say anything more about joint CDFs (see textbook)

Joint Expectation

- Similarly, for the expectation, we can define the joint expectation:

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

if X and Y are discrete, or

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

if X and Y are continuous.

Covariance

Let X and Y be two random variables. Then the covariance of X and Y is

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[(X - \mu_X)\,(Y - \mu_Y)\right]$$

where $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$.

Covariance

Let X and Y be two random variables. Then the covariance of X and Y is

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[(X - \mu_X)\,(Y - \mu_Y)\right]$$

where $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$.

Remark:

$$\mathrm{Cov}(X, X) = \mathbb{E}\left[(X - \mu_X)\,(X - \mu_X)\right] = \mathrm{Var}[X]$$

Covariance

Let X and Y be two random variables. Then the covariance of X and Y is

$$\text{Cov}(X, Y) = \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right]$$

where $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$.

Remark:

$$\text{Cov}(X, X) = \mathbb{E}\left[(X - \mu_X)(X - \mu_X)\right] = \text{Var}[X]$$

Theorem

Let X and Y be two random variables. Then,

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Interesting properties

- For any X and Y

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + 2\,\mathrm{Cov}(X,Y) + \mathrm{Var}[Y]$$

Correlation coefficient

Definition

Let X and Y be two random variables. The correlation coefficient is

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}}$$

- ρ is always between 0 and 1 , i.e., $1 \leq \rho \leq 1$. This is due to the cosine angle definition.
- When $X = Y$ (fully correlated), $\rho = +1$.
- When $X = -Y$ (negatively correlated), $\rho = -1$.
- When X and Y uncorrelated then $\rho = 0$.

Independence

Theorem

If X and Y are independent, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

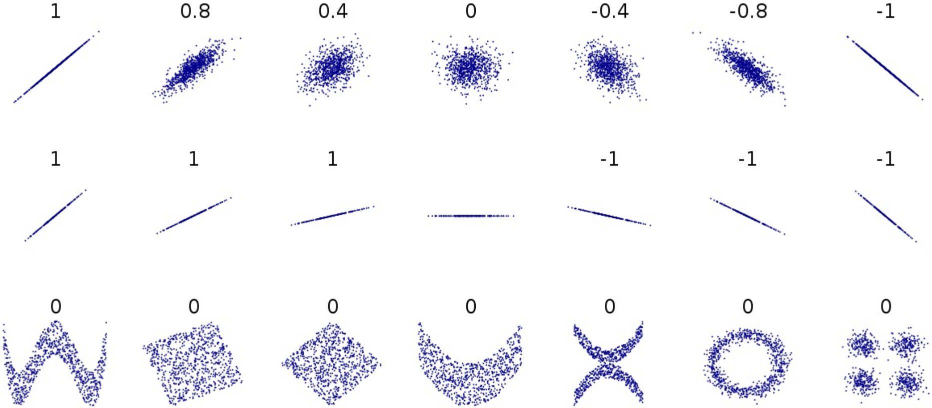
- Consider the following two statements:

a. X and Y are independent;

b. $\text{Cov}(X, Y) = 0$.

⇨ It holds that (a) implies (b), but (b) does not imply (a). Thus, independence is a stronger condition than correlation.

Independence



Ideal vs Empirical

Theory:

$$\rho = \frac{\mathbb{E}[XY] - \mu_X \mu_Y}{\sigma_X \sigma_Y}.$$

Practice:

$$\widehat{\rho} = \frac{\frac{1}{N}\sum_{n=1}^{N} x_n y_n - \overline{x}\,\overline{y}}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})^2}\sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \overline{y})^2}},$$