# Section 5: Joint distributions

STA 35C – Statistical Data Science III

**Instructor:** Akira Horiguchi

Fall Quarter 2025 (Sep 24 – Dec 12)
MWF, 12:10 PM – 1:00 PM, Olson 158
University of California, Davis

# Overview

Based on Chapter 5 of textbook: `https://www.probabilitycourse.com/`

- Contains problems with solutions, and problems without solutions.

1. Introduction
2. For discrete random variables
3. For continuous random variables
4. Covariance and correlation

# Introduction

# Introduction

We are often interested in several random variables that are related to each other.

- A person's height and weight are typically related.
- A person's age and SAT score are typically related.

First we will study two random variables, but easy to extend to $\geq 2$ random variables.

# For discrete random variables

## Definition: Joint probability mass function (PMF)

The *joint PMF* of two discrete random variables $X$ and $Y$ is the function $f_{XY}$ defined as

$$f_{XY}(a,b) = P(X = a, Y = b).$$

Can be expressed as "the probability that $X = a$ and $Y = b$."

- $P(X = a, Y = b)$ can be very different from $P(X = a)\,P(Y = b)$.
  Consider drawing an animal from a population of orange giraffes and purple fish.

The *joint range* for $X$ and $Y$ is defined as

$$R_{XY} = \{(a,b) \mid f_{XY}(a,b) > 0\}.$$

- $R_{XY}$ is always a subset of $R_X \times R_Y$, but $R_X \times R_Y$ might have pairs that $R_{XY}$ does not.

# Marginal PMF

The joint PMF $f_{XY}$ contains all the information regarding the distributions of $X$ and $Y$. We can obtain the *marginal PMF* of $X$ by

$$\underbrace{f_X(a)}_{\text{marginal PMF}} = P(X = a) = \sum_{b \in R_Y} P(X = a, Y = b) = \sum_{b \in R_Y} \underbrace{f_{XY}(a, b)}_{\text{joint PMF}}.$$

Similarly, we can obtain the *marginal PMF* of $Y$ by

$$f_Y(b) = \sum_{a \in R_X} f_{XY}(a, b).$$

E.g. suppose we draw a student from a high school and consider their year and the math class that the student is currently taking.

Consider two random variables $X$ and $Y$ with joint PMF

**Table 5.1 Joint PMF of $X$ and $Y$ in Example 5.1**

| | $Y=0$ | $Y=1$ | $Y=2$ |
|---|---|---|---|
| $X=0$ | $\frac{1}{6}$ | $\frac{1}{4}$ | $\frac{1}{8}$ |
| $X=1$ | $\frac{1}{8}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

$X$ & $Y$ are indep iff $P(X=a, Y=b) = P(X=a)P(Y=b)$ for all $a \in R_X$, $b \in R_Y$

$P(X=0) = \frac{1}{6} + \frac{1}{4} + \frac{1}{8} = \frac{13}{24}$

$P(X=1) = \frac{1}{8} + \frac{1}{6} + \frac{1}{6}$

$P(Y=0) = \frac{1}{6} + \frac{1}{8} = \frac{7}{24}$ " "

1. Find $P(X = 0, Y \leq 1)$. $= P(X=0, Y=0) + P(X=0, Y=1) = \frac{1}{6} + \frac{1}{4} = \frac{1}{3}$
2. Find the marginal PMFs of $X$ and $Y$.
3. Find $P(Y = 1 \mid X = 0)$. $= \frac{P(Y=1, X=0)}{P(X=0)} = \frac{1/4}{13/24} = \frac{6}{13}$
4. Are $X$ and $Y$ independent? NO

$P(X=0, Y=0) = \frac{1}{6}$ $P(X=0)P(Y=0) = \frac{13}{24} \cdot \frac{7}{24}$

# Conditional PMF

Often, we observe the value of a random variable $X$, and we want to update the PMF of another random variable $Y$ whose value has not yet been observed.
E.g., regression of response $Y$ on covariate $X$.

## Definition: Conditional probability mass function (PMF)

For two discrete random variables $X$ and $Y$ with respective marginal PMFs $f_X$ and $f_Y$, the *conditional PMF of Y given X* is the function $f_{Y|X}$ defined as

$$P(Y=b \mid X=a) = f_{Y|X}(b|a) = \boxed{\frac{f_{XY}(a,b)}{f_X(a)}}, \quad \text{for any } a \in R_X \text{ and } b \in R_Y.$$

Similarly (by symmetry), the *conditional PMF of X given Y* is the function $f_{X|Y}$ defined as

$$f_{X|Y}(a|b) = \frac{f_{XY}(a,b)}{f_Y(b)}, \quad \text{for any } a \in R_X \text{ and } b \in R_Y.$$

$$\boxed{P(Y=b \mid X=a) = \frac{P(Y=b, X=a)}{P(X=a)} = \boxed{\frac{f_{XY}(a,b)}{f_X(a)}}}$$

# Conditional Expectation

Often want to find the mean of a conditional PMF.
Similar to finding a "regular" expected value, but replace PMF with the conditional PMF.

## Definition: Conditional expectation

For two discrete random variables $X$ and $Y$, the *conditional expectation of $Y$ given $X = a$* is

$$\boxed{E[Y \mid X = a] = \sum_{b \in R_Y} \overset{\text{value}}{b} \cdot \underbrace{f_{Y|X}(b|a)}_{\text{weight}}.}$$

still a weighted average

Similarly (by symmetry), the *conditional expectation of $X$ given $Y = b$* is

$$E[X \mid Y = b] = \sum_{a \in R_X} a \cdot f_{X|Y}(a|b).$$

# For continuous random variables

### Definition: Joint probability density function (PDF)

Two random variables $X$ and $Y$ are *jointly continuous* if there exists a nonnegative function $f_{XY} \colon \mathbb{R}^2 \to \mathbb{R}$, such that, for any set $A \subset \mathbb{R}^2$, we have

$$P\Big((X, Y) \in A\Big) = \iint_A f_{XY}(a, b)\,\mathrm{d}a\,\mathrm{d}b.$$

The function $f_{XY}$ is called the *joint PDF* of $X$ and $Y$.

The *joint range* for $X$ and $Y$ is defined as

$$R_{XY} = \big\{(a, b) \mid f_{XY}(a, b) > 0\big\}.$$

The joint PDF $f_{XY}$ contains all the information regarding the distributions of $X$ and $Y$.
We can obtain the *marginal PDF* of $X$ by

sum over possible values of Y

$$\underbrace{\overset{f_X}{f_X(a)}}_{} = \underbrace{\int_{-\infty}^{\infty} f_{XY}(a, b)\,\mathrm{d}b}_{}, \quad \text{for all } a.$$

Similarly, we can obtain the *marginal PDF* of $Y$ by

$$f_Y(b) = \int_{-\infty}^{\infty} f_{XY}(a, b)\,\mathrm{d}a, \quad \text{for all } b.$$

## Definition: Conditional probability density function (PDF)

For two jointly continuous random variables $X$ and $Y$ with respective marginal PDFs $f_X$ and $f_Y$, the *conditional PDF of Y given X* is the function $f_{Y|X}$ defined as

$$f_{Y|X}(b|a) = \frac{f_{XY}(a,b)}{f_X(a)} \quad \text{for any } a \in R_X \text{ and } b \in R_Y.$$

Similarly (by symmetry), the *conditional PDF of X given Y* is the function $f_{X|Y}$ defined as

$$f_{X|Y}(a|b) = \frac{f_{XY}(a,b)}{f_Y(b)} \quad \text{for any } a \in R_X \text{ and } b \in R_Y.$$

# Conditional expectation and variance

For these definitions, suppose $X$ and $Y$ are two jointly continuous random variables.

## Definition: Conditional expectation

The *conditional expectation of $Y$ given $X = a$* is

$$E[Y \mid X = a] \;=\; \int_{-\infty}^{\infty} b \cdot f_{Y|X}(b|a)\,\mathrm{d}b.$$

## Definition: Conditional variance

The *conditional variance of $Y$ given $X = a$* is

$$Var(Y \mid X = a) \;=\; E\left[Y^2 \mid X = a\right] - \left(E[Y|X = a]\right)^2$$

$$Var(Y) = E[Y^2] - (EY)^2$$

# Covariance and correlation

# Motivation

A joint PMF/PDF characterizes the relationship between two random variables $X$ and $Y$.

- Simpler characterizations are the *covariance* and *correlation* of $X$ and $Y$.

## Definition 4: Covariance $Cov(\cdot,\cdot)$

Let $X, Y$ be random variables with $E(X^2), E(Y^2) < \infty$. Then the *covariance* between $X$ and $Y$ is defined as

$$Cov(X,Y) := E\big[(X - EX)(Y - EY)\big]. \tag{1}$$

- The covariance between $X$ and $Y$ can also be calculated as

$$Cov(X,Y) = E(XY) - E(X)E(Y). \tag{2}$$

- We say $X$ and $Y$ are *uncorrelated* if $Cov(X,Y) = 0$. Then $X$ and $Y$ have no linear relationship, and $E(XY) = E(X)E(Y)$.
- $Cov(X,Y) > 0$ indicate a positive linear relationship between $X$ and $Y$.
- $Cov(X,Y) < 0$ indicate a negative linear relationship between $X$ and $Y$.
- Covariance is symmetric: $Cov(X,Y) = Cov(Y,X)$.

From either (1) or (2), we can deduce that $Cov(X,X) = Var(X)$.

- In general, we can use covariance to prove/deduce many results for variance.

## Definition 5: Correlation coefficient $\rho(\cdot,\cdot)$

Let $X, Y$ be random variables with $E(X^2), E(Y^2) < \infty$. Then, the *correlation coefficient* between $X$ and $Y$ is defined as, provided $Var(X) > 0$ and $Var(Y) > 0$,

$$\rho(X,Y) := \frac{Cov(X,Y)}{\underbrace{\sqrt{Var(X)}}\underbrace{\sqrt{Var(Y)}}} \in [-1,1]. \tag{3}$$

- $\rho(X,Y) = 0 \Rightarrow$ between $X$ and $Y$ is no linear relationship.
- $\rho(X,Y) = -1$ $(1) \Rightarrow$ all values of $X$ and $Y$ lie on a line with negative (positive) slope.
- If $\rho(X,Y)$ is close to -1 (1), there is a strong negative (positive) linear relationship between $X$ and $Y$.

# Functions of Two Random Variables

# Expected value

Suppose we are interested in the expected value of a function $Z = g(X, Y)$ of random variables $X$ and $Y$, where $g: \mathbb{R}^2 \to \mathbb{R}$ is a known function.

- Examples: $Z = XY$, where $X$ and $Y$ are the random width and length of a rectangle, respectively.
- To compute the expected value $E[Z]$, we can use LOTUS:

$$E[g(X, Y)] = \sum_{(a,b)\in R_{XY}} g(a, b) f_{XY}(a, b) \tag{4}$$

  - This holds for both continuous and discrete random variables.
  - This allows us to compute $E[Z]$ without computing the PMF/PDF of $Z$.
- A useful special case to know:

## Properties of independent random variables

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

$$E[XY] = E[X]\,E[Y]. \tag{5}$$

(Independence really helps us simplify calculations!)