CS 440/ECE 448 Lecture 2: Random Variables

Mark Hasegawa-Johnson, 1/2023

Outline

- Notation: Probability, Probability Mass, Probability Density
- Jointly random variables
- Conditional Probability and Independence
- Expectation
- Covariance Matrix

Notation: Probability

If an experiment is run an infinite number of times, the probability of event A is the fraction of those times on which event A occurs.

Axiom 1: every event A has a non-negative probability.

$$\Pr(A) \geq 0$$

Axiom 2: If an event Ω always occurs, we say it has probability 1.

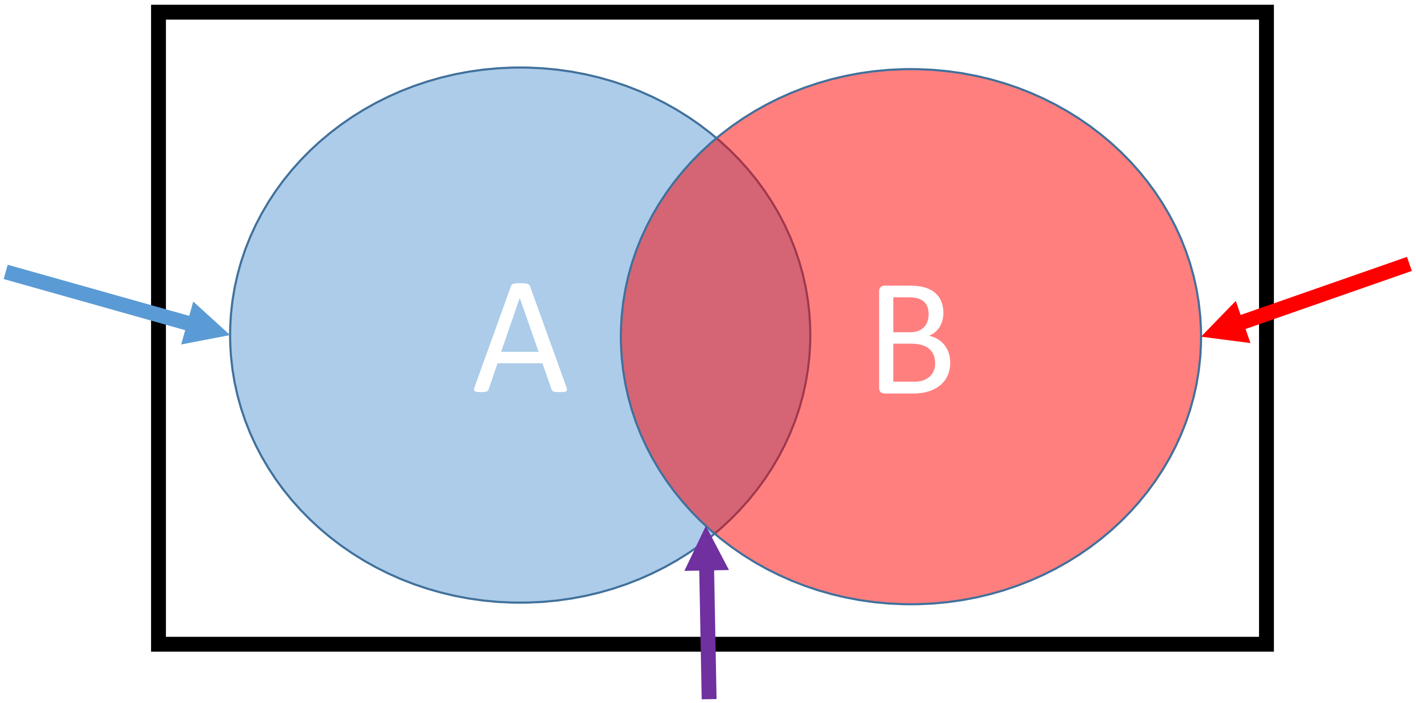
$$\Pr(\Omega) = 1$$

Axiom 3: probability measures behave like set measures.

$$\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A \wedge B)$$

Axiom 3: probability measures behave like set measures.

Area of the whole rectangle is $P(\Omega) = 1$.



Area of this circle is $P(A)$.

Area of this circle is $P(B)$.

Area of their intersection is $P(A \wedge B)$.
Area of their union is $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Notation: Random Variables

A **random variable** is a function that summarizes the output of an experiment. We use **capital letters** to denote random variables.

- Example: every Friday, Maria brings a cake to her daughter's pre-school. X is the number of children who eat the cake.

We use a **small letter** to denote a particular **outcome** of the experiment.

- Example: for the last three weeks, each week, 5 children had cake, but this week, only 4 children had cake. Estimate $P(X = x)$ for all possible values of x.

Notation: $P(X = x)$ is a number, but $P(X)$ is a distribution

- $P(X = 4)$ or $P(4)$ is the probability mass or probability density of the outcome "$X = 4$." For example:

$$P(X = 4) = \frac{1}{4}$$

- $P(X)$ is the complete **distribution**, specifying $P(X = x)$ for all possible values of x. For example:

$P(X) =$

x	4	5
$P(x)$	$\frac{1}{4}$	$\frac{3}{4}$

Discrete versus Continuous RVs

- X is a **discrete random variable** if it can only take countably many different values.
  - Example: X is the number of people living in a randomly selected city
    $$X \in \{1,2,3,4, \ldots\}$$
  - Example: X is the first word on a randomly selected page
    $$X \in \{\text{the, and, of, bobcat}, \ldots\}$$
  - Example: X is the next emoji you will receive on your cellphone
    $X \in \{$😀, 😃, 😄, 😁, 😆, 😅, 🤣, …$\}$
- X is a **continuous random variable** if it can take uncountably many different values
  - Example: X is the energy of the next object to collide with Earth
    $$X \in \mathbb{R}^+ \text{ (the set of all positive real numbers)}$$

Probability Mass Function (pmf) is a type of probability

- If X is a **discrete random variable,** then $P(X)$ is its **probability mass function (pmf)**.
- A probability mass is just a probability. $P(X = x) = \Pr(X = x)$ is the just the probability of the outcome "$X = x$." Thus:

$$0 \leq P(X = x)$$

$$1 = \sum_x P(X = x)$$

Probability Density Function (pdf) is NOT a probability

- If X is a **density random variable,** then $P(X)$ is its **probability density function (pdf)**.
- A probability density is NOT a probability. Instead, we define it as a density ($P(X = x) = \frac{d}{dx}\Pr(X \leq x)$). Thus:

$$0 \leq P(X = x)$$
$$1 = \int_{-\infty}^{\infty} P(X = x)dx$$

Outline

- Notation: Probability, Probability Mass, Probability Density
- Jointly random variables
- Conditional Probability and Independence
- Expectation
- Covariance Matrix

Jointly Random Variables

- Two or three random variables are “jointly random” if they are both outcomes of the same experiment.
- For example, here are the temperature (Y, in °C), and precipitation (X, symbolic) for six days in Urbana:

	X=Temperature (°C)	Y=Precipitation
January 11	4	cloud
January 12	1	cloud
January 13	-2	snow
January 14	-3	cloud
January 15	-3	clear
January 16	4	rain

Joint Distributions

Based on the data on prev slide, here is an estimate of the joint distribution of these two random variables:

P(X=x,Y=y)		Y			
		snow	rain	cloud	clear
X	-3	0	0	1/6	1/6
	-2	1/6	0	0	0
	1	0	0	1/6	0
	4	0	1/6	1/6	0

Notation: Vectors and Matrices

- A normal-font capital letter is a random variable, which is a function mapping from the outcome of an experiment to a measurement
- A normal-font small letter is a scalar instance
- A boldface small letter is a vector instance
- A boldface capital letter is a matrix instance

Notation: Vectors and Matrices

$P(X = \boldsymbol{x})$ is the probability that random variable X takes the value of the vector $\boldsymbol{x}$. This is just a shorthand for the joint distribution of $x_1, \cdots, x_n$:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \qquad P(X = \boldsymbol{x}) = P(X_1 = x_1, \cdots, X_n = x_n)$$

Similarly, for a random matrix, we could write:

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,n} \end{bmatrix}, \qquad P(X = \boldsymbol{X}) = P\big(X_{1,1} = x_{1,1}, \ldots, X_{m,n} = x_{m,n}\big)$$

Marginal Distributions

Suppose we know the joint distribution $P(X, Y)$. We want to find the two **marginal distributions** $P(X)$:

- If the unwanted variable is discrete, we marginalize by adding:

$$P(X) = \sum_{y} P(X, Y = y)$$

- If the unwanted variable is continuous, we marginalize by integrating:

$$P(X) = \int P(X, Y = y) dy$$

Marginal Distributions

Here are the marginal distributions of the two weather variables:

$P(X,Y)$	**snow**	**rain**	**cloud**	**clear**
-3	0	0	1/6	1/6
-2	1/6	0	0	0
1	0	0	1/6	0
4	0	1/6	1/6	0

$P(X)$
1/3
1/6
0
1/3

$P(Y)$	1/6	1/6	1/2	1/6

Outline

- Notation: Probability, Probability Mass, Probability Density
- Jointly random variables
- Conditional Probability and Independence
- Expectation
- Covariance Matrix

Joint and Conditional distributions

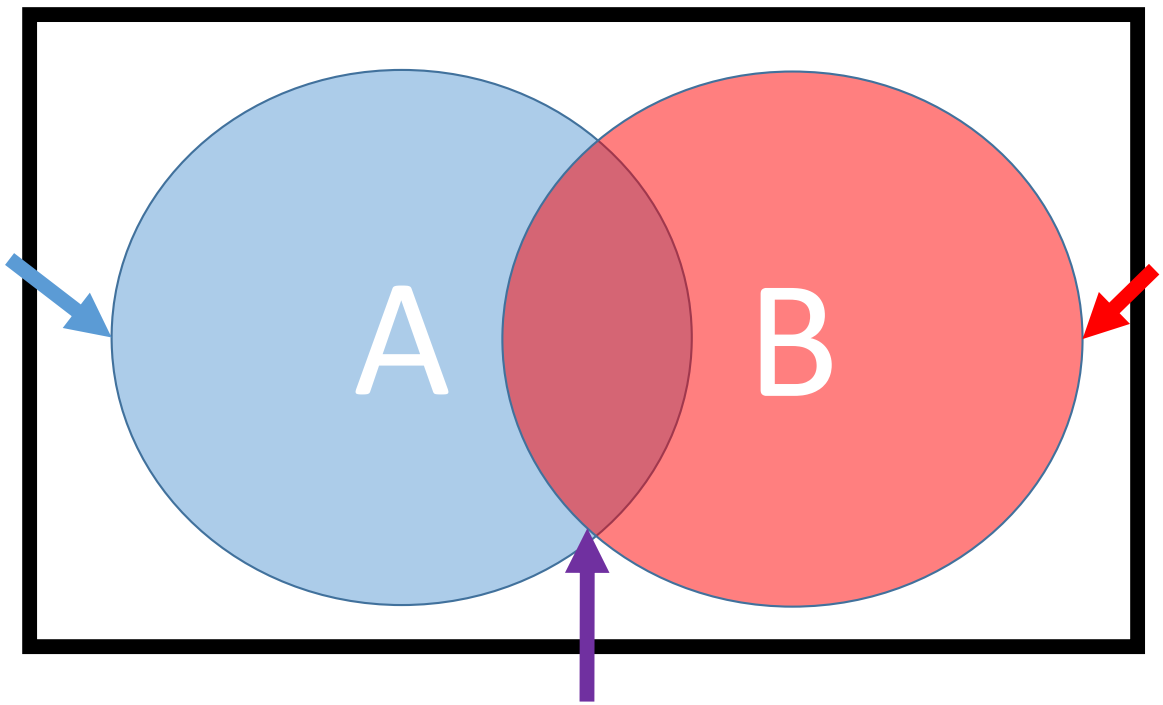
- $P(X, Y)$ is the probability (or pdf) that $X = x$ and $Y = y$, over all x and y. This is called their **joint distribution**.

- $P(Y|X)$ is the probability (or pdf) that $Y = y$ happens, given that $X = x$ happens, over all x and y. This is called the **conditional distribution** of Y given X.

Joint probabilities are usually given in the problem statement

Area of the whole rectangle is $\Pr(\Omega) = 1$.

Suppose $\Pr(A) = 0.4$



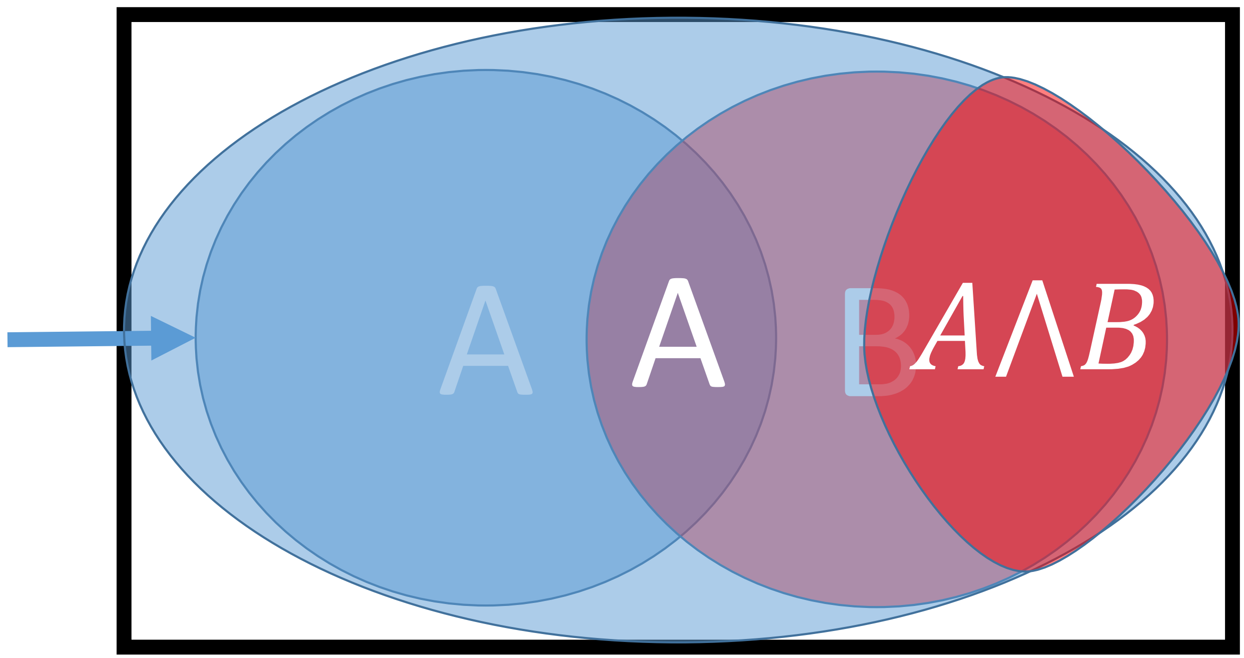
Suppose $\Pr(B) = 0.2$

Suppose $\Pr(A \wedge B) = 0.1$

Conditioning events change our knowledge!
For example, given that A is true…

Most of the events in this rectangle are no longer possible!

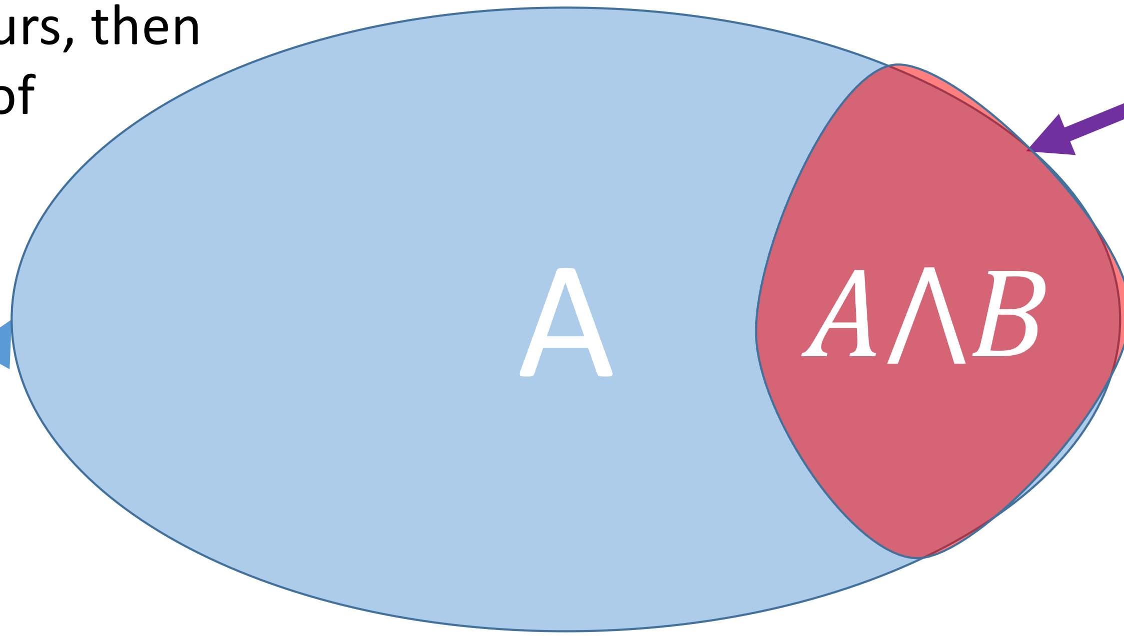
Only the events inside this circle are now possible.



Conditioning events change our knowledge!
For example, given that A is true…

If A always occurs, then by the axioms of probability, the probability of A=T is 1.
We can say that

$\Pr(A|A) = 1.$



The probability of B, given A, is the size of the event $A \wedge B$, expressed as a fraction of the size of the event A:

$$\Pr(B|A) = \frac{\Pr(A \wedge B)}{\Pr(A)}$$

Joint and Conditional distributions of random variables

- $P(X, Y)$ is the **joint probability distribution** over all possible outcomes $P(X = x, Y = y)$.
- $P(X|Y)$ is the **conditional probability distribution** of outcomes $P(X = x|Y = y)$.
- The **conditional** is the **joint** divided by the **marginal:**

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

Conditional is the joint divided by the marginal:

$$P(X|Y = \text{cloud}) = \frac{P(X, Y = \text{cloud})}{P(Y = \text{cloud})} = \frac{\begin{bmatrix}\frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6}\end{bmatrix}}{1/2}$$

	snow	**rain**	**cloud**	**clear**
-3	0	0	1/6	1/6
-2	1/6	0	0	0
1	0	0	1/6	0
4	0	1/6	1/6	0

P(X|Y=cloud)
1/3
0
1/3
1/3

Joint = Conditional×Marginal

$$P(X,Y) = P(X|Y)P(Y)$$

Independent Random Variables

Two random variables are said to be independent if:

$$P(X|Y) = P(X)$$

In other words, knowing the value of Y tells you nothing about the value of X.

… and a more useful definition of independence…

Plugging the definition of independence,

$$P(X|Y) = P(X),$$

…into the "Joint = Conditional×Marginal" equation,

$$P(X,Y) = P(X|Y)P(Y)$$

…gives us a more useful definition of independence.

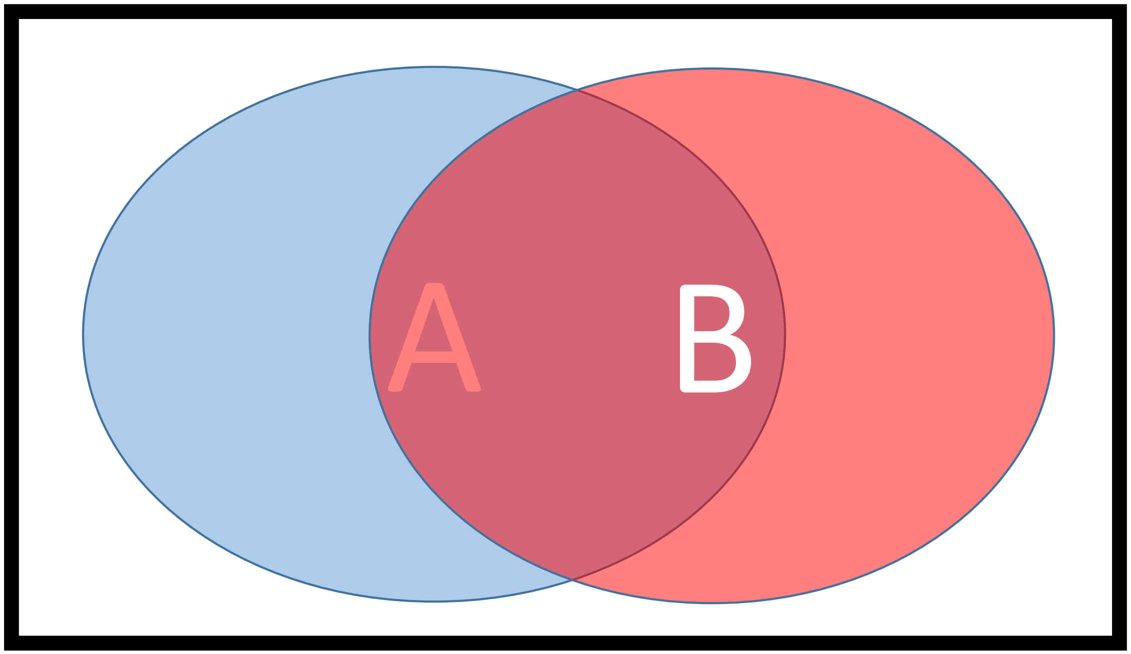
<u>Definition of Independence</u>: Two random variables, X and Y, are independent if and only if

$$P(X,Y) = P(X)P(Y)$$

Independent events

Independent events occur with equal probability, regardless of whether or not the other event has occurred:

$$\Pr(A|B) = \Pr(A)$$
$$\Pr(A \land B) = \Pr(A)\Pr(B)$$



Outline

- Notation: Probability, Probability Mass, Probability Density
- Jointly random variables
- Conditional Probability and Independence
- Expectation
- Covariance Matrix

Expectation

- The expected value of a function is its weighted average, weighted by its pmf or pdf.
- If X and Y are discrete, then

$$E[f(X,Y)] = \sum_{x,y} f(x,y)P(X = x, Y = y)$$

- If X is continuous, then

$$E[f(X,Y)] = \iint_{-\infty}^{\infty} f(x,y)P(X = x, Y = y)dxdy$$

Quiz question

Go to https://us.prairielearn.com/pl/course_instance/129874/

Take the quiz called “20-Jan”

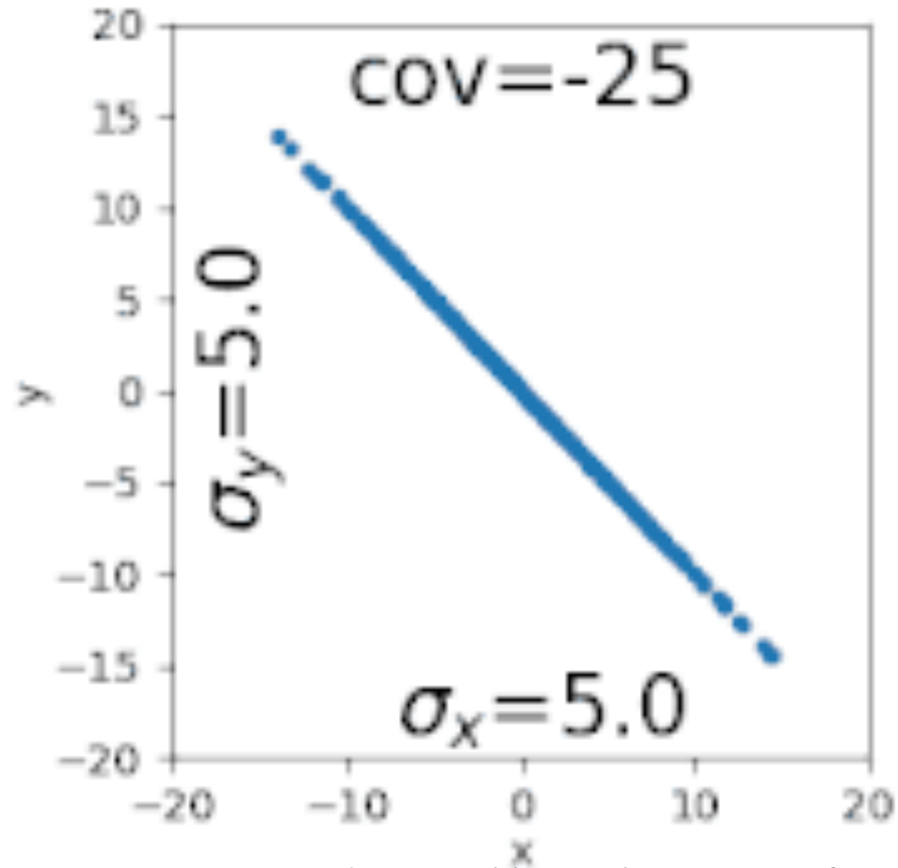
Outline

- Notation: Probability, Probability Mass, Probability Density
- Jointly random variables
- Conditional Probability and Independence
- Expectation
- Covariance Matrix

Covariance

The covariance of two random variables is the expected product of their deviations:

$$\text{Covar}(X, Y) \\ = E[(X - E[X])(Y - E[Y])]$$



Two zero-mean random variables, with variances of 25, and with various values of covariance.
Public domain image, https://commons.wikimedia.org/wiki/File:Varianz.gif

Covariance Matrix

Suppose $X = [X_1, \dots, X_n]^T$ is a random vector. Its matrix of variances and covariances (a.k.a. covariance matrix) is

$$\Sigma = E[(X - E[X])(X - E[X])^T] = \begin{bmatrix} \mathrm{Var}(X_1) & \cdots & \mathrm{Covar}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \mathrm{Covar}(X_1, X_n) & \cdots & \mathrm{Var}(X_n) \end{bmatrix}$$

$$= \begin{bmatrix} E[(X_1 - E[X_1])^2] & \cdots & E[(X_1 - E[X_1])(X_n - E[X_n])] \\ \vdots & \ddots & \vdots \\ E[(X_1 - E[X_1])(X_n - E[X_n])] & \cdots & E[(X_n - E[X_n])^2] \end{bmatrix}$$

Summary

- Probability Mass and Probability Density

$$P(X = x) = \Pr(X = x) \quad \dots \quad \text{or} \quad \dots \quad P(X = x) = \frac{d}{dx}\Pr(X \le x)$$

- Jointly random variables

$$P(X = \boldsymbol{x}) = P(X_1 = x_1, \cdots, X_n = x_n)$$

- Conditional Probability and Independence

$$P(X, Y) = P(X|Y)P(Y)$$
$$P(X|Y) = P(X) \Leftrightarrow P(X, Y) = P(X)P(Y)$$

- Expectation

$$E[f(X, Y)] = \sum_{x,y} f(x, y)P(X = x, Y = y) \quad \dots \quad \text{or} \quad \dots$$
$$E[f(X, Y)] = \iint_{-\infty}^{\infty} f(x, y)P(X = x, Y = y)dxdy$$

- Mean, Variance and Covariance

$$\Sigma = E[(X - E[X])(X - E[X])^T]$$