CS 440/ECE 448 Lecture 2: Random Variables

Mark Hasegawa-Johnson, 1/2025

Outline

- Probability
- Random Variables
- Jointly random variables
- Conditional Probability and Independence

Notation: Probability

If an experiment is run in an infinite number of parallel universes, the probability of event A is the fraction of those universes in which event A occurs.

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

$$\mathrm{P}(A) \geq 0$$

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

$$\mathrm{P}(\Omega) = 1$$

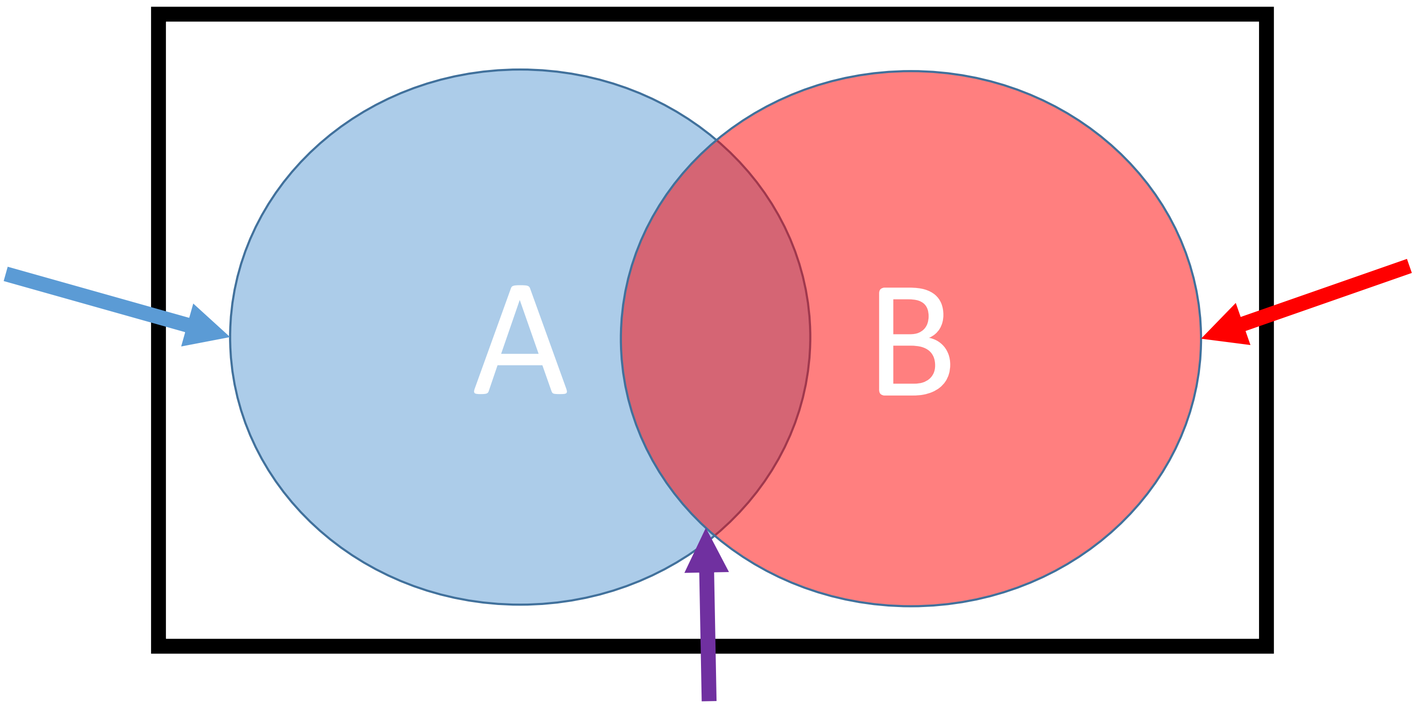
Axiom 3: probability measures behave like set measures.

$$\mathrm{P}(A \vee B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \wedge B)$$

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

Outline

- Probability
- Random Variables
- Jointly random variables
- Conditional Probability and Independence

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

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

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

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

$P(X) =$

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

Domain and Cardinality

- $\mathcal{X}$ is the domain of X, i.e., the set of its possible values.

$$\sum_{x \in \mathcal{X}} P(X = x) = 1$$

- $|\mathcal{X}|$ is the cardinality of X, i.e., the number of possible values
- The probability of an "average" outcome is $P(X = x) = \frac{1}{|\mathcal{X}|}$

Expectation

The expected value of a function is its probability-weighted average.

$$E[f(X)] = \sum_{x \in \mathcal{X}} f(x) P(X = x)$$

What’s it good for?

- How do you use probability in your life?
- Why does an AI need to know about probability?

Outline

- Probability
- Random Variables
- Jointly random variables
- Conditional Probability and Independence

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 (x, in °C), and precipitation (y, 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 Distribution

Based on the data on previous 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

Marginal Distribution

Suppose we know the joint distribution $P(X, Y)$. The **marginal distribution** is $P(X)$:

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

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

- Probability
- Random Variables
- Jointly random variables
- Conditional Probability and Independence

Joint and Conditional distributions

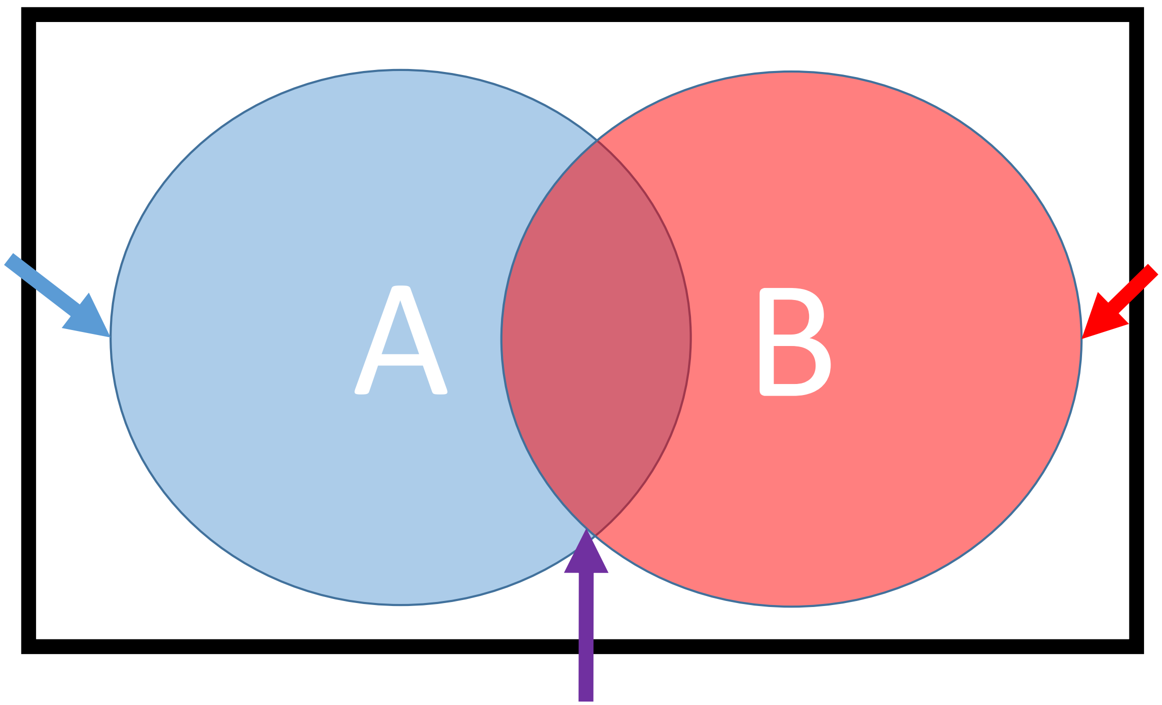
- **Joint distribution** $P(X = x, Y = y)$ is the probability that $X = x$ and $Y = y$.

- **Conditional distribution** $P(Y = y|X = x)$ is the probability that $Y = y$ given that $X = x$.

Joint probabilities are usually given in the problem statement

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

Suppose $\Pr(A) = 0.4$

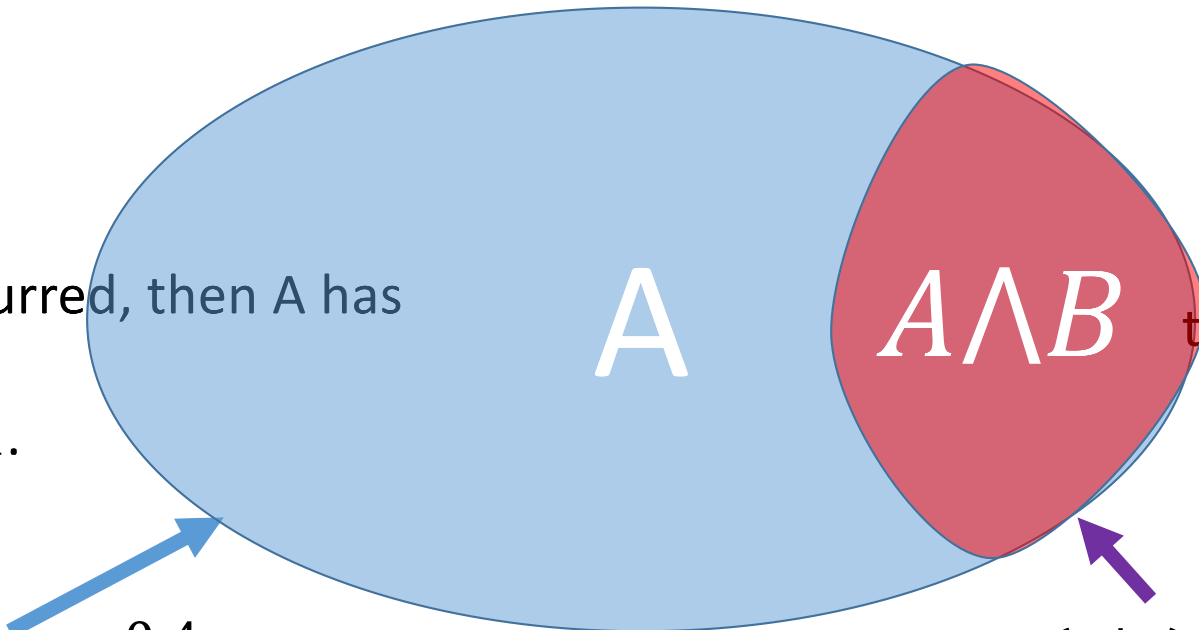


Suppose $\Pr(B) = 0.2$

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

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

If A has occurred, then A has occurred.
$P(A|A) = 1$.
We can say that



$$\mathrm{P}(A|A) = \frac{0.4}{0.4} = 1$$

If A has occurred, then the only way for B to occur is if both A and B occur:

$$\Pr(B|A) = \frac{\Pr(A \wedge B)}{\Pr(A)} = \frac{0.1}{0.4}$$

Joint and Conditional distributions

- **Joint distribution** $P(X = x, Y = y)$ is the probability that $X = x$ and $Y = y$.

- **Conditional distribution** $P(Y = y|X = x)$ is the probability that $Y = y$ given that $X = x$.

- **Jokes are clever manipulations**: Joint = conditional $\times$ marginal

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

This equation is sometimes called Bayes' rule. It is also sometimes called the definition of conditional probability.

Independent Random Variables

Two random variables are said to be independent if:

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

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,

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

…into "jokes are clever manipulations:"

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

…gives us a more useful definition of independence.

Definition of Independence: Two random variables, X and Y, are independent if and only if

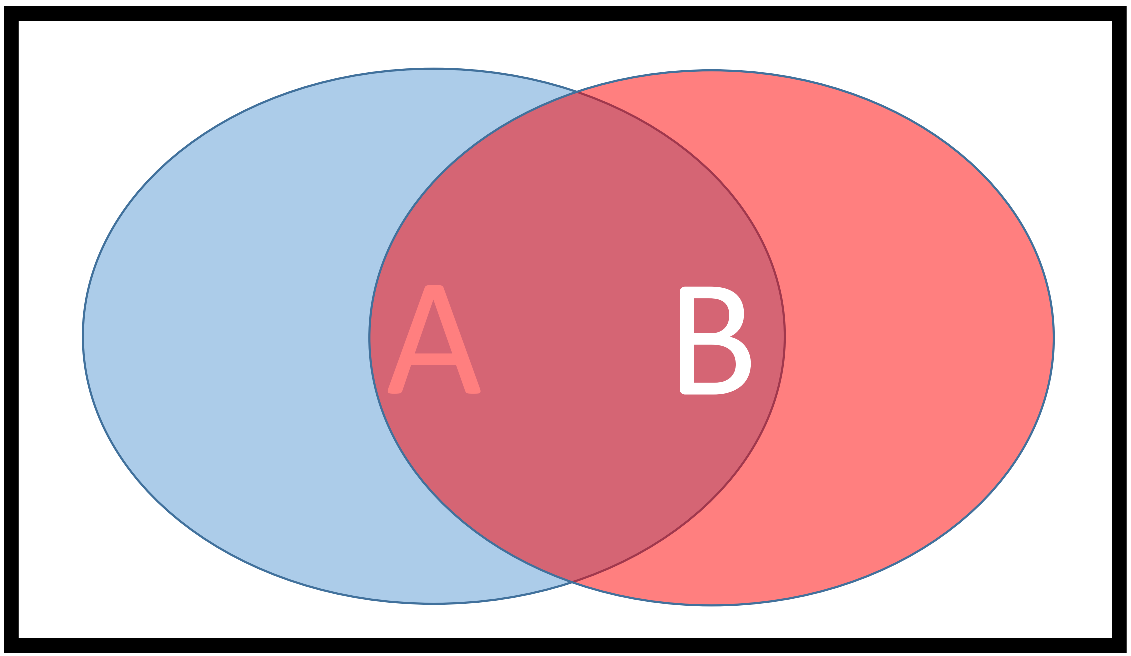
$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

Independent events

Independent events occur with equal probability, regardless of whether the other event has occurred:

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

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



Quiz question

Go to https://us.prairielearn.com/pl/course_instance/174920/assessment/2503895

Take the quiz called “24-Jan”

Summary

- Axioms of probability:

$$\mathrm{P}(A) \geq 0$$

$$\mathrm{P}(\Omega) = 1$$

$$\mathrm{P}(A \vee B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \wedge B)$$

- Domain of a random variable:

$$\sum_{x \in \mathcal{X}} P(X = x) = 1$$

- Jokes are clever manipulations (Joint = conditional times marginal):

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

- Independence:

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

- Expectation:

$$E[f(X)] = \sum_{x \in \mathcal{X}} f(x)P(X = x)$$