02462 – Signals and data

Technical University of Denmark,
DTU Compute, Institut for Matematik og Computer Science.

Overview

Introduction to Signals and Data

DTU

Ugeskema for 2. semester (jf. det overordnede studieforløb, ret til ændringer forbeholdes)

	MANDAG	TIRSDAG	ONSDAG	TORSDAG	FREDAG
8-12	10022 **Fysik 1** Lyngby Campus		01005 **Matematik 1** Lyngby Campus 10-12 (C-version)		02464 **Kunstig intelligens og menneskelig kognition** Lyngby Campus
12-13	Pause	Pause	Pause	Pause	Pause
13-17	02403 **Introduktion til matematisk statistik** Lyngby Campus (Juni)	02462 **Signaler og data** Lyngby Campus	01005 **Matematik 1** Lyngby Campus (C-version)		01005 **Matematik 1** Lyngby Campus (C-version)

General course objectives

To provide the participants knowledge of:

- ⋆ Signal and data types used in intelligent systems
- ⋆ Representations of signals with a focus on images, sound, video, and text
- ⋆ Signal processing and probabilistic methods to enhance signals and suppres noise and confounders
- ⋆ Visualization tools for debugging and communication of signal analysis
- ⋆ Python as a tool for sensing, representation and analysis of signals

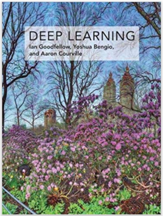
Course plan

3 weeks	• Probabilities and statistics
3 weeks	• Representation
3 weeks	• Signal models (audio, images, text)
3 weeks	• Deep Learning
1 week	• Recap, exam preparation
	• Oral exam

Probability and probability rules

Reading material for week 1-3

Deep Learning, by Ian Goodfellow et al.



free html-version at `http://deeplearningbook.org/`

Read chapter 3, in particular:

- 3.1 - 3.9 Intro and probability rules
- 3.9.3 Normal distributions
- 3.11 Bayes' rule

Use of probability theory in AI

"Probability theory is a mathematical framework for representing uncertain statements. It provides a means of quantifying uncertainty as well as axioms for deriving new uncertain statements.
In artificial intelligence applications, we use probability theory in two major ways. ***First****, the laws of probability tell us how AI systems should reason, so we design our algorithms to compute or approximate various expressions derived using probability theory.* ***Second****, we can use probability and statistics to theoretically analyze the behavior of proposed AI systems."*
(Deep Learning, chapter 3)

What is a probability

A probability is a number between 0 and 1 used to represent

Degree of belief How likely you think some event is (rational belief according to some model)

Frequency If you repeat an experiment, how often does some event occur (on average)

What is a probability

A probability is a number between 0 and 1 used to represent

- **Degree of belief** How likely you think some event is (rational belief according to some model)
- **Frequency** If you repeat an experiment, how often does some event occur (on average)

Probability of an event

we define the probability of an event as the limit of a frequency in a sampling process.

What is a probability

A probability is a number between 0 and 1 used to represent

Degree of belief How likely you think some event is (rational belief according to some model)

Frequency If you repeat an experiment, how often does some event occur (on average)

Probability of an event

we define the probability of an event as the limit of a frequency in a sampling process.

Example:
Flip a coin, count the number of heads (N_{head}):

$$\mathbb{P}(head) = \lim_{N \to \infty} \frac{N_{head}}{N}$$

Terminology

Experiment	Activity with an observable result
Trials	Repetition of experiment
Outcome	Result of each trial
Sample space	Set of all possible outcomes
Sample points	Elements in the sample space
Event	Subset of the sample space

Professor Serna. Introduction to probability. `https://bit.ly/1iB1DS8`

Terminology

Experiment	Activity with an observable result
Trials	Repetition of experiment
Outcome	Result of each trial
Sample space	Set of all possible outcomes
Sample points	Elements in the sample space
Event	Subset of the sample space

Example:
Rolling a six-sided die.

$$S = \{\epsdice{1}, \epsdice{2}, \epsdice{3}, \epsdice{4}, \epsdice{5}, \epsdice{6}\}$$

Professor Serna. Introduction to probability. `https://bit.ly/1iB1DS8`

Terminology

Experiment	Activity with an observable result
Trials	Repetition of experiment
Outcome	Result of each trial
Sample space	Set of all possible outcomes
Sample points	Elements in the sample space
Event	Subset of the sample space

Example:
Rolling a six-sided die.

$$S = \{\epsdice{1}, \epsdice{2}, \epsdice{3}, \epsdice{4}, \epsdice{5}, \epsdice{6}\}$$

An event is a subspace of S:

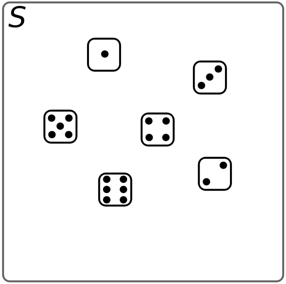
$$even = \{\epsdice{2}, \epsdice{4}, \epsdice{6}\}$$

Professor Serna. Introduction to probability. `https://bit.ly/1iB1DS8`

Probabilities

Example:

$$S = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$$

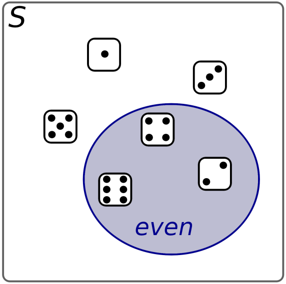


Probabilities

Example:

$S = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$

$even = \{$⚁, ⚃, ⚅$\}$

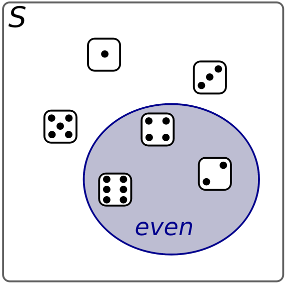


Probabilities

Example:

$$S = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$$

$$even = \{⚁, ⚃, ⚅\}$$



When all elements in the sampling space have equal probability; what is the probability:

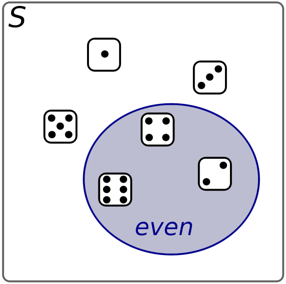
$\mathbb{P}(even)$

Probabilities

Example:

$$S = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$$

$$even = \{⚁, ⚃, ⚅\}$$



When all elements in the sampling space
have equal probability; what is the probability:

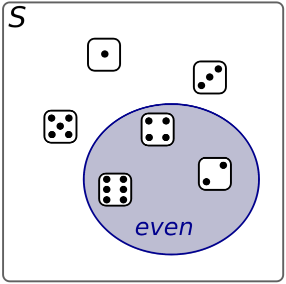
$$\mathbb{P}(even) = \frac{\#\text{elements in subset}}{\#\text{elements in sampling space}}$$

Probabilities

Example:

$$S = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$$

$$even = \{⚁, ⚃, ⚅\}$$



When all elements in the sampling space
have equal probability; what is the probability:

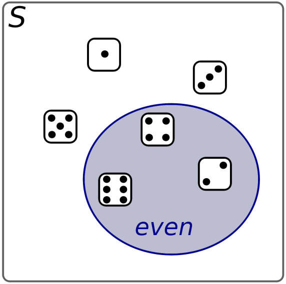
$$\mathbb{P}(even) = \frac{\text{\#elements in subset}}{\text{\#elements in sampling space}} = \frac{3}{6} = 50\%$$

Probabilities

Example:

$$S = \{\text{⚀}, \text{⚁}, \text{⚂}, \text{⚃}, \text{⚄}, \text{⚅}\}$$

$$even = \{\text{⚁}, \text{⚃}, \text{⚅}\}$$



When all elements in the sampling space have equal probability; what is the probability:

$$\mathbb{P}(even) = \frac{\#\text{elements in subset}}{\#\text{elements in sampling space}} = \frac{3}{6} = 50\%$$

As limit of frequency in a sampling process:

$$\mathbb{P}(even) = \lim_{N\to\infty} \frac{N_{\text{⚁}} + N_{\text{⚃}} + N_{\text{⚅}}}{N}$$

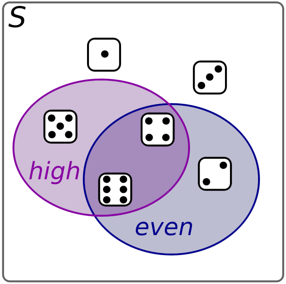
Conditional probability

Example:

$$S = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$$

$$even = \{⚁, ⚃, ⚅\}$$

$$high = \{⚃, ⚄, ⚅\}$$



Conditional probability

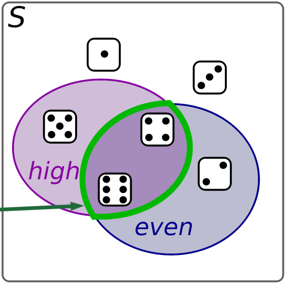
Example:

$$S = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$$

$$even = \{⚁, ⚃, ⚅\}$$

$$high = \{⚃, ⚄, ⚅\}$$

$$even \cap high = \{⚃, ⚅\}$$



$A \cup B$ is the *union* between A and B

$A \cap B$ is the *intersection* of A and B

Conditional probability

Example:

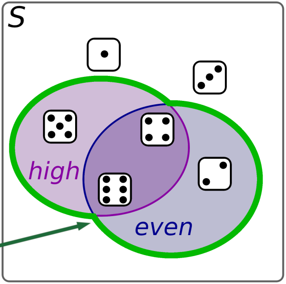
$$S = \{\text{⚀}, \text{⚁}, \text{⚂}, \text{⚃}, \text{⚄}, \text{⚅}\}$$

$$even = \{\text{⚁}, \text{⚃}, \text{⚅}\}$$

$$high = \{\text{⚃}, \text{⚄}, \text{⚅}\}$$

$$even \cap high = \{\text{⚃}, \text{⚅}\}$$

$$even \cup high = \{\text{⚁}, \text{⚃}, \text{⚄}, \text{⚅}\}$$



$A \cup B$ is the *union* between A and B

$A \cap B$ is the *intersection* of A and B

Conditional probability

Example:

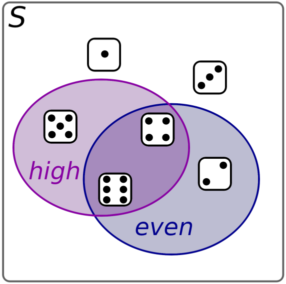
$$S = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$$

$$\mathit{even} = \{⚁, ⚃, ⚅\}$$

$$\mathit{high} = \{⚃, ⚄, ⚅\}$$

$$\mathit{even} \cap \mathit{high} = \{⚃, ⚅\}$$

$$\mathit{even} \cup \mathit{high} = \{⚁, ⚃, ⚄, ⚅\}$$



When all elements in the sampling space have equal probability; what is the conditional probability:

$\mathbb{P}(\mathit{high} \mid \mathit{even})$

Conditional probability

Example:

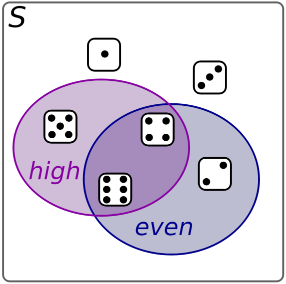
S = {⚀, ⚁, ⚂, ⚃, ⚄, ⚅}

even = {⚁, ⚃, ⚅}

high = {⚃, ⚄, ⚅}

even ∩ *high* = {⚃, ⚅}

even ∪ *high* = {⚁, ⚃, ⚄, ⚅}



When all elements in the sampling space have equal probability; what is the conditional probability:

$$\mathbb{P}(high \mid even) = \frac{\#\text{elements in } even \cap high}{\#\text{elements in } even} = \frac{2}{3}$$

Conditional probability

Example:

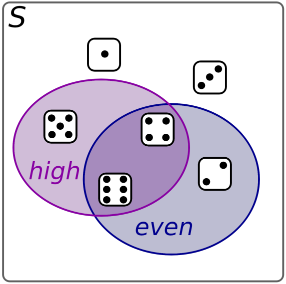
$$S = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$$

$$\mathit{even} = \{⚁, ⚃, ⚅\}$$

$$\mathit{high} = \{⚃, ⚄, ⚅\}$$

$$\mathit{even} \cap \mathit{high} = \{⚃, ⚅\}$$

$$\mathit{even} \cup \mathit{high} = \{⚁, ⚃, ⚄, ⚅\}$$



When all elements in the sampling space have equal probability; what is the conditional probability:

$$\mathbb{P}(\mathit{high} \mid \mathit{even}) = \frac{\#\text{elements in } \mathit{even} \cap \mathit{high}}{\#\text{elements in } \mathit{even}} = \frac{2}{3}$$

As limit of frequency in a sampling process:

$$\mathbb{P}(\mathit{high} \mid \mathit{even}) = \lim_{N \to \infty} \frac{N_{⚃} + N_{⚅}}{N_{⚁} + N_{⚃} + N_{⚅}}$$

Conditional probability

Example:

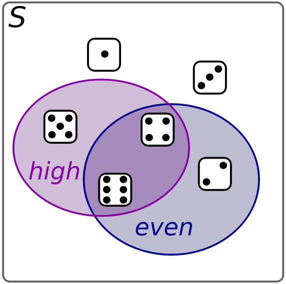
$S = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$

$\textit{even} = \{$⚁, ⚃, ⚅$\}$

$\textit{high} = \{$⚃, ⚄, ⚅$\}$

$\textit{even} \cap \textit{high} = \{$⚃, ⚅$\}$

$\textit{even} \cup \textit{high} = \{$⚁, ⚃, ⚄, ⚅$\}$



Conditional probability

If A and B are two events in the sample space S, then the **conditional probability of A given B** is defined as:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{when } \mathbb{P}(B) > 0.$$

Product rule

Example:

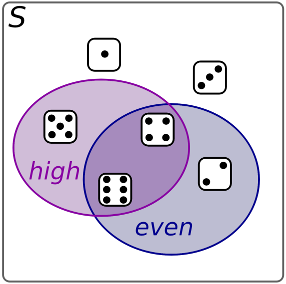
$$S = \{\text{⚀}, \text{⚁}, \text{⚂}, \text{⚃}, \text{⚄}, \text{⚅}\}$$

$$\mathit{even} = \{\text{⚁}, \text{⚃}, \text{⚅}\}$$

$$\mathit{high} = \{\text{⚃}, \text{⚄}, \text{⚅}\}$$

$$\mathit{even} \cap \mathit{high} = \{\text{⚃}, \text{⚅}\}$$

$$\mathit{even} \cup \mathit{high} = \{\text{⚁}, \text{⚃}, \text{⚄}, \text{⚅}\}$$



Product rule

If A and B are two events in the sample space S, then the **product of the probabilities for** A **and** B is defined as:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A)$$

Sum rule

Example:

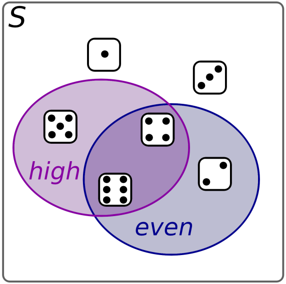
S = {⚀, ⚁, ⚂, ⚃, ⚄, ⚅}

even = {⚁, ⚃, ⚅}

high = {⚃, ⚄, ⚅}

even ∩ *high* = {⚃, ⚅}

even ∪ *high* = {⚁, ⚃, ⚄, ⚅}



Sum rule

If A and B are two events in the sample space S, then the **sum of the probabilities for** A **and** B is defined as:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Exercises

First exercise session

1. Writet a stochastic simulator in Python.
2. Illustrate the *axioms of probability*.
3. Become familiar with the sum- and product-rule.
4. Show *Bayes Theorem* for conditional probabilities:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

Bayes' theorem

Suppose we know $\mathbb{P}(B|A)$, but are interested in the probability $\mathbb{P}(A|B)$

Bayes' theorem

- For any two events A and B in the sample space S, where $\mathbb{P}(B) \neq 0$, it holds that:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)}$$

- Let $A_1, A_2, ..., A_K$ be a *partition* of the sample space S. Using the *law of total probability* for $\mathbb{P}(B)$, it then holds that:

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(B|A_j) \cdot \mathbb{P}(A_j)}{\sum_k \mathbb{P}(B|A_k) \cdot \mathbb{P}(A_k)}$$

Random variables, probability mass functions and joint probabilities

Random variable

- Real-world sample spaces and events are often complicated and abstract. To say anything meaningful we need to map it to something quantitative.
- This is done by introducing functions that "asks questions" about the samples.

Example:
Let the function **`numberOfDots`** map the outcome of a die-roll to real values by counting the "number of dots":

$$\texttt{numberOfDots}(\text{⚀}) = 1$$
$$\texttt{numberOfDots}(\text{⚁}) = 2$$
$$\texttt{numberOfDots}(\text{⚂}) = 3$$
$$\texttt{numberOfDots}(\text{⚃}) = 4$$
$$\texttt{numberOfDots}(\text{⚄}) = 5$$
$$\texttt{numberOfDots}(\text{⚅}) = 6$$

Random variable

- The concept of a *random variable* links the sample space and events to numeric data, necessary for applying statistics and machine-learning methods.

Random variable

A **random variable** X is a function that assigns a real number $X(s)$ to each outcome s in the sample space S

$$X(s) \in \mathbb{R}$$

Example:
Flip a coin 5 times, let $X(s)$ be the number of heads in the event s:

- if $s_1 =$ H H T H T, then $X(s_1) = 3$
- if $s_2 =$ T T T T T, then $X(s_2) = 0$
- if $s_3 =$ H T T H H, then $X(s_3) = 3$

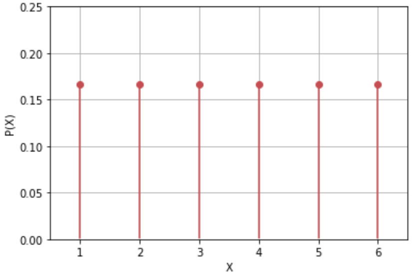
Probability Mass Function

Probability Mass Function (PMF)

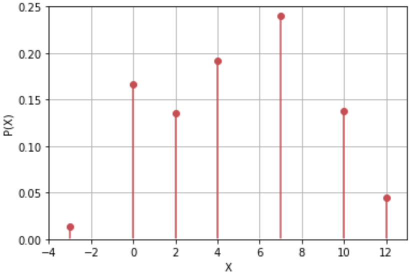
The *probability mass function* $P(x)$ assigns a probability to each possible value x of the random variable X

$$P(x) = \mathbb{P}(X = x)$$

Example:



Roll of a fair 6-sided die.



Grades in Physics 1 (s2018).

Joint probabilities

- Outcome of trial characterized by two "values"
- We observe pairs (x, y):
  - Length of an object vs subjective notion of "large"
  - Image of surface vs quality

Joint probabilities

- Outcome of trial characterized by two ”values”
- We observe pairs (x, y):
  - Length of an object vs subjective notion of ”large”
  - Image of surface vs quality

Joint probability

The *joint probability mass function,* $P(x, y)$*,* assigns a probability to each possible joint outcome x and y of the two random variables X and Y.

For ease of reading we often write the PMF as $P(X = x, Y = y)$

Working with joint probabilities

Example: counts for $N = 50$ trials, sampling the two variables X and Y:

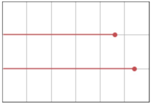
	$X = 1$	$X = 2$	$X = 3$	
$Y = 1$	12	7	4	23
$Y = 2$	3	11	13	27
	15	18	17	50

Working with joint probabilities

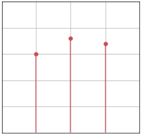
Joint and marginal probabilities for X and Y:

	$X=1$	$X=2$	$X=3$	
$Y=1$	.24	.14	.08	.46
$Y=2$	.06	.22	.26	.54
	.30	.36	.34	1

$P(Y=y)$



$P(X=x)$



Marginal probability

Let Y, X be random variables. The *marginal probability* for $Y=y$ is given by:

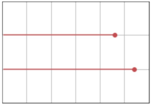
$$P(Y=y)=\sum_{x} P(Y=y, X=x)$$

Working with joint probabilities

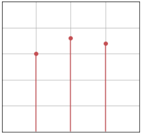
Example: computing the *conditional* probability $P(x|Y=1)$

	$X=1$	$X=2$	$X=3$	
$Y=1$	.24	.14	.08	.46
$Y=2$	.06	.22	.26	.54
	.30	.36	.34	1

$P(Y=y)$



$P(X=x)$



Conditional probability

Let X, Y be random variables. The *conditional probability* for $X=x$ is given by:

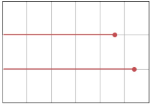
$$P(X=x|Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}$$

Working with joint probabilities

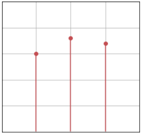
Example: computing the *conditional* probability $P(x|Y=1)$

	$X=1$	$X=2$	$X=3$	
$Y=1$	.52	.31	.17	1
$Y=2$	.06	.22	.26	.54
	.30	.36	.34	1

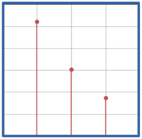
$P(Y=y)$



$P(X=x)$



$P(X=x|Y=1)$



Conditional probability

Let X, Y be random variables. The *conditional probability* for $X=x$ is given by:

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

Law of total probability

Law of total probability for random variables

Let X, Y be random variables where x, y represent possible values, it holds that:

$$P(x) = \sum_y P(x, y) = \sum_y P(x|y) \cdot P(y)$$

Independent random variables

The definition of the joint distribution can be extended to any number of random variables $X_1, X_2, \ldots, X_N$.
Often a higher number occurs when modelling *independent random variables*.

Independent random variables

Two random variables X, Y are independent if, for every x, y:

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

Example:
Rolling two fair dice simultaneously, what is the probability of rolling two 6′s?

$$P(X = 6, Y = 6) \quad = \quad P(X = 6) \cdot P(Y = 6) \quad = \quad \frac{1}{6} \cdot \frac{1}{6} \quad = \quad \frac{1}{36}$$

Exercises

Second exercise session

1. Learn to evaluate marginal and conditional probabilities and reason about their interpretation.
2. Build a probabilistic classifier using *Bayes' theorem*.

Learning objectives

Learning log

- Understand probabilities as frequencies in a sampling process.
- Understand the definition of a sample space and identify sample spaces of random processes.
- Use random generator in NumPy to write programs that simulates a random process.
- Understand intersections, unions, complements of basics sets and compute probabilities using the sum- and product-rule.
- Understand the definition of conditional probabilities and their inverse relation (Bayes' theorem).
- Describe a *random variable* as a mapping that assigns a *real number* to each outcome in a sample space.
- Describe the *probability mass function* for a discrete random variable.
- Understand the joint sample space for independent and dependent random variables.