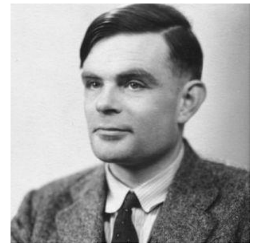
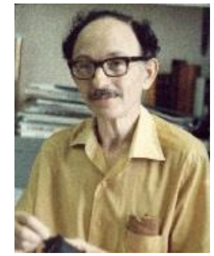
Decision Theory







Thomas Bayes, Pierre Simon de Laplace,, Bruno de Finetti, Alan Turing, Irving Good, Leonard Jimmie Savage, Dennis Lindley, Arnold Zellner, Kathryn Chaloner, Susie Bayarri , Daniel Kahneman

Who am I?

Anders Nordgaard

Reader and Forensic specialist in statistics
Swedish Police Authority – National Forensic Centre.

Former senior lecturer and director of studies at the Division of Statistics (and Machine Learning), LiU.

Nowadays, adjunct lecturer at this division (up to 20 % of full time)

Teaching this course
Examiner of Master's thesis work

Easiest way of contact: andno100@gmail.com

A course on decision making under uncertainty – Reasoning with probabilities

- <u>Course responsible and tutor:</u>

  Anders Nordgaard (andno100@gmail.com, Anders.Nordgaard@liu.se)

- <u>Course web page:</u>

  www.ida.liu.se/~732A66

  Note: There is no course room in Lisam for this course (due to ignorance with the course responsible)

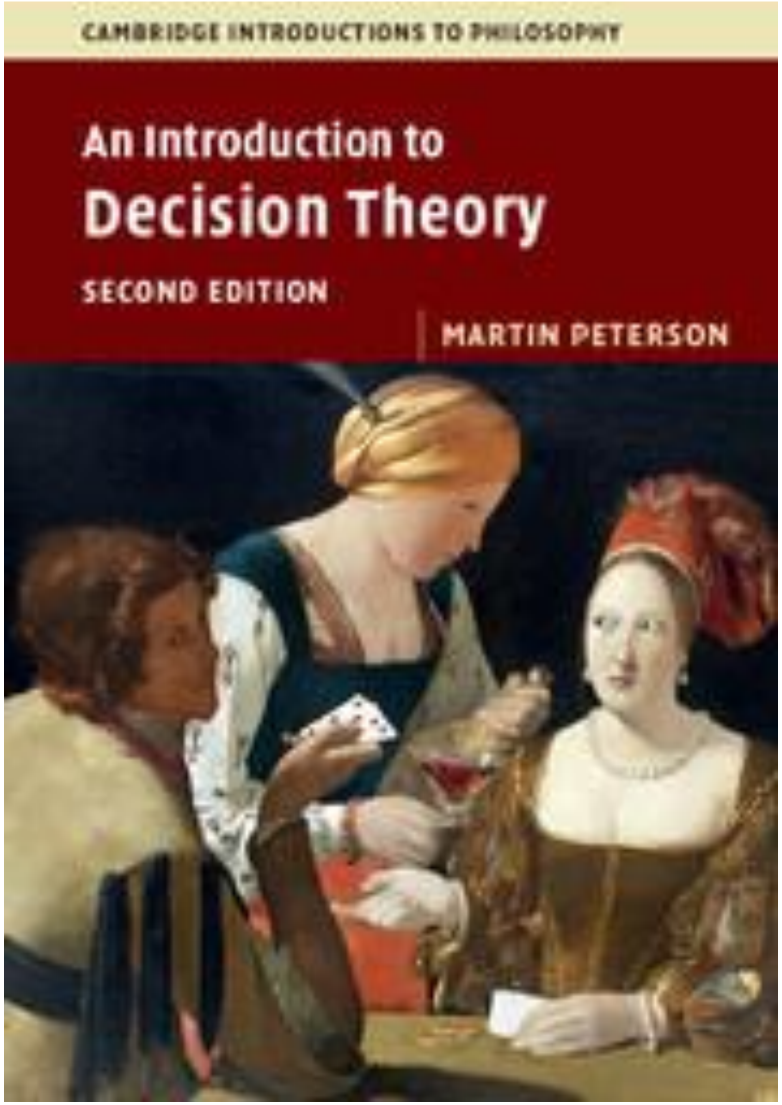
- <u>Teaching:</u>

  Lectures on theory

  Seminars with complex problems

  Discussion of assignments

- Course literature:
  - Peterson M.: *An Introduction to Decision Theory* 2nd ed. Cambridge University Press, 2017. ISBN 9781316606209 (paperback), 9781316585061 (digital)



- Course literature:
  - Peterson M.: *An Introduction to Decision Theory* 2nd ed. Cambridge University Press, 2017. ISBN 9781316606209 (paperback), 9781316585061 (digital)

  Former course literature also works:
  - Winkler R.L.: *An Introduction to Bayesian Inference and Decision* 2nd ed. Probabilistic Publishing, 2003 ISBN 0-9647938-4-9
  - Electronic version available for purchase or lending: https://archive.org/details/introductiontoba00robe/page/n8/mode/1up
  - *The relevant exercises from this book will temporarily be uploaded to the course web*
- Additional literature:
  - Taroni F., Bozza S., Biedermann A., Garbolino P., Aitken C. : Data analysis in forensic science – A Bayesian decision perspective, Chichester: Wiley, 2010
  - Gittelson S. (2013). Evolving from Inferences to Decisions in the Interpretation of Scientific Evidence. Thèse de Doctorat, Série criminalistique LVI, Université de Lausanne. ISBN 2-940098-60-3. Available at http://www.unil.ch/esc/files/live/sites/esc/files/shared/These_Gittelson.pdf

- Examination:
  - Assignments (compulsory to pass)
  - Final oral exam (compulsory, decides the grade)

Assignments:

- There will be 3-4 assignments
- Co-working is permitted…
- …but each student must submit their own solution
- Insufficient solutions will need supplementary submission

Oral exam:

- Normally in a group of 2 students (occasionally 1 student, never 3 or more)
- A discussion on the course contents and concepts with practical examples
- 2 hours duration (1 student: 1 hour)
- Individual feedback and grading

Outcome of Evaliuate course evaluation for study year 2021/22

- Response rate: 22%
- No questions sticking out in the multiple choice questions
- Free-text answers on question 6 and 7:

6. What changes do you consider to be possible that would improve the course with respect to, for example, content, teaching principles, administration, teachi ng methods, or examination forms?

Make it a normal paper exam or only examination project

7.
Give examples of content, teaching principles, teaching methods, examination forms, or any other aspect of the course that you consider to have been particularly successful.

oral exam

Opinions taken up at oral exams:

- Connections to machine laerning
- Prepared topics seminar
- Shorter sessions
- Shorter course period (finish entire course by Christmas)
- Mixture of lecture and problem discussion
- One more assignment – less difficult
- More on game theory
- Use Lisam: Extra exercises, QA:s, summaries of key points
- Case examples
- More software use – less maths
- Tell the purpose of each lecture at the beginning
- Presentations by students
- Follow-up on all assignments
- Explain the notation
- Larger assignment
- Repetition – problem discussion
- Roadmap

Who are you?

Name

Background

Expectations on this course

Lecture 1: Probabilities

Purpose: To repeat and extend previous knowledge of probability calculus.

The concept of probability

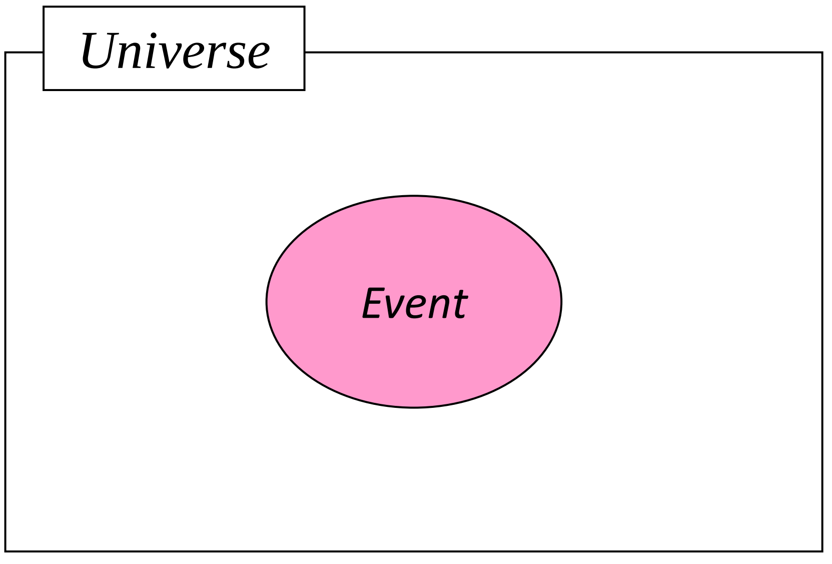




Category	*Frequency*	*Probability ?*
	9	0.6
	3	0.2
	3	0.2

The *probability* of an event is…

- the degree of belief in the event (that the event has happened)
- a measure of the size of the event relative to the size of the universe



Probability of event= $P(Event)$

- $0 \leq P(Event) \leq 1$
- $P(Universe) = 1$
- If two events, A and B are mutually exclusive then

$$P(A \text{ or } B) = P(A) + P(B)$$

"Kolmogorov axioms" (finite additivity variant)

The universe, all events in it and the probabilities assigned to each event constitute the ***probability space***.

This does *not* mean that…

“probabilities and stable relative frequencies are equal” (*Frequentist definition of probability*)

merely…

If any event is assigned a probability, that probability must satisfy the axioms.

Example: Coin tossing

Suppose you toss a coin. One possible event is “heads”, another is “tails”

If you assign a probability p to “heads” and a probability q to “tails they both must be between 0 and 1.

As “heads” cannot occur simultaneously with “tails”, the probability of “heads or tails” is $p + q$.

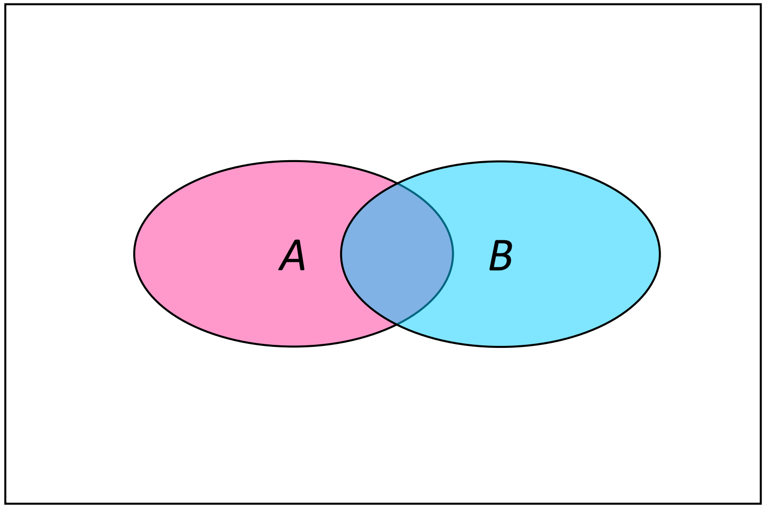
If no other event is possible then “heads or tails” = Universe ➔ $p + q = 1$

Relevance, Conditional probabilities

An event B is said to be *relevant* for another event A <u>if</u> the probability (degree of belief) that A is true depends on the state of B.

The *conditional* probability of A <u>given</u> that B is true is

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



If B is true then its *complement* $\bar{B}$ $(B^C, \neg B)$ is *irrelevant* to consider.

If A is to be true under these conditions, only the part of A inside B should be considered.

This part coincides with (A,B)

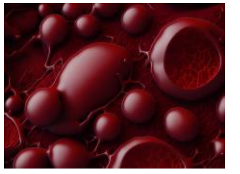
The measure of the size of this event must be relative to the size of B

Example:

Assume you believe that approx. 1% of all human beings carry both a gene for developing disease *A* and a gene for developing disease *B*.

Further you believe that

- 8 % of all human beings carry the gene for developing disease A



- 10% of all human beings carry the gene for developing disease *B.*

Then as a consequence your degree of belief that a person who has developed disease *B* also carries the gene for developing disease *A* should be 10% (0.01/0.10)

Since 10 % is different from 8 %, carrying the gene for *B* is relevant for carrying the gene for *A.*

Reversing the definition of conditional probability:

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

"The multiplication law of probability"

but also… $P(A,B) = P(B|A) \cdot P(A)$

$$\Rightarrow P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \text{ and } P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

➔ For sorting out conditional probabilities it is not necessary to assign the probabilities of intersections

“All probabilities are conditional…”

How a probability is assigned depends on background knowledge.

E.g. if you assign the probability 0.5 for the event “heads” in a coin toss, you have assumed that

- the coin is fair
- the coin cannot land endways

…but it may be the case that you cannot assign any probability to the background knowledge

Let I denote all background knowledge *relevant* for A

$$\Rightarrow P(A) = P(A|I)$$

Extensions:

$$P(A, B|I) = P(A|B, I) \cdot P(B|I)$$
$$P(A_1, A_2, \dots, A_n|I) =$$
$$= P(A_1|I) \cdot P(A_2|A_1, I) \cdot \dots \cdot P(A_n|A_1, A_2, \dots, A_{n-1}, I)$$

***Example*:** Suppose you randomly pick 3 cards from a well-shuffled deck of cards. What is the probability you will in order get a spade, a hearts and a spade?

I = The deck of cards is well-shuffled $\Rightarrow$ It does not matter how you pick your cards.

Let A_1 = First card is a spade; A_2 = Second card is a hearts; A_3 = Third card is a spade

$$\Rightarrow P(A_1, A_2, A_3|I) = P(A_1|I) \cdot P(A_2|A_1, I) \cdot P(A_3|A_1, A_2, I) =$$
$$= \frac{13}{52} \cdot \frac{13}{51} \cdot \frac{12}{50} \approx 0.015$$

Relevance and (conditional) independence

If B is relevant for A then $P(A|B,I) \neq P(A|I)$

If B is *irrelevant* for A then $P(A|B,I) = P(A|I)$

which in turn gives $P(A,B|I) = P(A|I) \cdot P(B|I)$

In this case A and B are said to be *conditionally independent* events. (In common statistical literature only *independent* is used as term.)

Note that it is the background knowledge I that determines whether this holds or not.

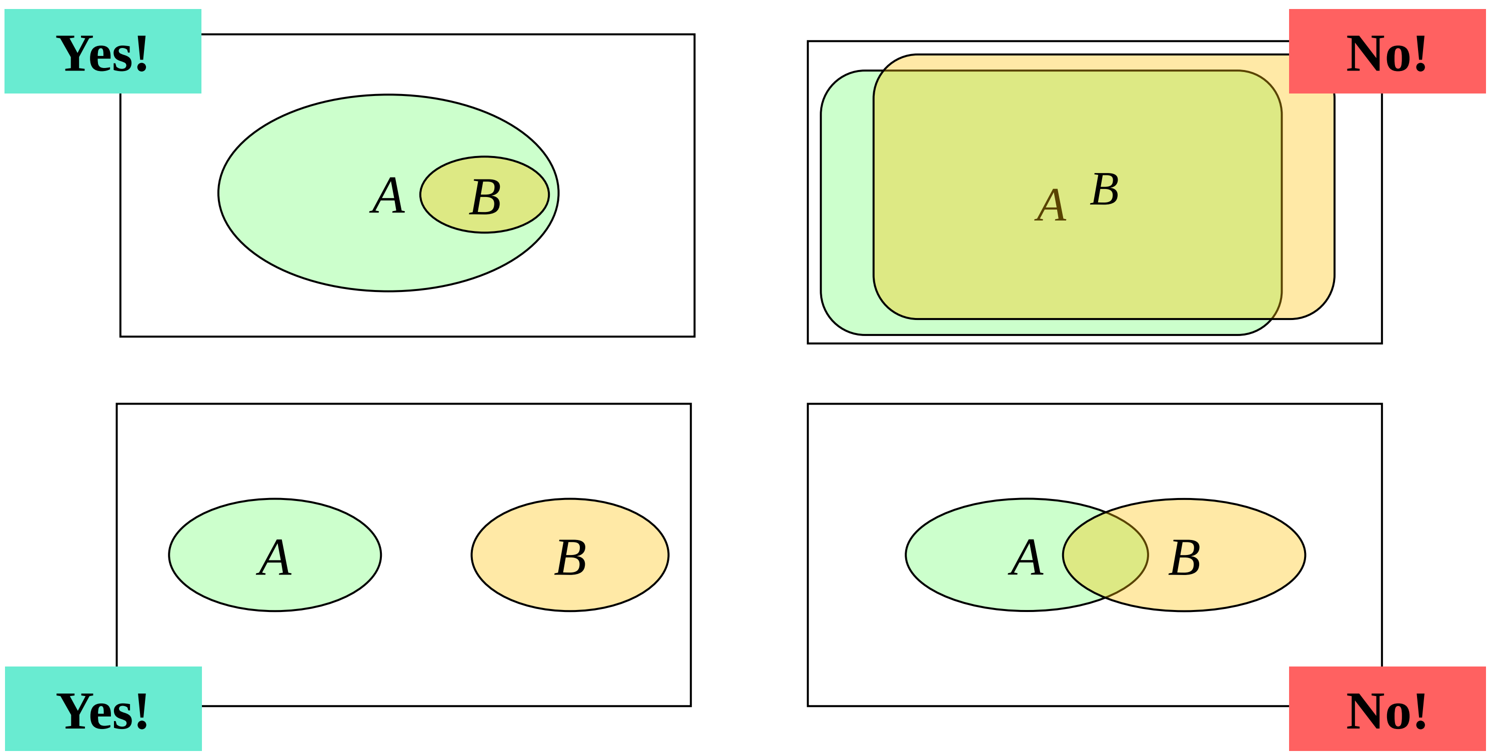
Note also that if $P(A|B,I) = P(A|I)$ then $P(B|A,I) = P(B|I)$

Irrelevance is reversible!

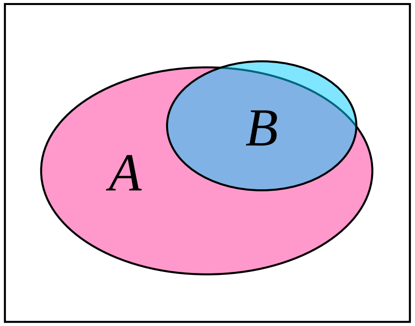
Below are four rectangles. Each rectangle represents the universe, so its area is equal to one (1=100%)

Assume that the sets A (green) and B (yellowish) are drawn according to scale (the sizes of the sets are proportional to the probabilities of the events).

In which of the cases below are A and B definitely conditionally dependent (given I)?

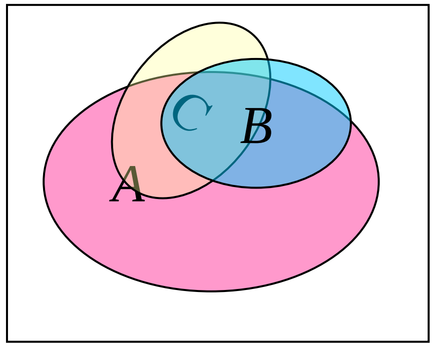


Further conditioning…



$$P(A, B|I) \neq P(A|I) \cdot P(B|I)$$

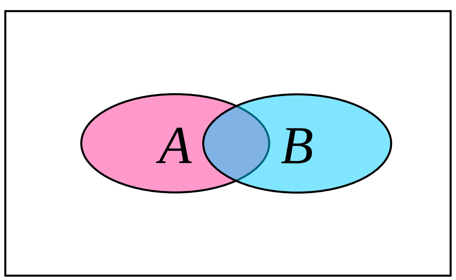
Area of $A \cap B$ divided by area of A is not equal to area of B divided by area of rectangle.



<u>Inside C</u> the area of <u>$A \cap B$</u> divided by the area of A is equal to the area of B divided by the area of C.

$$P(A, B|C, I) = P(A|C, I) \cdot P(B|C, I)$$

Two events that are conditionally dependent under one set of assumptions may be conditionally *independent* under another set of assumptions



The law of total probability:

$$P(A|I) = P(A, B|I) + P\left(A, \overline{B}|I\right) =$$
$$= P(A|B, I) \cdot P(B|I) + P\left(A\middle|\overline{B}, I\right) \cdot P\left(\overline{B}|I\right)$$

$\Rightarrow$ Bayes' theorem:

$$P(A|B, I) = \frac{P(B|A, I) \cdot P(A|I)}{P(B|A, I) \cdot P(A|I) + P\left(B\middle|\overline{A}, I\right) \cdot P\left(\overline{A}|I\right)}$$