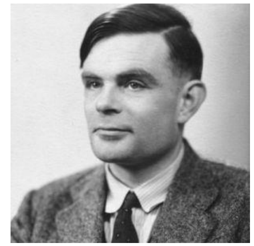
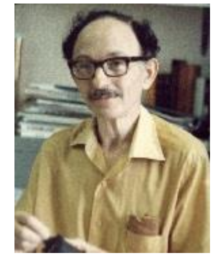
Decision Theory







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
Supervision 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

  Practical exercises

  Discussion of assignments

- Course literature:
  - 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 (substitutes the amount from a book that a teacher is entitled to distribute in paper copies to students)

- 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

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

  Assignments:
  - There will be 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
  - Time-point is flexible, but not before end of teaching and preferably before June 15
  - Depending on the development of the Covid-19 situation, the exam may be online, but a physical meeting is highly preferred

Amendments due to last year's course evaluation

- Intention to have as much of the meetings in campus
- Keep (and update) recording of lectures
- Som adjustments of the assignments

...but there were very few suggestions for improvement in the course evaluation.

Lecture 1:
Repeat and extend…

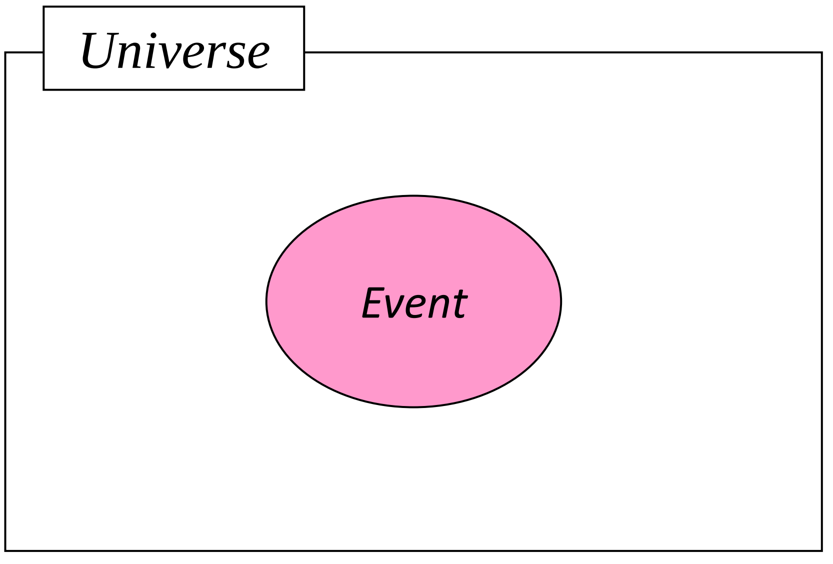
Probability and likelihood

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



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

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)

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 $p + q$.

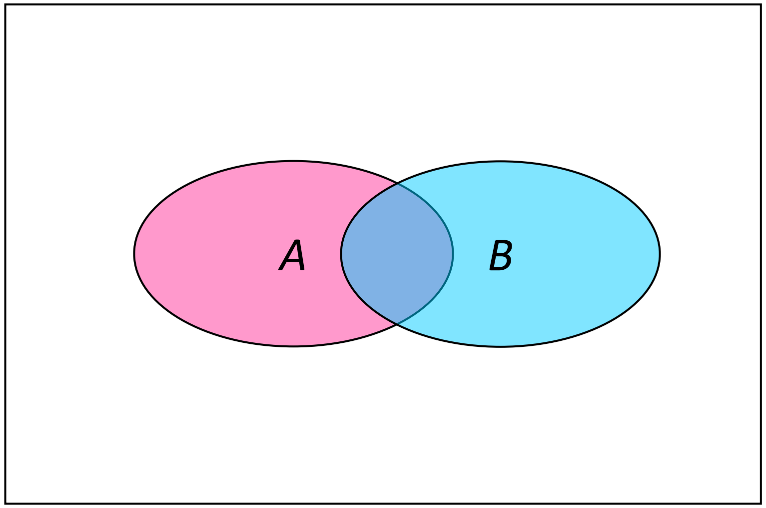
If no other event is possible then "heads or tails" = Universe ➔ $p + q = 1$

Relevance, Conditional probabilities

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

The *conditional* probability of A given that B is true is

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



If B is true then $\overline{B}$ 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 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)

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… $$P(A,B) = P(B|A) \cdot P(A)$$

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

➔ 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 ⇒ 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* is said to be *conditionally independent* events. (In common statistical literature only *independent* is used as term.)

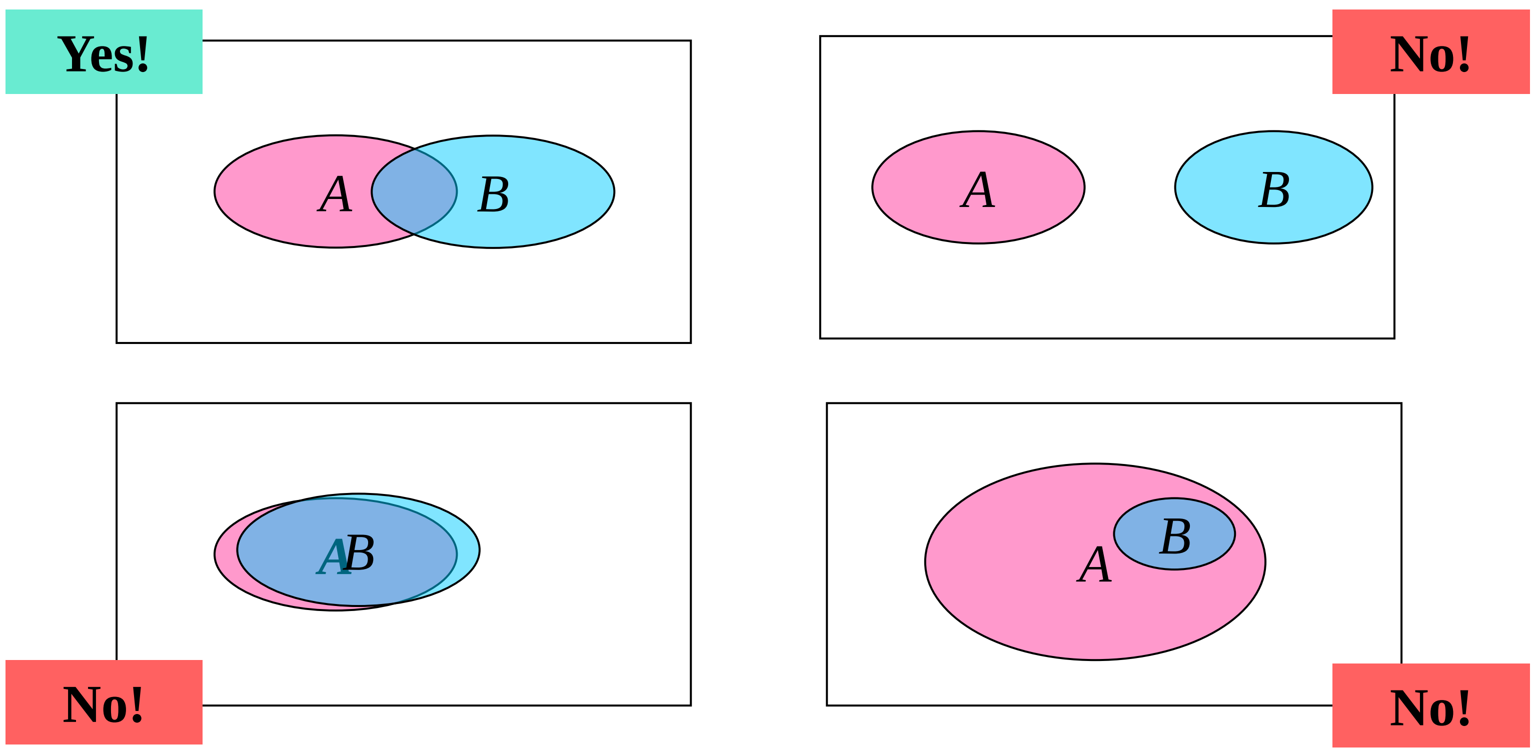
Note that it is the background knowledge *I* that determines whether this holds or not.

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

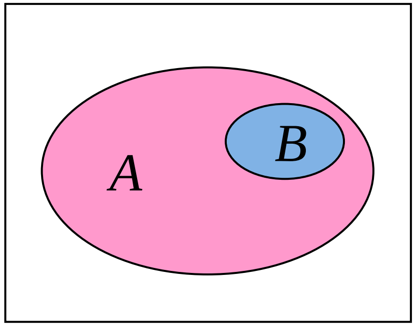
Irrelevance is reversible!

Assume that the sets below are drawn according to scale (the sizes of the sets are proportional to the probabilities of the events).

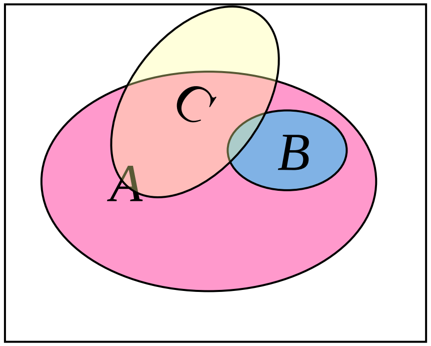
In which of the cases below may *A* and *B* be conditionally independent (given *I*)?



Further conditioning…

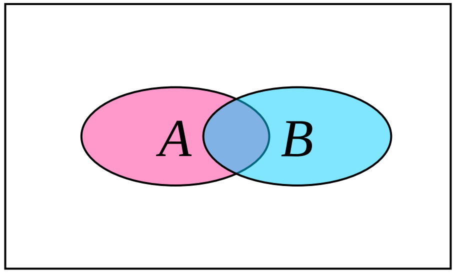


$$P(A,B|I) \neq P(A|I) \cdot P(B|I)$$



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

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:

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

Example:

Assume a method for detecting a certain kind of dye on banknotes is such that

- it gives a positive result (detection) in 99 % of the cases when the dye is present, i.e. the proportion of false negatives is 1%

- it gives a negative result in 98 % of the cases when the dye is absent, i.e. the proportion of false positives is 2%

The presence of dye is rare: prevalence is about 0.1 %

Assume the method has given positive result for a particular banknote.

What is the conditional probability that the dye is present?

Solution:

Let A = "Dye is present" and B = "Method gives positive result"

What about I ?

• We must assume that the particular banknote is as equally likely to be exposed to dye detection as any banknote in the population of banknotes.

• **Is that a realistic assumption?**

Now, $P(A) = 0.001; \quad P(B|A) = 0.99; \quad P\left(B\,\middle|\,\overline{A}\right) = 0.02$

Applying Bayes' theorem gives

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

$$= \frac{0.99\cdot 0.001}{0.99\cdot 0.001 + 0.02\cdot 0.999} = \underline{\qquad\qquad}$$

Odds and Bayes' theorem on odds form

The *odds* for an event A "is" a quantity equal to the probability:

$$Odds(A) = \frac{P(A)}{P(\overline{A})} = \frac{P(A)}{1 - P(A)} \quad \Rightarrow \quad P(A) = \frac{Odds(A)}{Odds(A) + 1}$$

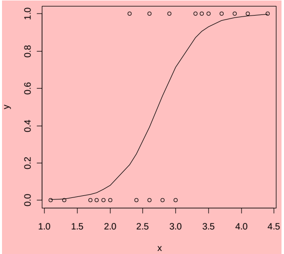
Why two quantities for the same thing?

Example: An "epidemiological" model

Assume we are trying to model the probability p of an event (i.e. the prevalence of some disease).

The *logit link* between p and a set of k explanatory variables $x_1, x_2, \ldots, x_k$ is

$$\text{logit}(p) = \ln \frac{p}{1-p} = \beta_0 + \beta_1 \cdot x_1 + \cdots + \beta_k \cdot x_k$$



This link function is common in *logistic regression analysis*.

Note that we are modelling the natural logarithm of the odds instead of modelling p.

As the odds can take any value between 0 and ∞ the logarithm of the odds can take any value between $-\infty$ and ∞ ➔ Makes the model practical.

Conditional odds

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

expresses the *updated* belief that *A* holds when we take into account that *B* holds

Like probabilities, all odds are conditional if we include background knowledge *I* as our basis for the calculations.

$$Odds(A|I) = \frac{P(A|I)}{P(\bar{A}|I)}; \quad Odds(A|B,I) = \frac{P(A|B,I)}{P(\bar{A}|B,I)}$$

The odds ratio:

$$OR = \frac{Odds(A|B,I)}{Odds(A|I)} = \frac{\frac{P(A|B,I)}{P(\bar{A}|B,I)}}{\frac{P(A|I)}{P(\bar{A}|I)}}$$

expresses *how* the belief that *A* holds updates when we take into account that *B* holds.

Now

$$\underline{Odds(A|B,I)} = \frac{P(A|B,I)}{P(\bar{A}|B,I)} = \frac{\frac{P(B|A,I)\cdot P(A|I)}{P(B|I)}}{\frac{P(B|\bar{A},I)\cdot P(\bar{A}|I)}{P(B|I)}} =$$

$$= \frac{P(B|A,I)}{P(B|\bar{A},I)}\cdot\frac{P(A|I)}{P(\bar{A}|I)} = \underline{\frac{P(B|A,I)}{P(B|\bar{A},I)}\cdot Odds(A|I)}$$

"*Bayes' theorem on odds form*"

The ratio $\dfrac{P(B|A,I)}{P(B|\overline{A},I)}$

is a special case of what is called a *likelihood ratio* (the concept of “likelihood” will follow)

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

where we have substituted C for $\bar{A}$ and we no longer require A and C to be complementary events (not even mutually exclusive).

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

always holds, but the ratios involved are not always odds

“The updating of probability ratios when a new event is observed goes through the likelihood ratio based on that event.”

Probability and Likelihood – Synonyms?

An event can be *likely* or *probable*, which for most people would be the same.
Yet, the definitions of probability and likelihood are different.

In a simplified form:

- The probability of an event measures the degree of belief that this event is true and is used for reasoning about not yet observed events
- The likelihood of an event is a measure of how likely that event is in light of another *observed* event
- Both are objected to probability calculus

More formally…

Consider the *unobserved* event *A* and the *observed* event *B*.

There are probabilities for both representing the degrees of belief for these events in general: $P(A|I)$, $P(B|I)$

However, as *B* is observed we might be interested in

$$P(A|B,I)$$

which measures the *updated* degree of belief that *A* is true once we know that *B* holds. Still a probability, though.

How interesting is

$$P(B|A,I)$$

?

$P(B \mid A, I)$ might look meaningless to consider as we have actually observed *B*.

However, it says something about *A*.

We have observed *B* and if *A* is relevant for *B* we may compare $P(B \mid A, I)$ with $P(B \mid \bar{A}, I)$.

Now, even if we have not observed *A* or $\bar{A}$, one of them must be true (as a consequence of *A* and *B* being relevant for each other).

If $P(B \mid A, I) > P(B \mid \bar{A}, I)$ we may conclude that *A* is more *likely* to have occurred than is $\bar{A}$, or better phrased:

"*A* is a better *explanation* to why *B* has occurred than is $\bar{A}$".

$P(B \mid A, I)$ is called the *likelihood* <u>of *A*</u> given the observed *B* (and $P(B \mid \bar{A}, I)$ is the likelihood of $\bar{A}$).

Note! This is different from the conditional probability of *A* given *B*: $P(A \mid B, I)$.

Potential danger in mixing things up:

When we say that an event is the more likely one in light of data we do not say that this event has the highest probability.

Using the likelihood as a measure of how likely is an event is a matter of *inference to the best explanation*.

Logics: Implication:

$$A \to B$$

- If A is true then B is true, i.e. $P(B \mid A, I) \equiv 1$
- If B is false then A is false, i.e. $\mathrm{P}(A \mid \overline{B}, I) \equiv 0$
- If B is true we cannot say anything about whether A is true or not (implication is different from equivalence)

“Probabilistic implication”:

$$A \xrightarrow{P} B$$

- If A is true then B *may* be true, i.e. $P(B|A, I) > 0$
- If B is false the A may still be true, i.e. $P(A|\overline{B}, I) > 0$
- If B is true then we may decide which of A and $\bar{A}$ that is the best explanation

Inference to the best explanation:

- B is observed
- $A_1, A_2, \ldots, A_m$ are potential alternative explanations to B
- If for each $j \neq k$ $P(B \mid A_k, I) > P(B \mid A_j, I)$ then A_k is considered the best explanation for B and is provisionally accepted