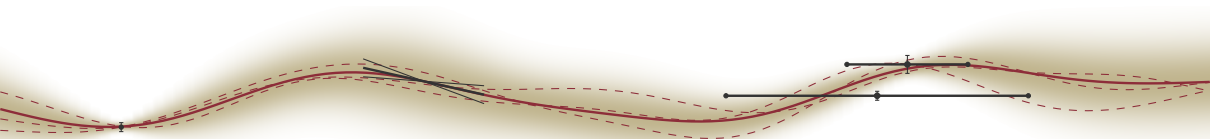
Probabilistic Inference and Learning
Lecture 00
Introduction - Reasoning under Uncertainty

Philipp Hennig

15 October 2018

Eberhard Karls
Universität Tübingen

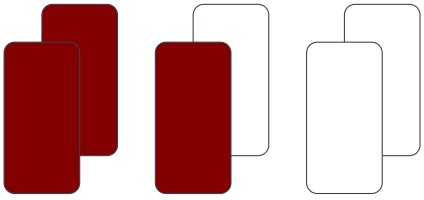
Faculty of Science
Department of Computer Science
Chair for the Methods of Machine Learning



Which Card?

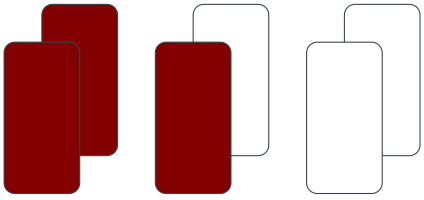
an opening experiment

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Which Card?

an opening experiment



- $\frac{1}{2}$
- $2/3$
- something else
- I don't know yet

An *inference* problem requires statements about the value of an *unobserved* (latent) variable x based on observations y which are **related** to x, but may not be sufficient to fully determine x. This requires a notion of **uncertainty**.

Life is Uncertain

the ability to reason under uncertainty is the hallmark of intelligence

Life is Uncertain

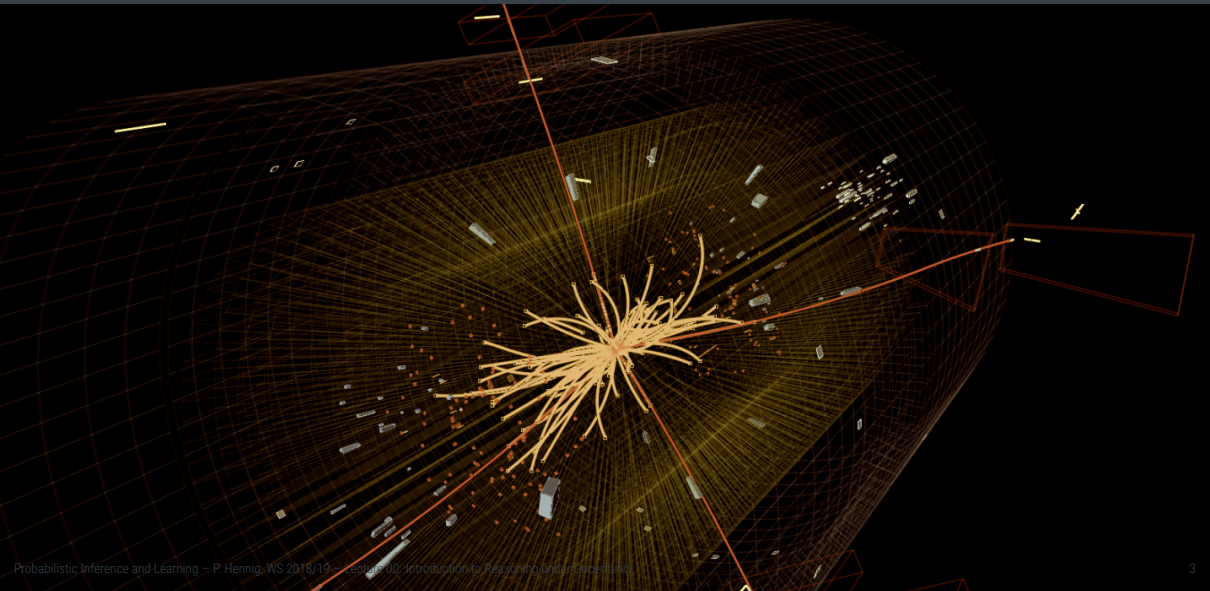
the ability to reason under uncertainty is the hallmark of intelligence

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

[ICJ den Haag, 2006]

Life is Uncertain

the ability to reason under uncertainty is the hallmark of intelligence



[© CERN / CMS Collaboration, 2011]

Life is Uncertain

the ability to reason under uncertainty is the hallmark of intelligence

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

[Luke Fildes, 1891 (Tate Modern)]

The actual science of logic is conversant at present only with things either certain, impossible, or entirely doubtful, none of which (fortunately) we have to reason on. Therefore the true logic for this world is the calculus of Probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind.

James Clerk Maxwell (1831–1879)
[image source: BBC / public domain]

Deductive and Plausible Reasoning

Limits of propositional logic

[adapted from E.T. Jaynes, 2003, §1 & 2]

$A =$ "it will begin to rain by 6pm"
$B =$ "the sky will become cloudy before 6pm"

$$A \Rightarrow B$$

if A is true, the B is true

Deductive Reasoning:

A is true	thus	B is true	modus ponens
B is false	thus	A is false	modus tollens

Plausible Reasoning:

B is true	thus	A becomes more plausible
A is false	thus	B becomes less plausible

A = "it is raining"
B = "the sky is cloudy"

$$p(B \mid A) > p(B)$$

if A is true, the B becomes more plausible

Plausible Reasoning:

A is true	thus	B becomes more plausible
B is false	thus	A becomes less plausible
B is true	thus	A becomes more plausible
A is false	thus	B becomes less plausible

The goal of this course:

- Establish a *formal framework* for **probable reasoning**
- Use it to build *powerful* inference mechanisms for **real-world problems**
- Develop the *technical tools* necessary to implement **inference** in practice

Cox's Axioms

A formalization of **common sense**

[adapted from Jaynes, 2003, §1.7]

Definition (Cox's axioms)

1. The *plausibility* of B assuming A is true **is a real number**, denoted by $p(B \mid A)$. Larger numbers correspond to higher plausibility.
2. **Plausibility complies with common sense**:
   If $p(A \mid C') > p(A \mid C)$ and $p(B \mid A \wedge C') = p(B \mid A \wedge C)$ then

   $$p(A \wedge B \mid C') \geq p(A \wedge B \mid C) \quad \text{and also} \quad p(\neg A \mid C') < p(\neg A \mid C).$$

3. **Plausibility is consistent:**
   - 3.1 If a conclusion can be reached in several ways, then every possible way must lead to the same result.
   - 3.2 All available evidence must be taken into account, none can be excluded.
   - 3.3 Equivalent (isomorphic) states of knowledge must be represented by the same plausibility.

Richard T. Cox (1898–1991)
Photo: Jack Engeman

Cox's Theorem

Warning: Proof is not completely rigorous

[for details, cf. Jaynes, 2003, §2]

Cox's axioms imply

- up to monotonic transformations, plausibility of A must be represented by $0 \leq p(A) \leq 1$,

  where $p(A) = 1 \equiv$ **certainty** that A is true,

  $p(A) = 0 \equiv$ **certainty** that $\neg A$ is true.

  We will call $p(A)$ the probability of A being true. Notation: $p(A, B) \equiv p(A \wedge B)$.
- the **product rule**

$$p(A, B \mid C) = p(A \mid B, C) \cdot p(B \mid C) = p(B \mid A, C) \cdot p(A \mid C)$$

- the **sum rule**

$$p(A \mid C) + p(\neg A \mid C) = 1$$

Corollary (Bayes' Theorem)

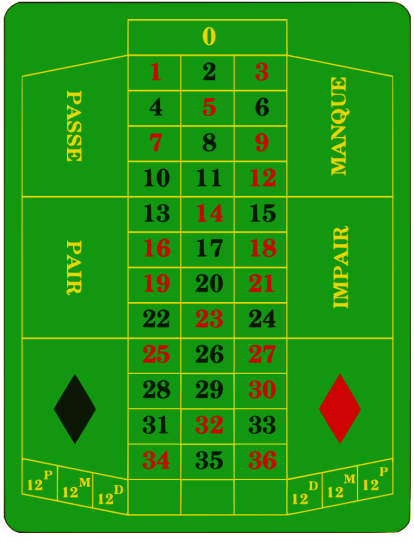
$$p(A \mid B) = \frac{p(B \mid A) \cdot p(A)}{p(B)} = \frac{p(B \mid A) \cdot p(A)}{p(B, A) + p(B, \neg A)}$$

A Different Viewpoint

Probability can be derived philosophically or defined mathematically

[image: Wikipedia]

- Cox's axioms aim to construct *autonomous agents* whose reasoning is consistent with human "common sense"
- Alternative idea: Treat *truth* as a finite (unit) amount of *mass* that can be *spread*, or *distributed*, over a set of mutually exclusive events.



Kolmogorov's Axioms

Plausibility as a Measure

[A.N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, 1933]

§ 1. Axiome[2].

Es sei E eine Menge von Elementen $\xi, \eta, \zeta, \ldots$, welche man *elementare Ereignisse* nennt, und $\mathfrak{F}$ eine Menge von Teilmengen aus E; die Elemente der Menge $\mathfrak{F}$ werden weiter *zufällige Ereignisse* genannt.

I. $\mathfrak{F}$ *ist ein Mengenkörper*[3].

II. $\mathfrak{F}$ *enthält die Menge* E.

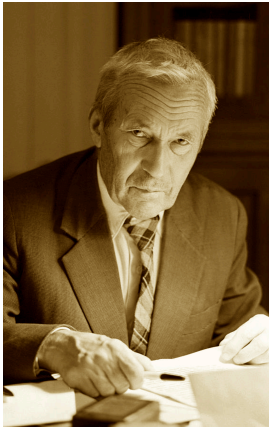
III. *Jeder Menge* A *aus* $\mathfrak{F}$ *ist eine nichtnegative reelle Zahl* $\mathsf{P}(A)$ *zugeordnet. Diese Zahl* $\mathsf{P}(A)$ *nennt man die Wahrscheinlichkeit des Ereignisses* A.

IV. $\mathsf{P}(E) = 1$.

V. *Wenn* A *und* B *disjunkt sind, so gilt*

$$\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B).$$

Ein Mengensystem $\mathfrak{F}$ mit einer bestimmten Zuordnung der Zahlen $\mathsf{P}(A)$, welche den Axiomen I—V genügt, nennt man ein *Wahrscheinlichkeitsfeld*.



Andrey N. Kolmogorov
(1903–1987)

Kolmogorov's Axioms

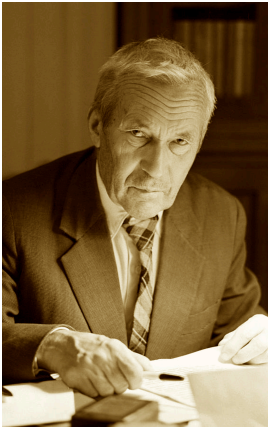
Plausibility as a Measure

[A.N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, 1933]

Definition (Kolmogorov's axioms, simplified form)

Let Ω be a space of possible events (samples,propositions,…), and let F be the set of all possible subsets of Ω. The **probability** p of an even $A \in F$ is a real map $p : F \to \mathbb{R}$ that has the following three properties

1. **non-negativity:** $0 \leq p(A) \leq 1$ for all $A \in F$.
2. **normalization:** $p(\Omega) = 1$
3. **additivity:** if A and B are mutually exclusive, then $p(A \vee B) = p(A) + p(B)$.



Andrey N. Kolmogorov
(1903–1987)

Kolmogorov's Axioms

Plausibility as a Measure

[A.N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, 1933]

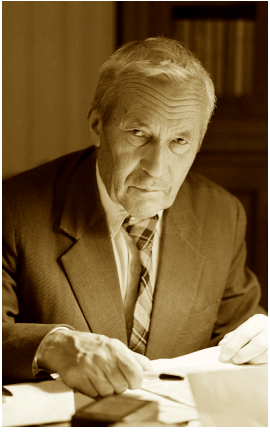
Definition (Kolmogorov's axioms, contemporary formal form)

Let (Ω, F, p) be a **measure space** (aka. Borel space). That is,

1. F is a σ-**algebra** on Ω. That is, F is a collection of subsets of Ω that is closed under countable unions (and intersections) and includes the empty set:
$$\varnothing \in F \text{ and } (f_n \in F \; \forall n \in \mathbb{N}) \Rightarrow \bigcap_{i=1}^{\infty} f_n \in F.$$
2. p is a **measure** on (Ω, F). That is, $p : F \to \mathbb{R}$ with $p(A) \geq 0$ for all $A \in F$, $p(\varnothing) = 0$ and
(σ-**additivity**): if $A_i \cap A_j = \varnothing \; \forall i, j \in \mathbb{N}$, then $p\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_i p(A_i)$.

The measure p is called a **probability measure** if

3. **normalization:** $p(\Omega) = 1$



Andrey N. Kolmogorov
(1903–1987)

Kolmogorov's Axioms

Plausibility as a Measure

[A.N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, 1933]

Definition (Kolmogorov's axioms, simplified form)

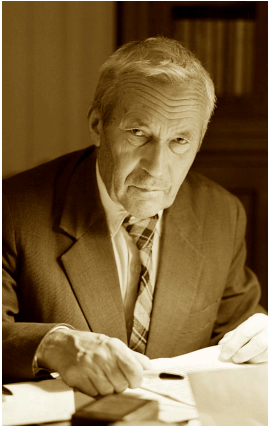
Let Ω be a space of possible events (samples,propositions,…), and let F be the set of all possible subsets of Ω. The **probability** p of an even $A \in F$ is a real map on $p : F \to \mathbb{R}$ that has the following three properties

1. **non-negativity:** $0 \leq p(A) \leq 1$ for all $A \in F$.
2. **normalization:** $p(\Omega) = 1$
3. **additivity:** if A and B are mutually exclusive, then $p(A \vee B) = p(A) + p(B)$.

Theorems: $p(A) + p(\neg A) = 1$ and $p(A) = p(A, B) + p(A, \neg B)$.
Definition: For $p(A) > 0$, define the **conditional probability**

$$p(B \mid A) := \frac{p(A, B)}{p(A)}$$

Bayes' Theorem follows directly.



Andrey N. Kolmogorov
(1903–1987)

[A.N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, 1933]

Wenn $\mathsf{P}(A) > 0$ ist, so nennt man den Quotienten

$$\mathsf{P}_A(B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(A)} \tag{5}$$

die *bedingte Wahrscheinlichkeit* des Ereignisses B unter der Bedingung A. Aus (5) folgt unmittelbar

$$\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}_A(B). \tag{6}$$

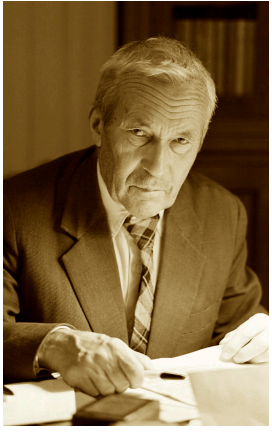
Aus (6) und der symmetrischen Formel

$$\mathsf{P}(AB) = \mathsf{P}(B)\,\mathsf{P}_B(A)$$

ergibt sich die wichtige Formel

$$\mathsf{P}_B(A) = \frac{\mathsf{P}(A)\,\mathsf{P}_A(B)}{\mathsf{P}(B)}, \tag{12}$$

welche eigentlich den Satz von BAYES enthält.



Andrey N. Kolmogorov
(1903–1987)

The Rules of Probability:

- the **Sum Rule:**

$$p(A) = p(A, B) + p(A, \neg B)$$

- the **Product Rule:**

$$p(A, B) = p(A \mid B) \cdot p(B) = p(B \mid A) \cdot p(A)$$

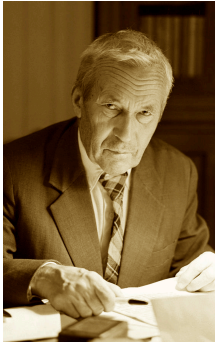
- **Bayes' Theorem:**

$$p(A \mid B) = \frac{p(B \mid A)p(A)}{p(B)} = \frac{p(B \mid A)p(A)}{p(B, A) + p(B, \neg A)}$$

Cox vs. Kolmogorov

Plausibilities are **probabilities**

Cox's Axioms	Kolmogorov's Axioms
plausibilities should comply with **common sense**	plausibilities should be a **measure** on a sample space
define $p(B \mid A)$	define $p(A)$ and $p(B \mid A)$
Bayes' theorem is derived from basic desiderata, thus very well justified	Bayes theorem is essentially defined (follows directly from definition of $p(B \mid A)$)
Philosophical argument from common sense	Rigorously derived from ad-hoc axioms



One can debate whether Bayes' theorem should be an axiom or a result. But taken together, Cox and Kolmogorov provide both a **philosophical** and a **mathematical** argument for representing plausibilities as **probabilities**.

Bayes' Theorem

Inverting Probabilities

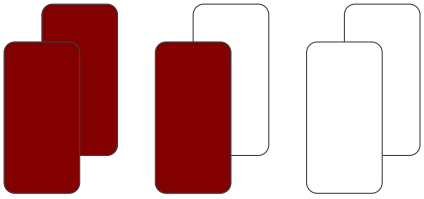
$$\underbrace{p(X \mid D)}_{\text{posterior for } X \text{ given } D} = \frac{\overbrace{p(X)}^{\text{prior for } X} \cdot \overbrace{p(D \mid X)}^{\text{likelihood for } X}}{\underbrace{p(D)}_{\text{evidence for the model}}} = \frac{p(X) \cdot p(D \mid X)}{\sum_{x \in \mathcal{X}} p(D \mid x)}$$

Bayes' Theorem tells us how to update the *belief* in a *hypothesis X* when observing *data D*.

- $p(D \mid X)$ is *the likelihood of X*, but the *(conditional) probability for D (given X)*
- the **model** is the entire thing – prior *and* likelihood
- despite the name, the prior is not necessarily what you know *before* seeing the data, but the marginal distribution $p(X) = \sum_{d \in \mathcal{D}} p(X, d)$ under *all* possible data.

Bayes' Theorem Appreciation Slides (1)

reasoning quantitatively under uncertainty



- $\frac{1}{2}$
- $2/3$
- something else
- I don't know yet

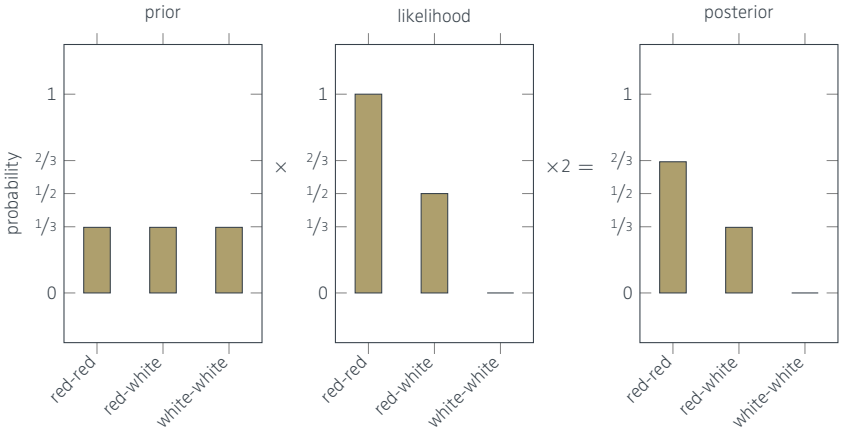
$$p(\text{card}|\text{color}) = \frac{p(\text{card}) \cdot p(\text{color} \mid \text{card})}{p(\text{color})} = \frac{p(\text{card}) \cdot p(\text{color} \mid \text{card})}{\sum_{i=1}^{3} p(\text{color} \mid \text{card} = i)}$$

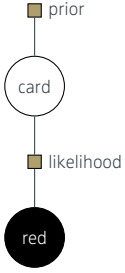
$$= \frac{1/3 \cdot p(\text{color} \mid \text{card})}{1/2} = \frac{2}{3} \cdot \{1 \quad 1/2 \quad 0\}$$

Note: It's all about the likelihood, not the prior!

Bayes' Theorem Appreciation Slides (1)

reasoning quantitatively under uncertainty



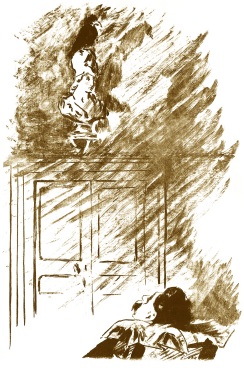


Bayes' Theorem Appreciation Slides (2)

rationality and nothing more

[Edgar Allan Poe, 1845 / Édouard Manet, 1875]

Once upon a midnight dreary, while I pondered, weak and weary,
Over many a quaint and curious volume of forgotten lore—
While I nodded, nearly napping, suddenly there came a tapping,
As of some one gently rapping, rapping at my chamber door.
"'Tis some visitor," I muttered, "tapping at my chamber door—
Only this and nothing more."



Given the tapping at your door, is it a visitor, bird, or something more? (Lenore?)

$$p(v \mid t) = \frac{p(t \mid v) \cdot p(v)}{p(t)} = \frac{p(t \mid v) \cdot p(v)}{p(t,v) + p(t,b) + p(t,m)}$$

$$p(b \mid t) = \frac{p(t \mid b) \cdot p(b)}{p(t)} = \frac{p(t \mid b) \cdot p(b)}{p(t,v) + p(t,b) + p(t,m)}$$

$$p(m \mid t) = \frac{p(t \mid m) \cdot p(m)}{p(t)} = \frac{p(t \mid m) \cdot p(m)}{p(t,v) + p(t,b) + p(t,m)}$$

Bayes' Theorem Appreciation Slides (2)

plausible reasoning

$$p(v \mid t) = \frac{p(t \mid v) \cdot p(v)}{p(t)} = \frac{p(t \mid v) \cdot p(v)}{p(t,v) + p(t,b) + p(t,m)}$$
$$p(b \mid t) = \frac{p(t \mid b) \cdot p(b)}{p(t)} = \frac{p(t \mid b) \cdot p(b)}{p(t,v) + p(t,b) + p(t,m)}$$
$$p(m \mid t) = \frac{p(t \mid m) \cdot p(m)}{p(t)} = \frac{p(t \mid m) \cdot p(m)}{p(t,v) + p(t,b) + p(t,m)}$$

- an a-priori very unlikely hypothesis can become likely if the observation is very unlikely under all other hypotheses
- even if a hypothesis strongly predicts the observation, it may remain unlikely if others predict it well, too
- at all times, the probability for all possible hypotheses sums to 1:

$$p(v) + p(b) + p(m) = 1 = p(v \mid t) + p(b \mid t) + p(m \mid t)$$

Plausible Reasoning, Revisited

have we succeeded in formalizing common sense?

$A =$ "it will begin to rain by 6pm"
$B =$ "the sky will become cloudy before 6pm"

$$A \Rightarrow B$$

if A is true, the B is true

Assume: if A is true, then B is true ($A \Rightarrow B$)

A is true	thus	B is true	(modus ponens)
B is false	thus	A is false	(modus tollens)
B is true	thus	A becomes more plausible	
A is false	thus	B becomes less plausible	

if A is true, B becomes more plausible ($p(B \mid A) > p(B)$)

A is true	thus	B becomes more plausible
B is false	thus	A becomes less plausible
B is true	thus	A becomes more plausible
A is false	thus	B becomes less plausible

Plausible Reasoning with Probabilities

we have constructed a formal system

Semantics

- **plausibility** of A is measured as **probability**, a real number $0 \leq p(A) \leq 1$
- $p(B \mid A)$ is the probability of B assuming that A is true
- $p(B)$ is the probability of B assuming nothing else
- $p(A) = 1$ is the statement that A is true assuming nothing
- $p(A) > p(B)$ is the statement that A is more probable than B

Basic laws of probability

$$p(A, B) = p(A \mid B)p(B) = p(B \mid A)p(A) \qquad \text{product rule}$$

$$p(A, B \mid C) = p(A \mid B, C)p(B, C) = p(B \mid A, C)p(A, C) \qquad \text{product rule}$$

$$p(A) + p(\neg A) = 1 \qquad \text{sum rule}$$

$$p(A \mid C) + p(\neg A \mid C) = 1 \qquad \text{sum rule}$$

Plausible Reasoning with Probabilities

In the limit of absolute certainty, probabilistic reasoning reverts to deduction.

Deductive Reasoning:

$$\frac{A \text{ is true}}{B \text{ is true}}$$

Assume if A is true, then B is true ($A \Rightarrow B$)

$$\frac{B \text{ is false}}{A \text{ is false}}$$

Probabilistic Reasoning:

- show $p(B \mid A) = 1$

Assume $p(B \mid A) = 1$

- show $p(\neg A \mid \neg B) = 1$

Plausible Reasoning with Probabilities

Probabilistic inference extends to **modes of reasoning** beyond deduction

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Plausible Reasoning:

Assume if A is true, then B is true ($A \Rightarrow B$)

$$\frac{B \text{ is true}}{A \text{ becomes more plausible}} \qquad \frac{A \text{ is false}}{B \text{ becomes less plausible}}$$

Probabilistic Reasoning:

Assume $p(B \mid A) = 1$

- show $p(B \mid A) \geq p(A)$
- show $p(B \mid \neg A) \leq p(B)$

Plausible Reasoning with Probabilities

Probabilistic inference extends beyond, but is consistent with, deduction

Lemma

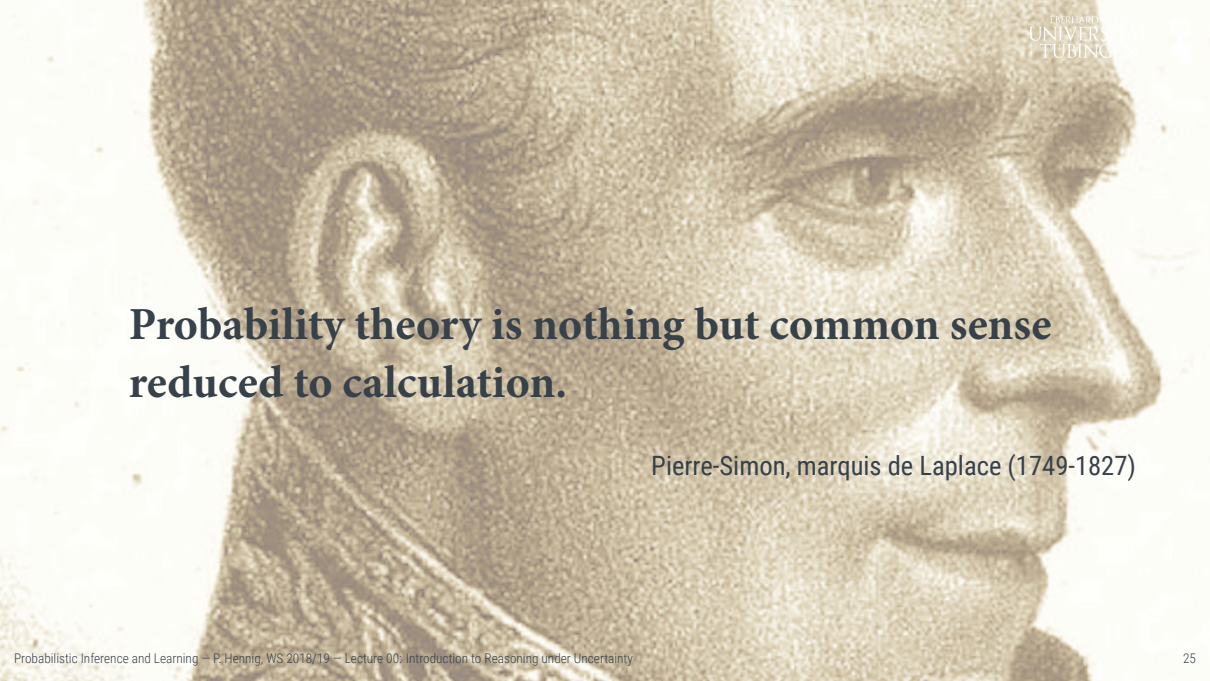
$p(B \mid A) = 1$ *implies*	*if A is true, then B is true*
✦ $p(B \mid A) = 1$ *"modus ponens"*	*A is true implies B is true*
✦ $p(B \mid \neg A) \leq p(B)$	*A is false implies B becomes less plausible*
✦ $p(A \mid B) \geq p(A)$	*B is true implies A becomes more plausible*
✦ $p(\neg A \mid \neg B) = 1$ *or, equivalently,* $p(A \mid \neg B) = 0$ *"modus tollens"*	*B is false implies A is false*

Plausible Reasoning with Probabilities

Probabilistic inference extends to **types of knowledge** where deductive reasoning does not apply

Lemma

$p(B \mid A) \geq p(B)$ *implies*	*if A is true, then B becomes more plausible*
✦ $p(B \mid A) \geq p(B)$	*A is true implies B becomes more plausible*
✦ $p(B \mid \neg A) \leq p(B)$	*A is false implies B becomes less plausible*
✦ $p(A \mid B) \geq p(A)$	*B is true implies A becomes more plausible*
✦ $p(\neg A \mid \neg B) \geq p(\neg A)$	*B is false implies A becomes less plausible*



Probability theory is nothing but common sense reduced to calculation.

Pierre-Simon, marquis de Laplace (1749-1827)

Administrative Stuff

Ilias course password: *Kolmogorov*

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Lectures	Mondays and Wednesdays, 8ct – 10
Exercise Groups	please vote now
Exercise Sheets	in Ilias, after lecture
Slides	in Ilias, *after* lecture
Break	please express opinions

Please fill out instant feedback after **every** lecture!

Exam

more admin

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

The Exam will be held on **Wednesday, 13 February 2019**, in the Kupferbau / Neue Aula

- You have to achieve **at least 30% of the exercise sheet points** to be admitted to the exam
- There are 12 Exercise sheets. Each sheet yields exactly 100 points
- The final grade consists entirely (100%) of the exam grade
- A second exam (*Nachklausur*) will only be held if necessary.
  Participation in first exam is mandatory to enter 2nd exam
  (Eventuelle Nachprüfung mit Hauptprüfungspflicht).
- **If** there is a second exam, it will be held on **Monday, 8 April 2019** (likely in F119 / Sand)

Some of the exercises are deliberately a bit harder, but also more instructive. Do not worry about the difficulty of the exercises. They are meant to help you learn, and are not representative for the exam. Remember that you only need 30%, and the exercise sheets do not count toward the exam grade.

Literature

only for further reading & reference

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

- **David J C MacKay** – Cambridge, 2003
  Information Theory, Inference, and Learning Algorithms
  `http://www.inference.phy.cam.ac.uk/itprnn/book.pdf`
- **David Barber** – Cambridge, 2012
  Bayesian Reasoning and Machine Learning
  `http://web4.cs.ucl.ac.uk/staff/D.Barber/textbook/270212.pdf`
- **Carl E Rasmussen & Christopher K I Williams** – MIT, 2006
  Gaussian Processes for Machine Learning
  `http://www.gaussianprocess.org/gpml/chapters/RW.pdf`
- **Edwin T Jaynes & G Larry Bretthorst** – Cambridge, 2003
  Probability Theory – the Logic of Science
  `https://bayes.wustl.edu/etj/prob/book.pdf` (first 70 pages only, but legal)

- **Judea Pearl** – Morgan Kaufmann, 1988
  Probabilistic Reasoning in Intelligent Systems
- **Christopher Bishop** – Springer, 2007
  Pattern Recognition and Machine Learning

Some ideas and techniques you will encounter in this course

Outlook (plan subject to adaptive changes)

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

- Connections between probabilistic inference and Boolean logic
- Learning functional relationships between variables. Code examples:
  - How to build a model for your body
  - How to rate a human's credit worthiness
- The world's fastest learning algorithms
- A generalization from "shallow" and "deep" to "structured" learning
- Formal and symbolic languages for artificial intelligence
- A general toolbox for encoding structured domain knowledge in a learning agent, and transferring it into a concrete algorithm
- And much more …

You may enjoy this course if

- you are looking for a **joint, connected, holistic** view on reasoning, inference, learning, intelligence
- you are not afraid of **math**, but see it more as a tool than the ultimate goal. I will try to balance **intuition and pictures** with mathematical analysis and code

Relationship to Other Approaches to Machine Learning

Why should you want to take this course?

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Statistical Learning Theory	Neural / Deep learning	Probabilistic Learning
formulate a loss-function *ad hoc*, mapping data to predictions/decisions. Then show that, under some external assumptions, this model has certain desirable properties.	build models from **intuition** or to emulate aspects of biological neural systems. Make them do cool things	formulate a **generative model**. Inference is then uniquely determined by Bayes' theorem. No need to question or analyse the paradigm over and over again
mathematical **analysis** in the foreground (often in the asymptotic or large-number limit)	**empirical** evaluation in the foreground	**numerical & computational** design in the foreground (the right model may be intractable)
statements about errors tend to focus on the **worst-case**	error analysis hard or non-existent, even ill-posed	structured and extensive **quantification of uncertainty** by the posterior, often core motivation

You are in the unique position to have access to lectures from all three viewpoints!

To be able to reason in an uncertain world,
whether to build intelligent machines or find scientific insight,
you have to understand probabilities.

- both philosophical (Cox) and formal (Kolmogorov) arguments lead to the same two rules:

  sum rule: $P(X) = p(X, Y) + p(X, \neg Y)$

  product rule: $P(X, Y) = P(X) \cdot P(Y \mid X)$

- their corollary, **Bayes' Theorem** provides the mechanism for **inference**:

$$\underbrace{p(X \mid D)}_{\text{posterior of } X \text{ given } D} = \frac{\overbrace{p(D \mid X)}^{\text{likelihood of } X \text{ under } D} \cdot \overbrace{p(X)}^{\text{prior of } X}}{\underbrace{p(D)}_{\text{evidence for the model}}}$$

- the fundamentality of probabilities has been debated at length. Probabilities are not the only inference system, but they are uniquely **general, expressive, and powerful**.
- **Machine learning and AI** can be approached in various ways. The probabilistic viewpoint is the closest we have to a **theory of everything** for ML.

Become a Bayesian

inference as a science, a technology, and an art

Bayesians are careful people. They want to know the assumptions encoded in their model, and the information provided by the data. They care about computations, structure, code.
There are no black-box Bayesians.