Welcome to the module

Decision Theory in Summer Semester 2023 !

Thomas Augustin & Christoph Jansen

Foundations of Statistics and Their Applications,
Department for Statistics, Ludwig-Maximilians-Universität Munich
http://www.foundstat.statistik.uni-muenchen.de/index.html

Off Topic

Welcome Event

for Freshers to the Master's Programme

Today 6.15 pm in M203 (main building)

Supporting material for the lecture

Decision Theory 2023

Thomas Augustin and Christoph Jansen

Foundations of Statistics and Their Applications,
Department for Statistics, Ludwig-Maximilians-Universität München (LMU Munich)
http://www.foundstat.statistik.uni-muenchen.de/index.html

Please note: This material is **not** designed as a self-explanatory script; rather, it is intended to facilitate focusing and note-taking during lectures by providing writing-intensive parts in typed form. Examples, crucial ideas and important threads of argumentation shall be elaborated interactively.

General Introduction

The omnipresence and basic structure of decisions

Fundamental learning objectives of the course

- Learn to design and solve decision problems in various areas of application

- Understand ‘common statistics and machine learning’ better by adopting an outside perspective
  + frame statistical and machine learning methods as decisions
  + reflect on the notion of uncertainty

The dilemma of the first lesson

Two quite opposite aspects

- It is necessary to discuss and clarify the organizational framework.

- It is a good academic tradition that the lecturer provides an overview of the planned lecture, including a “self-positioning”’.

The lecturers

- Thomas Augustin
- Christoph Jansen
- Appointment by mail `first_name.surname@stat.uni-muenchen.de`
- Questions, of course, also during or directly after the class.
- Please, make also use of the Moodle-Etherpad!

Options for obtaining ECTS points by attending this course

- This course consists of two parts

a) In the first two thirds of the lecture period (equivalent to 3 SWS lecture + 1 SWS exercise class): Basic concepts

b) In the last third: Deepening of some topics and overview of other areas, partly insight into current research

- Beyond this, further specialization in decision theory possible: seminar (in the break between winter and summer semester), Master's thesis; Methodological Research Colloquium in development

Options for obtaining ECTS points by attending this course

- Examination covering
  - ◇ a) only$\Rightarrow$: 6 ECTS for the module(-part) *decision theory*

or

  - ◇ a) and b)$\Rightarrow$
    - ◇ 6 ECTS for module(-part) *decision theory* + 3 ECTS to be used flexibly, with separate grades
    - ◇ 6 ECTS for module(-part) *decision theory* + 6 ECTS to be used flexibly, with separate grades, if in addition a term paper is submitted

    - ◇ a) and b) can also be split over two years

For students of Statistics / Statistics and Data Science this is already approved by the Board of Examiners. Students from other degree programmes have to approach the Board of Examiners responsible for them.

Planned division of the course

- TA will start and teach until May 23rd.
- CJ will continue after Whitsun, starting on June 5th.
- CJ will also mainly teach the 3 ECTS part, with some lectures in between delivered by TA.

A truly hard decision problem: Which format to choose for this course?

- The most intensive and successful learning is active learning by reflecting and discussing $\rightarrow$ high interactive character $\rightarrow$ face-to-face sessions.

- Videos offer new opportunities for dynamic, individual learning iff they are used in the right way $\rightarrow$ inverted classroom.

Start with a reasonable and serious attempt, and then adjust if necessary!

What do we suggest concretely?

- Face-to-face sessions, hopefully highly interactive

- If you wish we could make videos of last year's lecture available.

- A few parts may also be presented in an asynchronous way as video, especially if different prior knowledge exists or very technical aspects are taught. Then the lecture times will be adjusted accordingly.

Nota bene

- Resist the temptations involved!
- Videos just contain the frontal teaching part; students' contributions and discussions are cut.
- No videos on excercise classes
- Videos may be (slightly) outdated

Times and Locations

The reserved slots for the course are:

- Monday: 10:00 – 12:00 (A016 Main Building)
- Tuesday: 17:00 - 21:00 (A014 Main Building),

We have to cover approximately 270 minutes per week (minus a few special inverted classroom videos).

Supporting material for the lecture: slides

- Shall support note-taking and interactivity.
- Will be provided in advance: “5-minute brief preparation”
- Deliberately not a complete, self-explaining script
- Interactive elaboration of main ideas

Moodle as a communication platform

- Please enroll in Moodle course.
- Announcements to all participants
- Download of supporting material and exercise sheets
- Links to video material
- Forum, questions via easy to handle “etherpad”
- Infos concerning further planning
- . . .

NEW: In addition, two registrations in the LSF mandatory

- At the beginning of the lecture.
- For the exam

Code of conduct I

- Let's create a productive atmosphere, valuing the possibility of coming together again!
- Let's go beyond one-way teaching! Lecture between knowledge dissemination and knowledge production.
- Knowledge production is an emergent process where everyone's contribution is needed.
- Bring in your special individual background!
- Research lives from mutual exchange, creative ideas and their critics.
- Content-related questions are never silly; the same is true for corresponding answers.

Code of conduct II

- The spoken word is protected by law. The participants' privacy is to be respected under all circumstances (criminal prosecution!)
- No self-recording!
- Use of such recorded material for bullying, ridiculing, etc. someone: also massive consequences according to examination regulations

1 Introduction

1.1 Characterization of decision theory as a theory of rational decision making under uncertainty

1.1.1 Again: Omnipresence and basic structure of decision problems

Again: Fundamental learning objectives of the course

- Learn to design and solve decision problems in various areas of application

- Understand 'common statistics and machine learning' better by adopting an outside perspective
  + frame statistical and machine learning methods as decisions
  + reflect on the notion of uncertainty

1.1.2 Interdisciplinary importance and threads of development

- Rational Choice Theory
- „Uncertainty in Artificial Intelligence“, Expert Systems, Decision Support Systems
- Statistical Decision Theory

.

Rational Choice

- Macro-level by aggregation of individual actions.
- Modelling of goal-oriented (rational) actions of utility-maximizing actors (i.e., actors that fulfil their own preferences as well as possible) for a better understanding of real and fictitious phenomena and systems. (e.g. market events); relies mainly on game-theoretical considerations (strategic interdependencies).
- Paradigmatic shift in social sciences: no longer natural sciences as model.
- Major developments in economics
- In sociology e.g. Coleman: Foundations of Social Theory (1990) or Braun & Gautschi (2011): rational choice theory.
- Also strong impact on politology and practical philosophy

- „Modern“: Bounded Rationality, Behavioural Economics, Prospect-Theory (Kahneman & Tversky 1979, Nobel Prize 2002),
  - „Normality“ of „anomalies“.
  - Why do individuals not behave strictly rationally in the sense of classical economic theory?
  - Experimental labs in economics and social sciences MELESSA: Munich Experimental Laboratory for Economic and Social Sciences
  - MCMP (Munich Center for Mathematical Philosophy),

Second branch: „Uncertainty in Artificial Intelligence“, expert systems, decision support systems

- Modelling of the decision base of experts for optimal action in similar situations.
- Examples for application areas
  - business administration (investment decisions)
  - economics (portfolio management)
  - medicine (diagnostic systems)
  - engineering (control)

Expert systems: Typical tasks of knowledge representation

- Help to find optimal decisions in systems that are too large to oversee everything.
- Make expert knowledge widely available.
- Also: decrease in routine decisions to enable concentration on the essentials.

Decision support issues in automated driving systems

The following figures are taken from: Caballeroa, W.N., Ríos Insua, D., Banks, D. (2023): Decision support issues in automated driving systems. *International Transactions in Operational Research* **30**, 1216-1244. `DOI:10.1111/itor.12936`

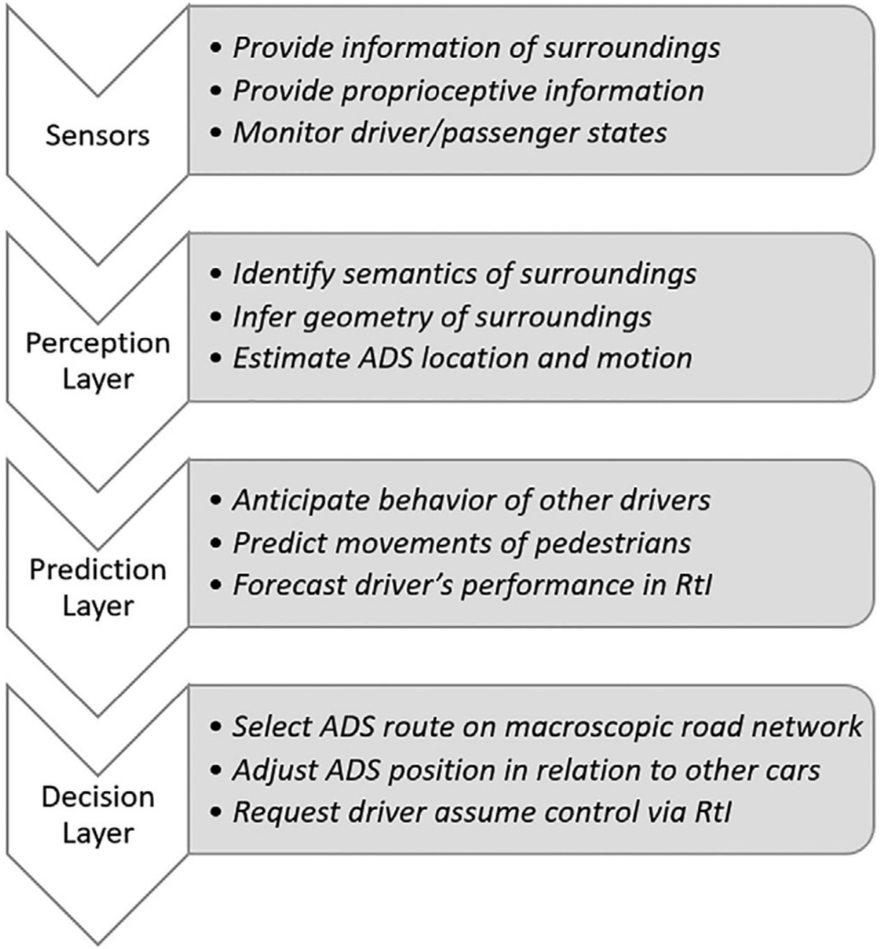
Levels of driving automation[1]

Levels of driving automation (Society of Automobile Engineers, 2018)

Level	Name	Description
0	No automation	The driver performs all driving tasks, but vehicle features may provide some decision support (e.g., blind spot and lane-departure warnings) and limited assistance.
1	Assisted automation	Automation for specific activities to include steering or braking support (e.g., lane keeping assistance and adaptive cruise control).
2	Partial automation	Combination of two or more level-1 features that provide steering and braking/acceleration support.
3	Conditional automation	Self-driving automation with full control of all critical safety functions under certain conditions (e.g., traffic jam chauffeur). The driver is still expected to take over in some instances.
4	High automation	Vehicles are fully self-driving, without need for human intervention. The automated driving system controls within a prescribed operational domain (e.g., local driverless taxi).
5	Full automation	The automated driving system can operate the vehicle under all on-road conditions with no design-based restrictions.

[1] Taken from Caballeroa, Ríos Insua, Banks (2023, p. 1217)

Layered ADS architecture[2]



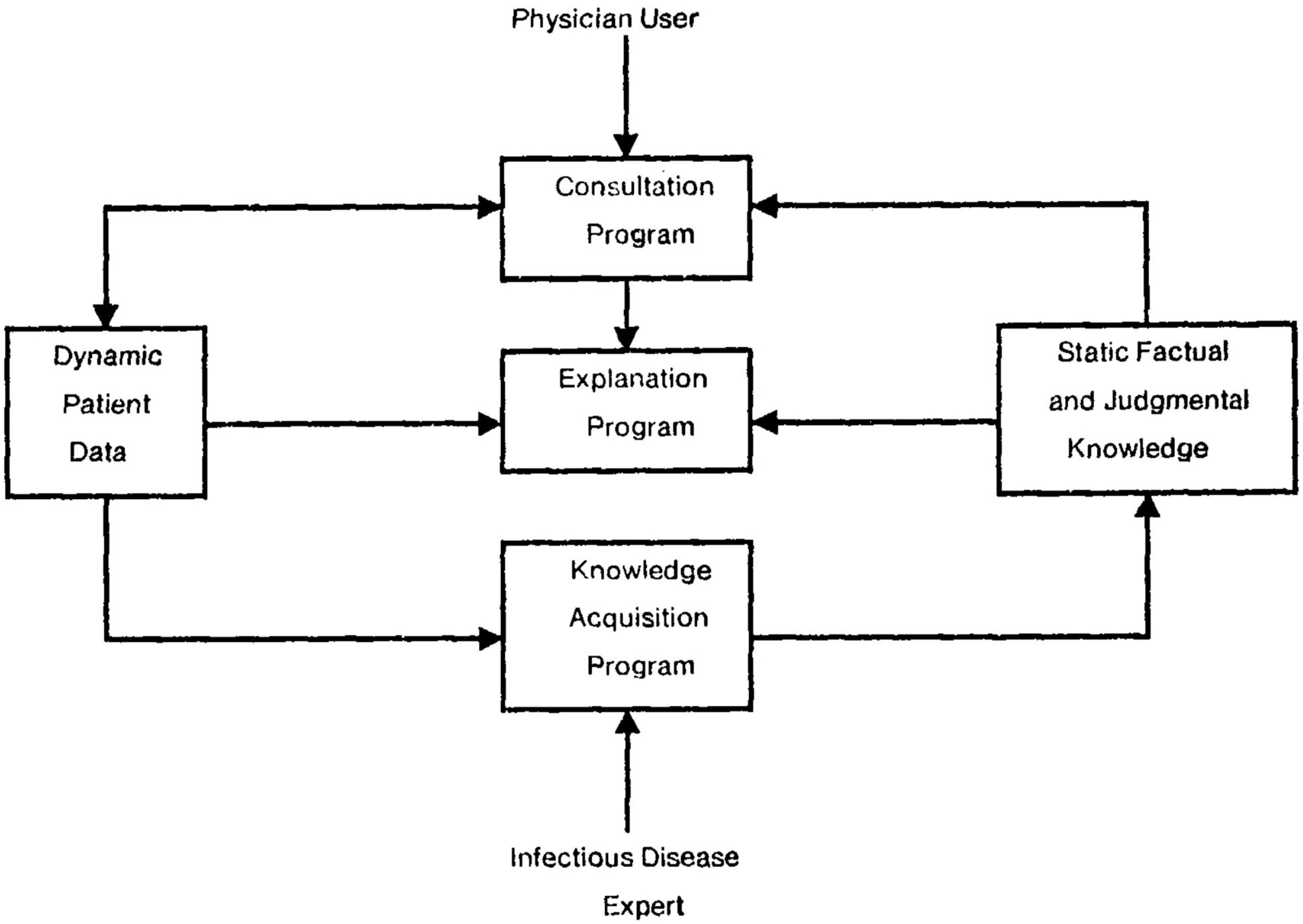
[2] Figure taken from Caballeroa, Ríos Insua, Banks (2023, p. 1220)

RtI: request-to-intervene

The MYCIN Project

- Expert support for physicians in efficient antibiotics treatment (name often -mycin).
- Excellent documentation of the research project: Buchanan, B.G. & Shortliffe, E.H. (ed.) (1984). *Rule Based Expert Systems. The MYCIN Experiment of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading (MA).
  Documented online via `www.shortliffe.net/Buchanan-Shortliffe-1984/MYCINBook.htm`, last access April 16th, 2023
- system applies typical 'question tree' of an expert and then, after some 50 to 60 questions, specifies likely bacteria and recommends a therapy

MYCIN: Basic architecture[3]



[3]Figure taken from van Melle, W. in Buchanan & Shortliffe (1984), Chapter 4, here p. 68

An exemplary session[4]

The Consultation Program 69

1) Patient's name: (first-last)
** **FRED BRAUN**
2) Sex:
** **M**
3) Age:
** **55**

(Questions 1-3 request background patient data.)

4) Are there any cultures for Fred Braun which may be related to the present illness, and from which organisms have been grown successfully in the microbiology laboratory?
** **Y**

--------CULTURE-1--------
5) From what site was the specimen for CULTURE-1 taken?
** **BLOD**
=BLOOD *(System corrects typing error.)*
6) Please give the date and time when CULTURE-1 was obtained. (mo/da/yr time)
** **JUN 20, 1977**
The first organism isolated from the blood culture of 20-JUN-77 (CULTURE-1) will be referred to as:

--------ORGANISM-1--------
7) Enter the laboratory-reported identity of ORGANISM-1:
****UNKNOWN** *(Typically identity is not yet known.)*
8) The stain (gram or Ziehl-Neelsen acid-fast) of ORGANISM-1:
** **NEG** *(Preliminary lab results give some clues.)*
9) Is ORGANISM-1 rod or coccus (etc.):
** **ROD**
10) What is the form of the individual organisms (e.g. lancet-shaped for cocci, fusiform for rods, etc.)?
** **FUSIFORM**

(. . . more questions follow in order to gather sufficient information to infer the identity and significance of the infecting organisms . . .)

23) What is the suspected portal of entry of the gramneg rod from the blood culture of 20-JUN-77 (ORGANISM-1) into this sterile site?
** **GI (6)** *(User indicates the uncertainty of his response.)*
24) Is Fred Braun a burn patient?
** **N**
25) Is Fred Braun a compromised host (e.g. alcoholic, sickle-cell-disease, immunosuppressed)?
** **N**
26) Is Fred Braun febrile?
** **Y**

(. . . after about 50–60 questions, MYCIN prints the diagnostic hypotheses on which therapy will be based . . .)

My therapy recommendations will be designed to treat for organisms that are either very likely or, although less likely, would have a significant effect on therapy selection if they were present. It is important to cover for the following probable infection(s) and associated organism(s):

INFECTION-1 is BACTEREMIA

<ITEM-1> E.COLI [ORGANISM-1]
<ITEM-2> KLEBSIELLA [ORGANISM-1]
<ITEM-3> ENTEROBACTER [ORGANISM-1]
<ITEM-4> KLEBSIELLA-PNEUMONIAE [ORGANISM-1]

FIGURE 4-2 Excerpts from a MYCIN consultation. (Comments in italics are not part of the actual interaction.)

70 **The Structure of the MYCIN System**

(. . . questions follow to evaluate possible therapy choices, and finally MYCIN prints its therapy recommendations . . .)

[REC-1] My preferred therapy recommendation is as follows:
In order to cover for items <1 2 3 4>:
Give: GENTAMICIN *(One drug covers 4 possible identities.)*
Dose: 119 mg (6.0 ml) q8h IV for 10 days [calculated on basis of 1.7 mg/kg]
Comments: Modify dose in renal failure.

FIGURE 4-2 continued

objects, known as *contexts* in MYCIN, are such things as individual cultures taken from the patient, organisms that grew out of them, and drugs the patient is currently receiving. Various attributes, termed *clinical parameters*, characterize these objects. Questions asked during the consultation attempt to fill in the values for relevant attributes of these objects. To represent the uncertainty of data or competing hypotheses, attached to each triple is a *certainty factor* (CF), a number between -1 and 1 indicating the strength of the belief in (or a measure of the importance of) that fact. A CF of 1 represents total certainty of the truth of the fact, while a CF of -1 represents certainty regarding the negation of the fact. While certainty factors are *not* conditional probabilities, they are informally based on probability theory (see Part Four). Some triples (with CF's) from a typical consultation might be as follows:

(IDENTITY ORGANISM-1 PSEUDOMONAS 0.8)
(IDENTITY ORGANISM-1 E. COLI 0.15)
(SITE CULTURE-2 THROAT 1.0)
(BURNED PATIENT-298 YES -1.0)

Here ORGANISM-1 is probably *Pseudomonas*, but there is some evidence to believe it is *E. coli*; the site of CULTURE-2 is (without doubt) the throat; and PATIENT-298 is known *not* to be a burn patient.

4.1.2 Production Rules

MYCIN reasons about its domain using judgmental knowledge encoded as production rules. Each rule has a *premise*, which is a conjunction of predicates regarding triples in the knowledge base. If the premise is true, the conclusion in the *action* part of the rule is drawn. If the premise is known with less than certainty, the strength of the conclusion is modified accordingly.

A typical rule is shown in Figure 4-3. The predicates (such as SAME) are simple LISP functions operating on associative triples, which match the declared facts in the premise clause of the rule against the dynamic data known so far about the patient. $AND, the multi-valued analogue of

[4] Copied from van Melle, W. in Buchanan & Shortliffe (1984), Chapter 4, here p. 68

MYCIN: Methodological Consequences

Modelling uncertain (expert) knowledge: *Uncertainty in Artificial Intelligence (UAI)*

- Probability theory as a concept fails in modelling expert knowledge

MYCIN: Methodological Consequences

Modelling uncertain (expert) knowledge: *Uncertainty in Artificial Intelligence (UAI)*

- Probability theory as a concept fails in modeling expert knowledge.
- Vagueness of concepts
- Probability evaluations require a (too) high degree of
  - precision and
  - internal consistency

Indivisible evidence versus additivity axiom

Beyond the traditional concept of probability[5]

"The drawbacks of pure probabilistic methods and of the certainty factor model have led us in recent years to consider alternate approaches. Particularly appealing is the mathematical theory of evidence developed by Arthur Dempster. [...]
We believe that the advantage of the Dempster-Shafer theory over previous approaches is its ability to model the narrowing of the hypothesis set with the accumulation of evidence, a process that characterizes diag- nostic reasoning in medicine and expert reasoning in general. An expert uses evidence that, instead of bearing on a single hypothesis in the original hypothesis set, often bears on a larger subset of this set. [....]

[5] From Gordon, J. & Shortliffe, E.H. in in Buchanan & Shortliffe (1984), Chapter 13, here p. 272f.

Continued

For example, in the search for the identity of an infecting organism, a smear showing gram-negative organisms narrows the hypothesis set of all possible organisms to a proper subset. This subset can also be thought of as a new hypothesis: the organism is one of the gram-negative organisms. However, this piece of evidence gives no information concerning the relative likelihoods of the organisms in the subset. Bayesians might assume equal priors and distribute the weight of this evidence equally among the gram-negative organisms, but, as Shafer points out, they would thus fail to distinguish between uncertainty, or lack of knowledge, and equal certainty. Because he attributes belief to subsets, as well as to individual elements of the hypothesis set, we believe that Shafer more accurately reflects the evidence-gathering process. [...]

Evolution of uncertainty theories (mind the plural!)

Klir and Wierman (Uncertainty-based Information, Physika, 1999^2, p. 1)

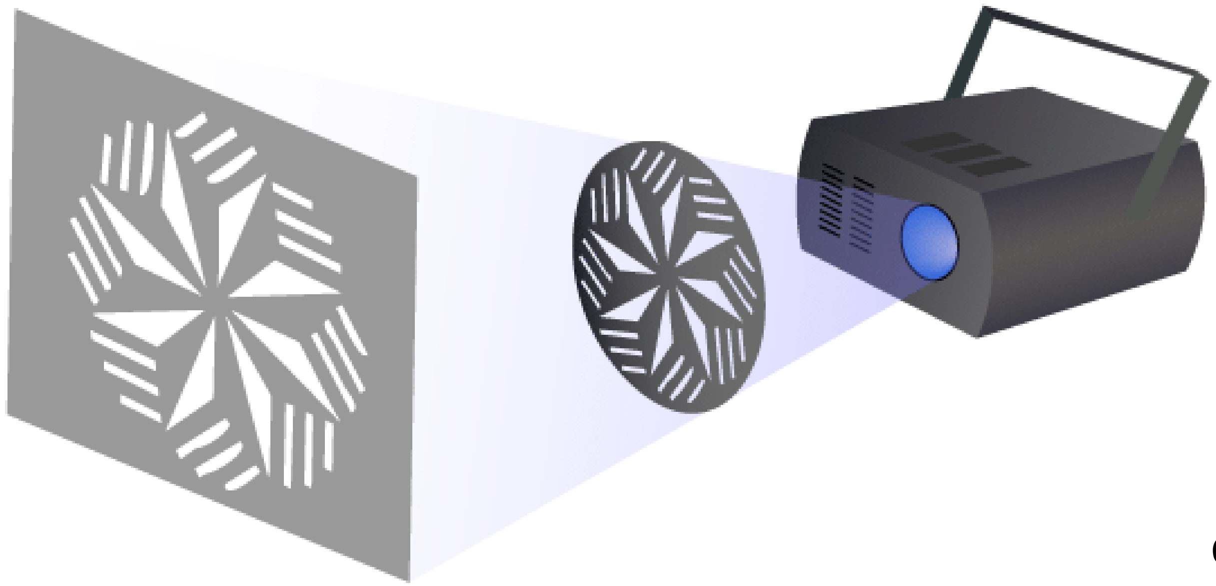
„For three hundred years [...] uncertainty was conceived solely in terms of probability theory. This seemingly unique connection between uncertainty and probability is now challenged [... by several other] theories, which are demonstrably capable of characterizing situations under uncertainty. [...]

[...] it become clear that there are several distinct types of uncertainty. That is, it was realized that uncertainty is a multidimensional concept. [.... That] multidimensional nature of uncertainty was obscured when uncertainty was conceived solely in terms of probability theory, in which it is manifested by only one of its dimensions“.

Projection

"[.... That] multidimensional nature of uncertainty was obscured when uncertainty was conceived solely in terms of [traditional [added: TA]] probability theory, in which it manifested by only one of its dimensions"

. . .



6

[6] https://commons.wikimedia.org/wiki/File:Gobo_projected_illustration.png, Jeremy Kemp, free [last access April 16th, 2023]

Uncertainty in Machine Learning

Mail 20.4.2020 (UAI-List) by Sebastian Destercke / Eyke Hüllermeier

```
------------------------------------------------
C A L L F O R P A P E R S

Uncertainty in Machine Learning

Workshop (combined with a tutorial) at ECML/PKDD 2020 September 14-18,
2020, Ghent, Belgium

https://sites.google.com/view/wuml-2020/ [last access April 24th, 2022]
```

“Motivation and Focus

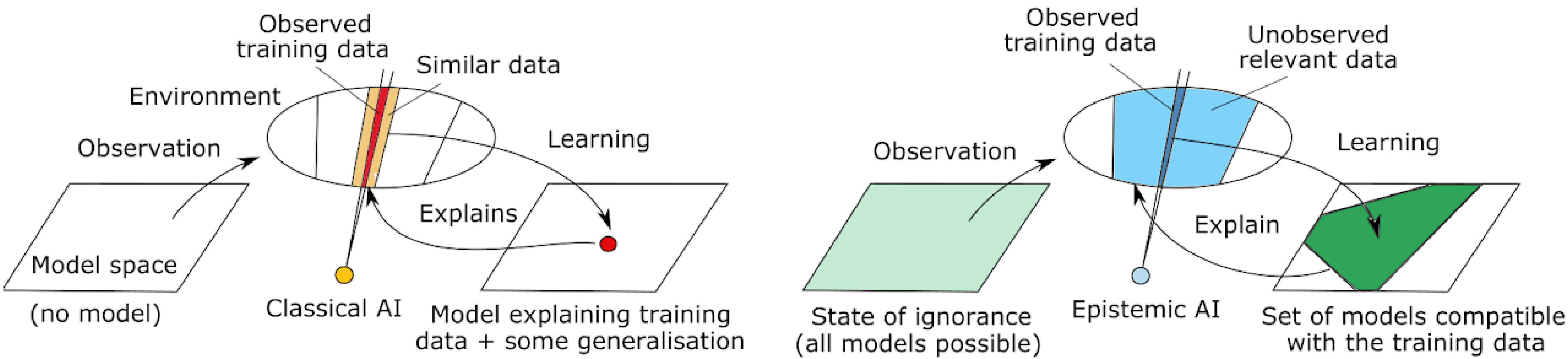
The notion of uncertainty is of major importance in machine learning and constitutes a key element of modern machine learning methodology. In recent years, it has gained in importance due to the increasing relevance of machine learning for practical applications, many of which are coming with safety requirements. In this regard, new problems and challenges have been identified by machine learning scholars, which call for new methodological developments. Indeed, while uncertainty has long been perceived as almost synonymous with standard probability and probabilistic predictions, recent research has gone beyond traditional approaches and also leverages more general formalisms and uncertainty calculi. For example, a distinction between different sources and types of uncertainty, such as aleatoric and epistemic uncertainty, turns out to be useful in many machine learning applications. The workshop will pay specific attention to recent developments of this kind. [...] ”

The epistemic artificial intelligence project

"While traditional ML learns from the (limited) available evidence a model able to describe it, with limited power of generalisation (see the figure [... on the next slide], left), epistemic AI (right) starts by assuming that the task at hand is (almost) completely unknown, because of the sheer imbalance between what we know and what we do not know. [...] Mathematically, Epistemic AI's principle translates into seeking to learn sets of hypotheses compatible with the (scarce) data available, rather than individual models. A set of models can provide, given new data, a robust set of predictions among which the most cautious one can be adopted, thus avoiding catastrophic results."

`https://www.epistemic-ai.eu`, last access April 16th, 2023

The epistemic artificial intelligence project



"Illustration of the concept of epistemic artificial intelligence. Epistemic AI's notion of learning (right), as opposed to that of traditional machine learning/artificial intelligence (left)."

`https://www.epistemic-ai.eu`, last access April 16th, 2023

1.1.3 Third branch: statistical decision theory

- Decision theory in statistics and ML
  - Probability considerations often play a role in the evaluation of the states of nature
  - data-dependent decisions
    - processing of data-based information
    - cost of information

  $\Rightarrow$ Role of decision theory as a statistical subdiscipline.

- On the other hand, decision theory can be seen as a formal „superstructure“ over testing, estimating and prediction: optimal statistical and ML procedures as optimal solution of a decision problem.

- Very successful and fashionable for a while; then more in the background. Renaissance in the context of machine learning.

1.1.4 Preliminary outline

Part 1: Fundamentals of Decision Theory

1. The basic structure of decision problems

1.1 Characterization of decision theory as a theory of rational decision making under uncertainty.

1.2 The basic form of a data-free decision problem

1.3 Typical examples

1.4 Randomized actions

1.5 Decision making based on data: statistical decision theory as a special case of data-free decision problems

2. Optimal decision making under uncertainty: decision criteria

2.1 Decision rules – decision principles; Dominance principle and admissibility (Pareto front)

2.2 Minimax decisions as 'virtual games against nature'

2.3 Bayes decisions in the 'virtual risk situation'

2.4 Some alternative rules in the context of classical decision theory

3. Basics for calculating optimal actions in large scale systems

3.1 Convex sets

3.2 Linear optimization

3.3 Determination of optimal (randomized) actions

3.3.1 Bayes criterion

3.3.2 Maximin criterion

3.3.3 Hodges and Lehmann criterion

4. Decision problems under a more general concept of probability

4.1 Generalized Probability

4.2 Decision criteria under generalized probability

Part 2: Advanced Topics

4.3 Efficient computation of optimal (randomized) actions under generalized probabilities via linear optimization

4.3.1 Max-E-Min criterion and Choquet expectation utility.

4.3.2 E-admissibility

5. Fundamentals of utility theory

5.1 Decision making under ordinal utility

5.2 Decision making under cardinal utility

5.3 Outlook: Partial cardinal utility

6. Outlook on other topics in decision theory

6.1 Robust Bayesian inference from a decision-theoretic point of view.

6.2 Likelihood-based decision theory

6.3 Decision theory and robust statistics

6.4 Further selected topics

Basic literature (of varying level of difficulty and fit)

- **Berger**, J.: *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, 1985^2.
- **Chernoff**, H., **Moses**, L.: *Elementary Decision Theory.* Wiley, New York, 1959, reprinted Dover, 2003.
- **Dimitrakakis**, C. & **Ortner**, R. (2022): *Decision Making Under Uncertainty and Reinforcement Learning.* Springer.
  preprint `www.cse.chalmers.se/~chrdimi/downloads/book.pdf`, Version of April 8th, 2021, last access April 15th, 2023.
- **Gilboa**, I.: *Rational Choice.* MIT Press, Cambridge (MA), 2010.
- **Liese**, F., **Mieschke**, K.J. (2008): *Statistical Decision Theory: Estimation, Testing, and Selection.* Springer, New York,

- **Parmigiani**, G., **Inoue**, L (2009): Decision Theory: Principles and Approaches. Wiley, Chichester.

- **Peterson**, M.: *An Introduction to Decision Theory.* Cambridge University Press, Cambridge, 2017^2.

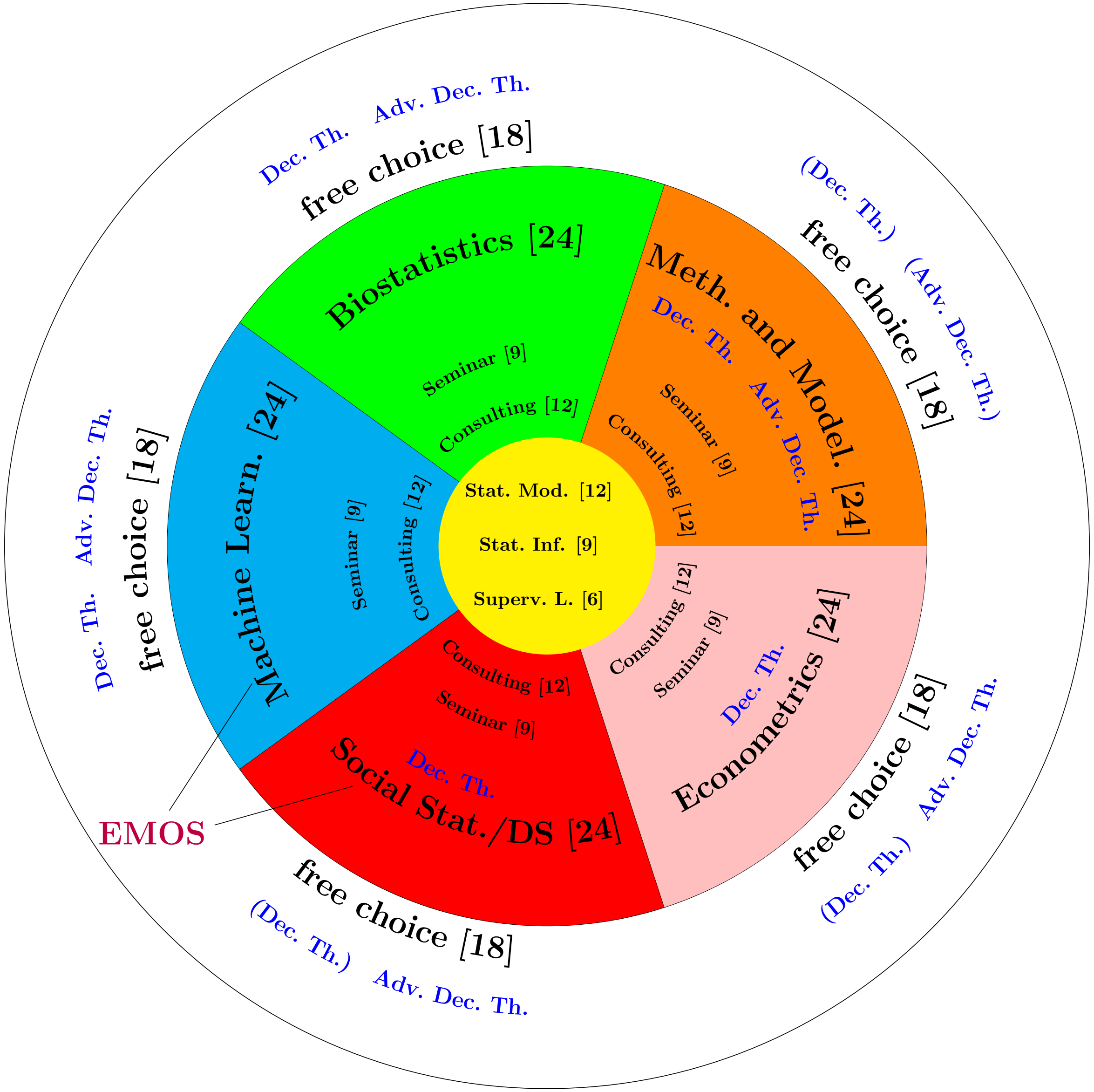
Chapter-specific literature is given there.

1.1.5 Further information on the organization of the module and its basic concept

The Module *Decision Theory* in the New Master's Programme

- machine learning, biostatistics: general elective (free 18 ECTS)
- social statistics and social data science: semi-narrow elective (at least 12 ECTS from...), general elective
- econometrics: semi-narrow elective (at least 6 ECTS from....), general elective
- methodology and modelling: narrow elective (at least 2 of 3 modules), semi-narrow and general elective

6 ECTS *Advanced Decision Theory* can be recognized, for instance, for *Methodological Discourses in Statistics and Data Science*.


Stat. Mod. [12]
Stat. Inf. [9]
Superv. L. [6]
Biostatistics [24]
Seminar [9]
Consulting [12]
free choice [18]
Dec. Th.
Adv. Dec. Th.
Meth. and Model. [24]
Dec. Th.
Adv. Dec. Th.
Seminar [9]
Consulting [12]
free choice [18]
(Dec. Th.)
(Adv. Dec. Th.)
Econometrics [24]
Dec. Th.
Seminar [9]
Consulting [12]
free choice [18]
Adv. Dec. Th.
(Dec. Th.)
Social Stat./DS [24]
Dec. Th.
Seminar [9]
Consulting [12]
free choice [18]
(Dec. Th.)
Adv. Dec. Th.
Machine Learn. [24]
Seminar [9]
Consulting [12]
free choice [18]
Dec. Th.
Adv. Dec. Th.
EMOS

Decision Theory in other Degree Programmes

- Narrow elective in “WiSo-Master”; elective in EMOS-variant
- Elective in “old” master’s programs statistics/biostatistics, and in the data science degree programme
- Elective in financial mathematics, computer science, sociology (advanced statistics (sic!)), . . .

Your background

- Who has a bachelor degree from an university different to LMU?
- Who has a bachelor degree from LMU but not in statistics?
- For whom is this semester the first one of her/his master's study?

A note on the degree of abstraction; fundamental thesis

Thesis (TA): Working with increasingly complex models, as important as it is and as many really significant successes are achieved in doing so, should not obscure the fact that many basic and fundamental issues in statistics and data science are still unresolved.

Core competencies of statisticians and data scientists

Learning objectives according to examination regulations

"Students should acquire an in-depth understanding of decision theory as a theory of rational decision-making under uncertainty and learn to critically classify common statistical methods from this general perspective."

Compare to above: Fundamental learning objectives of the course

- Learn to design and solve decision problems in various areas of application
- Understand 'common statistics and machine learning' better by adopting an outside perspective
  - \+ frame statistical and machine learning methods as decisions
  - \+ reflect on the notion of uncertainty

Implementation of learning objectives, examination modalities

- oral examination (ca. 25 min or for 6 ECTS 15 min).
- independent assessment of the 6 ECTS points part and the 3 ECTS points part.

1.1.6 Basic types of decision models

The decision problems considered in (the first part of) this module are restricted/characterized as follows:

- Homogeneous decisions maker

- Without loss of generality: single decision

- Here (almost) only decision problems with multiple states of nature are considered.

- There are different types of uncertainty! (TA. „**anchor of the whole decision theory**“)

  .

  – classical extremes of „generation of states of nature“.

  type I:) states of nature are the result of a perfect random mechanism with known probability distribution $\rightarrow$ „lotteries“ (<u>risk situation</u>) .

  versus

  type II:) state of nature created by actions of an hostile opponent (enemy)

– closely related to it, **[TA] but absolutely to distinguish from it:**

type I':) decision maker can specify exactly subjective probabilities for the states of nature, i.e., her/his own degrees of belief concerning the occurrence of the states. (Bayes(ian) decision situation)

type II':) uncertainty situation in the narrower/proper/strict sense , "ignorance", "Knightian Uncertainty", ((in German sometimes: Ungewissheitssituation))

- additional information
  We consider only two cases
  - without additional information
  - with additional information due to a random sample („statistical decision theory").

  We will see that actually the whole inductive statistics (hypothesis testing, parameter estimation) can be formally described as a decision problem under additional information.

- notion of rationality: „how to decide?"

1.2 The basic form of a data-free decision problem (No-data problem)

Def. 1.1. [Data-free decision problem]

A data-free decision problem (no-data problem) in utility (loss) form is a triple $(\mathbb{A}, \Theta, u(\cdot))$ or $(\mathbb{A}, \Theta, l(\cdot))$, consisting of

- of a nonempty set $\mathbb{A}$ („action set") ,
- of a nonempty set Θ („state set").
- and a mapping („utility function") (u $\hat{=}$ utility)

$$\begin{array}{rcccl} u & : & \mathbb{A} \times \Theta & \rightarrow & \mathbb{R} \\ & & (a, \vartheta) & \mapsto & u(a, \vartheta) \end{array} \tag{1.1}$$

or a mapping („loss function") (l $\hat{=}$ loss)

$$\begin{array}{rcccl} l & : & \mathbb{A} \times \Theta & \rightarrow & \mathbb{R} \\ & & (a, \vartheta) & \mapsto & l(a, \vartheta) \end{array} \tag{1.2}$$

Ex. 1.2. [The „omelet problem“ by Savage]

	ϑ_1	ϑ_2
a_1	12	-12
a_2	10	6
a_3	8	8
$(a_4$	4	4

Rem. 1.3. [Consequence function]

For some applications it is useful to add a step in between and – given $\mathbb{A}$ and Θ – first construct a consequence function

$$\begin{array}{cccc} c & : & \mathbb{A} \times \Theta & \to & \mathcal{C} \\ & & (a, \vartheta) & \mapsto & c(a, \vartheta) \end{array}$$

(with $\mathcal{C}$ being the set of potential consequences) to be considered and an utility evaluation on it.

$$\begin{array}{ccccc} u_{\mathcal{C}} & : & \mathcal{C} & \to & \mathbb{R} \\ & & c & \mapsto & u_{\mathcal{C}}(c) \end{array}$$

or a loss assessment

$$\begin{array}{ccccc} l_{\mathcal{C}} & : & \mathcal{C} & \to & \mathbb{R} \\ & & l & \mapsto & l_{\mathcal{C}}(c) \end{array}$$

to be determined.

$u(\cdot)$ and $l(\cdot)$ are then obtained by superposition of both functions as

$$u(a, \vartheta) = u_{\mathcal{C}}(c(a, \vartheta)) \qquad \text{resp.} \qquad l(a, \vartheta) = l_{\mathcal{C}}(c(a, \vartheta)) \,.$$

Rem. 1.4. [On utility and loss functions]

- In Definition 1.1, loss and utility functions (as functions of the form $\mathbb{A} \times \Theta \to \mathbb{R}$) formally have exactly the same structure. So it must always be made explicit which case applies. (utility: the more, the better; loss: the less, the better)
- This ambiguity is not least due to the fact that one should actually work with preference orders on the set of consequences as a fundamental entity. Then, utility theory teaches how, and under what conditions, one can derive from preference orders a cardinal (metric) utility. (More on this in part two.)
- The cardinality of the utility is not doubted (at first); thus, one can "calculate as usual" with utility units. In particular, equivalence of utility and loss views is usually assumed in what follows: by multiplying by (-1), one can then convert any utility function into a loss function reflecting exactly the same preference ordering. Therefore, in the following we will usually speak only of either utility or loss function.

Rem. 1.5. [Notation in the finite case; embedding in the $\mathbb{R}^m$]

In the case of a finite action set and a finite set of states of nature, the following notation is used:

$$\begin{aligned} \mathbb{A} &= \{a_1, ..., a_i, ..., a_n\} \\ \Theta &= \{\vartheta_1, ..., \vartheta_j, ..., \vartheta_m\} \end{aligned} \tag{1.3}$$

(so $|\mathbb{A}| = n$, $|\Theta| = m$; i running index in $\mathbb{A}$, j running index in Θ).

Utility functions, loss functions and consequence functions can then be represented as matrices $(u_{ij})_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$, $(l_{ij})_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$, $(c_{ij})_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$.

For example:

$$\begin{array}{c|cccccc}
 & \vartheta_1 & \vartheta_2 & .. & .. & \vartheta_m \\
\hline
a_1 & u_{11} & u_{12} & .. & .. & u_{1m} \\
a_2 & u_{21} & u_{22} & .. & .. & u_{2m} \\
.. & .. & .. & .. & .. & \\
.. & .. & .. & u_{ij} & .. & .. \\
.. & .. & .. & .. & .. & \\
a_n & u_{n1} & u_{n2} & .. & .. & u_{nm}
\end{array} \tag{1.4}$$

One then speaks of utility <u>table</u>, loss <u>table</u> or consequences <u>table</u>.

Embedding into $\mathbb{R}^m$:

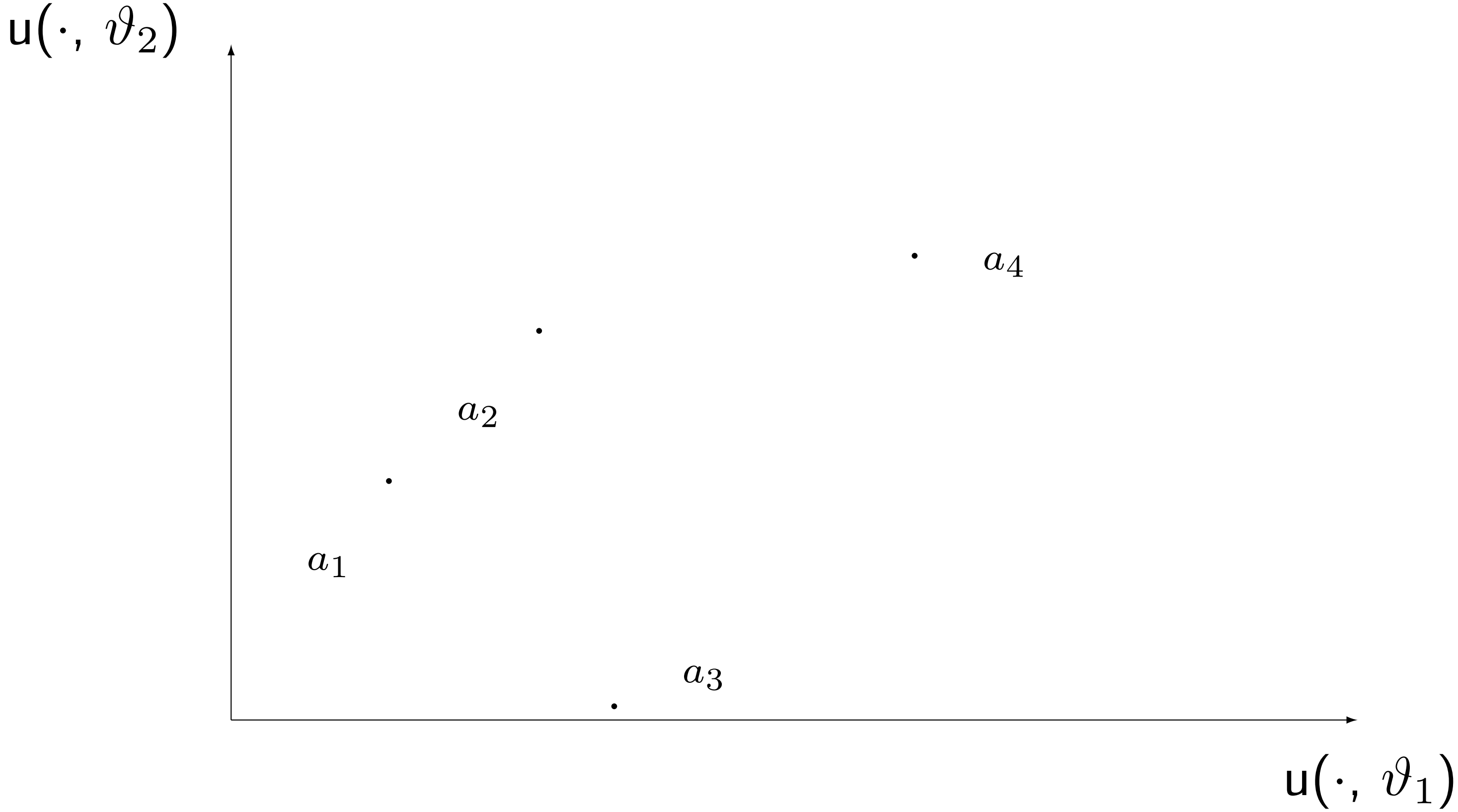
Each action $a_i \in \mathbb{A}$ can then be identified with the corresponding utility vector[7]

$$\vec{u}(a_i) = (u_{i1}, u_{i2}, ..., u_{ij}, ..., u_{im})^T \in \mathbb{R}^m \, . \tag{1.5}$$

If this view is taken, we speak of the direct utility or loss representation of the set of actions.

[7] In the subjective approach each vector is often called a *gamble*

Furthermore, if $m = 2$ or $m = 3$, a graphical representation is is possible:



Rem. 1.6. [Semantics of the data-free decision problem]

Abstracting from a concrete decision problem and summarizing it in the form $(\mathbb{A}, \Theta, u(\cdot))$ or $(\mathbb{A}, \Theta, l(\cdot))$ according to Definition 1.1, implicitly implies the following set of basic assumptions, which are to be critically questioned in the respective application.

a) $\mathbb{A}$ is known.

b) Θ is known (close-world-assumption).

c) The consequences are unique, i.e., the interaction of each $a \in \mathbb{A}$ and $\vartheta \in \Theta$ results in an unique certain consequence $c(a, \vartheta)$. Consequences may be probability distributions.

d) A uniquely determined[8] real-valued utility/loss function can be constructed fully reflecting the individual preferences of the decision maker.

- Cf. Remark 1.4: construction of cardinal (real-valued) utility functions from preferences satisfying certain regularity conditions $\rightarrow$ „utility theory"
- Throughout, it is assumed in particular that preferences allow a complete ordering and in this sense are/may be summarized as one-dimensional.
- Thus, there are no different non-mappable "utility dimensions" (e.g., a discrepancy between short- and long-term utility). Multidimensional utility assessments are the subject of so-called *multi-criteria / multi-target / multiple preferences decision theory*, whose embedding possibility will be briefly discussed in the examples section later on.

  There are also a number of approaches with fuzzy utility functions, which may be briefly addressed later in the in-depth section.

[8] Actually, uniqueness up to linear transformations would be enough.

- Utility is seen as an abstract quantity, in which, for instance, value dispositions are also integrated. Aspects like value orientation, fairness, reciprocity, etc. can play a role; they must be represented by the utility function.

- Caution: In the case of monetary consequences, utility is generally not identical with (or linear in) the quantity of money.
  Only if all the decision problems are far below the actual level of wealth, the utility function is (practically) linear in the monetary amounts. This is especially important when calculating expected utility/loss.

e) Actions and states are value-free. (Any valuations must be incorporated into the utility function).

f) The type of uncertainty is known (type I, type I', or type II, type II' uncertainty) (or later then generalizations).

g) No additional information is available (except the probability distribution on the states of nature for Type I or Type I' uncertainty). Formulate information acquisition via strategies $\rightarrow$ Statistical decision theory.

h) (Except for Type II uncertainty) The states of nature cannot be influenced by the choice of certain actions: „act(ion)-state independence“
If necessary, define "implication schemes" (defining states of natures is by no means always trivial!).

i) Actions are chosen in a one-time-process; there is no opportunity to correct the choice.

j) The decision situation is unique. There are no repetitions of the decision situation. (Formulate repeated decisions as one decision strategy!)