Learning from Weakly Structured Information

Foundations of Statistics and Their Applications

Thomas Augustin, Hannah Blocher, Cornelia Fütterer,
Christoph Jansen, Dominik Kreiß, Malte Nalenz,
Julian Rodemann, Georg Schollmeyer

Department of Statistics, LMU Munich

Munich, December 1st, 2021

Foundations of Statistics

and

Their Applications

Weakly Structured Information

- big data uncertainty, e.g., imprecise, error-prone, missing, conflicting, partial information
- non-standard data structures, e.g. partial ranking, multi attribute
- inexplicable, e.g. black box based

Preliminaries and Disclaimer

- Slides are available, just write to augustin@stat.uni-muenchen.de.
- many topics touched, invitation to co-operate and exchange
- Some statements will be formulated in a sharp, programmatic and sometimes even provocative manner in order to hopefully stimulate discussions. Such statements should by no means be misunderstood as a devaluation of the work of others.
- limited to statistics as learning from data, no data production

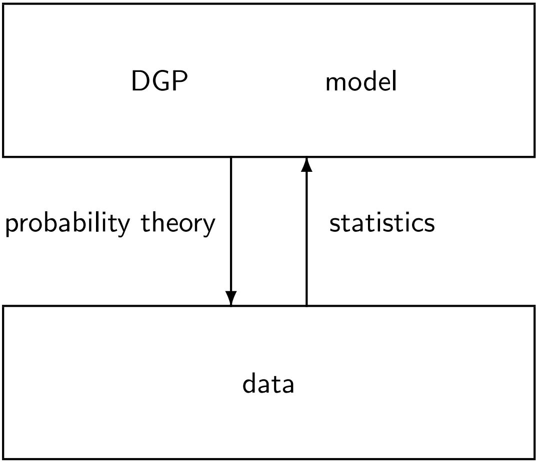
Table of contents

Table of contents

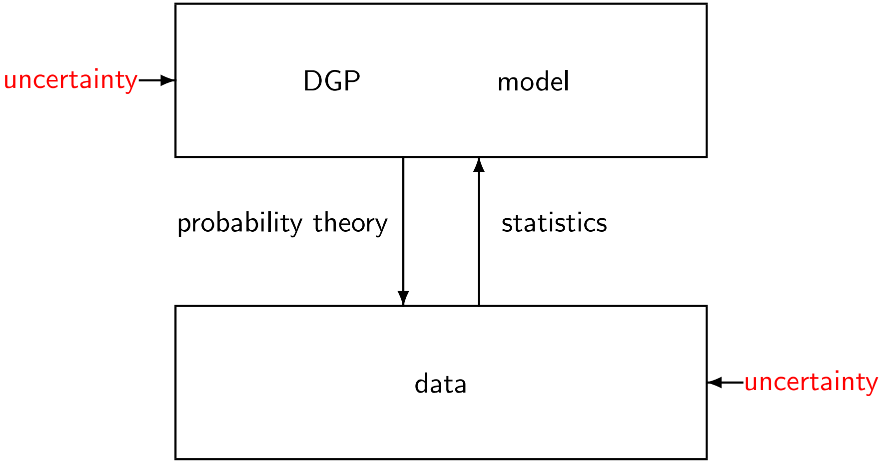
Be serious about the uncertainties involved!

- Statistics: learn from data and quantify uncertainty of conclusions by applying probability theory
- self-evident?
- questioned (outside statistics?)

Statistics as inverted probability theory



Statistics as inverted probability theory



Representing Uncertain Knowledge

Klir and Wierman (1998, Uncertainty-based Information, Physika, 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 became clear that there are several distinct types of uncertainty. That is, it was realized that* **uncertainty is a multidimensional concept** *[bold: TA]. [.... 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".*

Uncertainty in Machine Learning

"The notion of uncertainty is of major importance in machine learning and constitutes a key element of modern machine learning methodology. [...] 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."

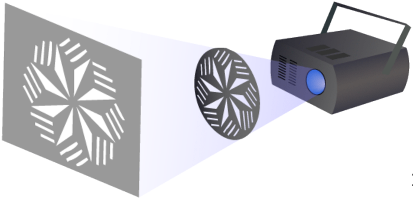
Destercke & Hüllermeier (2020, Web page of ECML/PKDD 2020 Tutorial and Workshop on Uncertainty in Machine Learning) [1]

[1] https://sites.google.com/view/wuml-2020/ [Nov 29th, 2021]

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" …

All points on the same light beam result in the same point on the projection surface.
Although actually different, they are indistinguishable on the low-dimensional projection surface.



[2]

[2] https://commons.wikimedia.org/wiki/File:Gobo_projected_illustration.png, Jeremy Kemp, free [Nov 23rd, 2021]

Complex Uncertainty

Uncertainty

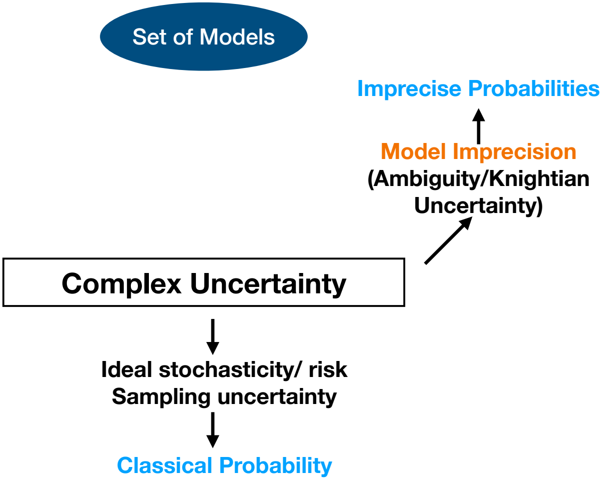
↓

Ideal stochasticity/ risk
Sampling uncertainty

↓

Classical Probability

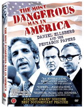
Complex Uncertainty



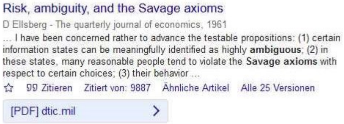
Knightian Uncertainty, Ambiguity: Ellsberg's "paradox"

- Are all uncertainties risks? Knight (1921)
- Ellsberg (1961, Quart.J.Econ): thought experiments among prominent statisticians and econometricians
- drawing balls from urns, partially under ambiguity
- majority violates the additivity axiom of classical probability and judges this as rational behavior

⟶ "Decision theory beyond expected utility"



[3]


Risk, ambiguity, and the Savage axioms
D Ellsberg - The quarterly journal of economics, 1961
… I have been concerned rather to advance the testable propositions: (1) certain information states can be meaningfully identified as highly **ambiguous**; (2) in these states, many reasonable people tend to violate the **Savage axioms** with respect to certain choices; (3) their behavior …
☆ 99 Zitieren Zitiert von: 9887 Ähnliche Artikel Alle 25 Versionen
[PDF] dtic.mil

Google scholar search: [Nov 24th, 2021]

[3] https://www.amazon.de/Most-Dangerous-Man-America-Ellsberg/dp/B00329PYGQ [Nov 24th, 2021]

Ambiguity in Decision Making – Empirical Evidence

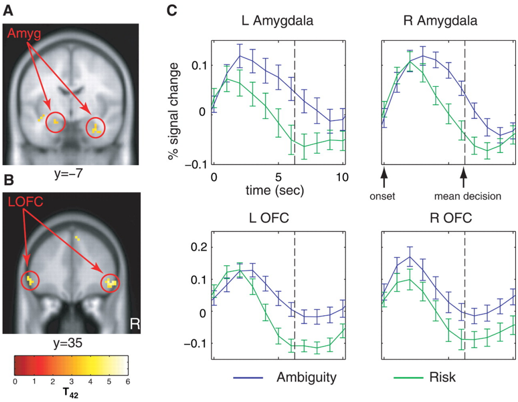
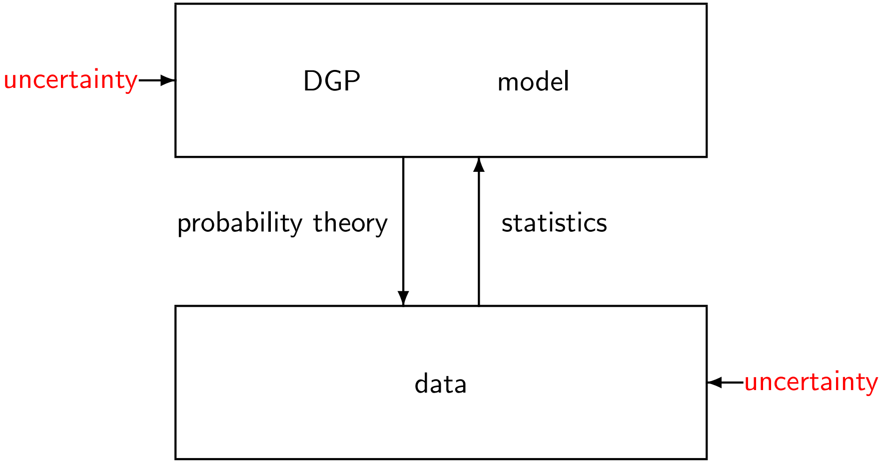


Figure: Hsu, Bhatt, Adolphs, Tranel & Camerer (2005, Science): Functional brain imaging corroborates Ellsberg's findings on the constitutive role of ambiguity in decision making.

Ambiguity and Non-additivity

- often non-divisible evidence: high certainty for compound event, but low certainty for each of the components
    - Bavarian or Austrian?
    - "Clearly a liver disease, but which one?"
    - compare later: yet undecided voters
- (relative certainty in $A \cup B$) >> (r. c. in A) + (r. c. in B)

Statistics as inverted probability theory (again)



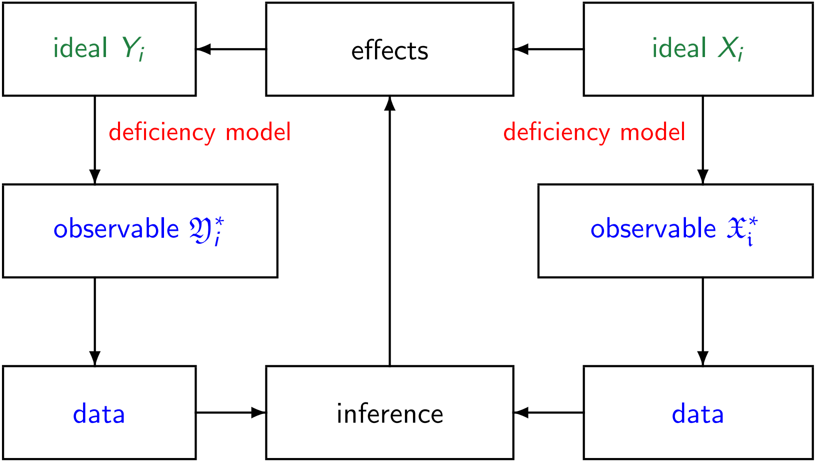
Different Kinds of Uncertainty

- Total Survey Error (TSE) concept: Groves & Lyberg (2010, Public Opinion Quarterly)
- learning from data as Chinese whispers game

Different Kinds of Uncertainty

- Total Survey Error (TSE) concept: Groves & Lyberg (2010, Public Opinion Quarterly)
- learning from data as Chinese whispers game
- sampling uncertainty
- “big data uncertainty”: does not vanish with increasing sample size

The two-layers perspective



Manski's Law of Decreasing Credibility

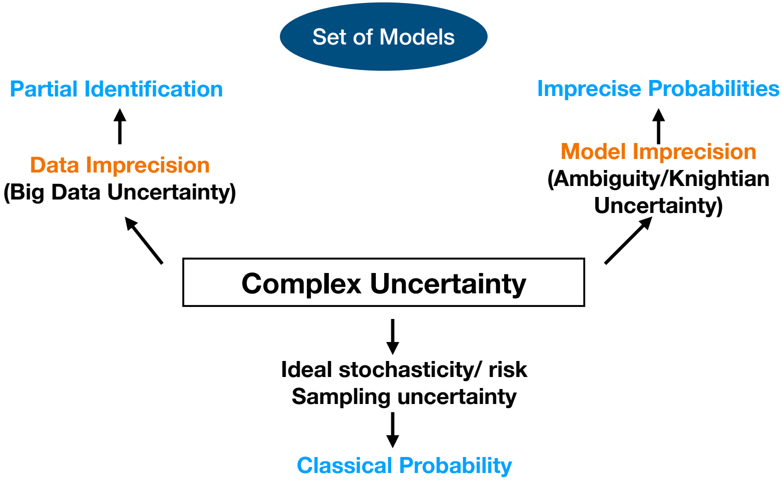
Credibility ?
"The credibility of inference decreases with the strength of the assumptions maintained." (Manski (2003, p. 1))

partial identification: Set of all models compatible with the data and tenable assumptions.

Charles Manski[4]

[4] http://faculty.wcas.northwestern.edu/~cfm754/; [Nov 22nd, 2021]

Complex Uncertainty



Imprecision in statistics

- hide/neglect imprecision!
- model imprecision away!

!! Take imprecision into account in a reliable way!

!! imprecision as a modelling tool

Quality of Information

"Let's Be Imprecise in Order to Be Precise (About What We Don't Know)"
Gong & Meng (2021, Statistical Science (Rejoinder), p. 210)

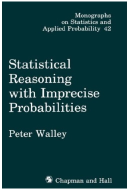
Xiao-Li Meng[5]

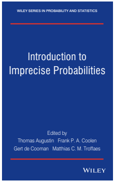
[5] https://statistics.fas.harvard.edu/people/xiao-li-meng; Nov 21st, 2021

Table of contents

Imprecise probability (IP) in a nutshell I

- unfortunate misnomer: actually IP claims to provide more precise (better) models
- uncertainty as a multidimensional concept
- theoretical foundations: Walley (1991, Chapman & Hall), Weichselberger (2001, Physika)
- intros: e.g. Augustin, Coolen, de Cooman, Troffaes (2014, eds, Wiley), ("ITIP"), Bradley (2019, Stanford Encyc. Phil)
- here: build simply on a very intuitive understanding
- from 1999 onwards biannual: ISIPTA: International Symposium on Imprecise Probabilities: Theories and Applications, www.sipta.org





Brief Digression: Kurt Weichselberger (1929-2016)

HiStaLMU Project; Augustin & Seising (2018, IJAR)

1968 Rector inaugural speech Berlin: interval-valued probabilities to resolve Fisher's fiducial argument

1974 Foundation of the Institute of Statistics and Philosophy of Science at LMU

1979 First Diploma Programme in Statistics

1991 With S. Pöhlmann: "A Methodology for Uncertainty in Knowledge-based Systems", Springer LN in AI

2001 "Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I", Physika

–2015 Work on the third volume on (interval-valued) logical probability

[6]

[6] Photo kindly provided by Weichselberger's family

Imprecise probability (IP) in a nutshell II

- *sets* of traditional probability models (credal sets) as the basic entity

  "$\Longleftrightarrow$"

  interval-valued probability $P(A) = [L(A), U(A)]$ of events A

- quality of information: "size of set", width of interval
  - traditional probability as the extreme case of perfect probabilistic information, real number, set with a single element
  - P(A)=[0;1] for all nontrivial events – set of all probability measures: complete ignorance, full ambiguity

- several updating/conditioning rules: here GBR only → conditioning element by element → *Robust Bayes*

- $L(\cdot)$ and $U(\cdot)$ are non-additive set-functions

IP in a nutshell III: Typical Applications

- direct modelling of partial knowledge: intervals of probabilities or expectations
- ordinal probabilities: $p(A) \leq p(B) \leq p(C)$...
- indivisible evidence ("Bavarian or Austrian?")
- handling of different granularities: unique extensions from any set-system to IP on the underlying measurable space
- unobserved heterogeneity

- decision making under ambiguity (Knightian uncertainty)
- reliability analysis in engineering

- robust statistics: "approximately true models"
- modelling prior knowledge/ignorance
- deficient, nay non-idealized, data (missingness, coarseness, measurement error, misclassification,...)

Table of contents

IP as a Superfluous Complication?

- Naturally, every abstraction yields some kind of imprecision.
- Is IP just a superfluous complication, invented by researchers who have not understood the nature of scientific abstraction?

The mantra of statistical modelling

Box & Draper (1987, Empirical Model Building and Response Surfaces, p. 424)

- "Essentially, all models are wrong,

The mantra of statistical modelling

Box & Draper (1987, Empirical Model Building and Response Surfaces, p. 424)

- "Essentially, all models are wrong,
- but some of them are useful",

The mantra of statistical modelling

Box & Draper (1987, Empirical Model Building and Response Surfaces, p. 424)

- “Essentially, all models are wrong,
- but some of them are useful”,
- and sometimes dangerous.

Assumptions may matter!

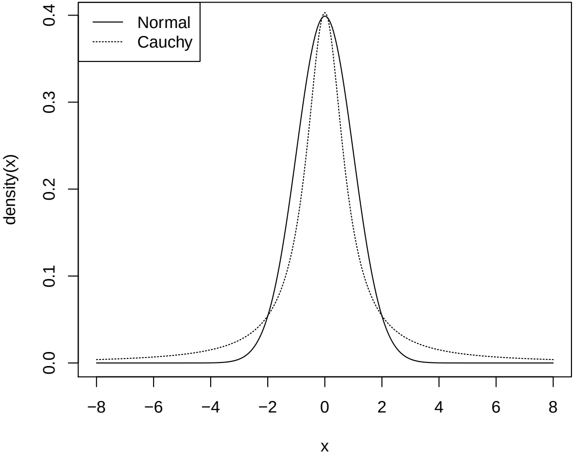


Figure: Densities of the Normal$(0, 1)$ and the Cauchy$(0, 0.79)$ distribution.

Assumptions may matter!

Consider sample mean $\overline{X}$.

- if $X_1, \ldots, X_n \sim N(\mu, 1)$ (normally distributed), then

$$\bar{X} \sim N(\mu, \frac{1}{n})$$

  Learning from the sample, with increasing sample size variance of $\overline{X}$ decreases.

- if $X_1, \ldots, X_n \sim \mathcal{C}(\mu, 1)$ (Cauchy-distributed), then

$$\overline{X} \sim C(\mu, 1)$$

  Distribution does not depend on n, no learning via sample mean possible

Assumptions may matter! Robustness

- many optimal procedures show very bad properties under minimal deviations from the ideal model
- instead of $f(x\|\vartheta)$: model "approximately $f(x\|\vartheta)$", i.e. consider all distributions "close to $f(x\|\vartheta)$"
  $\rightarrow$ neighbourhood models

Neighborhood-Based Ideas in Machine Learning

- classification trees: neighbourhood models help to avoid overfitting: lower entropy (Abellan & Moral (2003, IJUFKBS), Strobl (2005, ISIPTA))
- extended in Fink (2018, Diss LMU), Fink (2018, Imptree:CRAN)

- "cultivated random forests" for robust decision tree learning (Nalenz & Augustin (2021a, preprint)):
  - replacing internal nodes with two types of ensemble modules that pool together a set of decisions into a soft decision
    - option modules consisting of all reasonable variable choices at each step of the induction process
    - robust split modules including all elements of a neighbourhood of an optimal split-point as reasonable alternative split-points
    - $\longrightarrow$ ensemble centred around a single tree structure
  - abstain from predictions when the uncertainty is too high
  - better interpretability without loosing much predictive power

Neighborhood-Based Ideas in Machine Learning

connection between (imprecise) neighborhood models and ensemble learning

- Trained Tree Ensembles (Random Forests or Boosting) can often be approximated by imprecise decision rule and tree models.
- Often relatively simple imprecise model sufficient to capture the most central pattern found by the full ensemble: "compressed rule ensembles" (Nalenz & Augustin (2021b, under review))
- Carries over the low prediction variance and smooth decision boundaries, while being much more interpretable.

A Quick Look at Bayesian Methods

- Dominating branch based on Walley (1991, Chapman & Hall), Walley (1996, J R STAT SOC B)
- Relationship to Robust Bayesian Analysis

- Now distinction between variability and indeterminism possible
- Near ignorance priors
- Proper handling of prior-data conflict: Augustin, Walter & Coolen (2014, ITIP)

- Generalizations of Bayes factors: Ebner, Schwaferts &, Augustin (2019, ISIPTA), Schwaferts (2021, DissLMU)
- Extended (imprecise) empirical Bayes methods: Augustin & Schollmeyer (2021, Statistical Science)

Near-Ignorance Priors

- Any single prior distribution carries probabilistic information
- Use generalized priors such that “regular” events in the parameter space have probability $[0;1]$.
- “near-ignorance priors”
- IDM, multinomial inference: Walley (1996, JRSSB)
- exponential families: Banavoli & Zaffalon (2014, Statistics)
- Gaussian processes: Mangili (2015, ISIPTA; 2017, IJAR)

Prior-mean-RObust Bayesian Optimization (PROBO)

- Bayesian optimization with Gaussian process priors
- prior mean parameters have the highest influence on convergence among all prior components
- accounting for GP imprecision via a prior near-ignorance model
- generalized lower confidence bound: imprecision-adjusted acquisition function
- Rodemann (2021, MSc.LMU), Rodemann & Augustin (TR, under review)
- promising extensions

Statistics as inverted probability theory

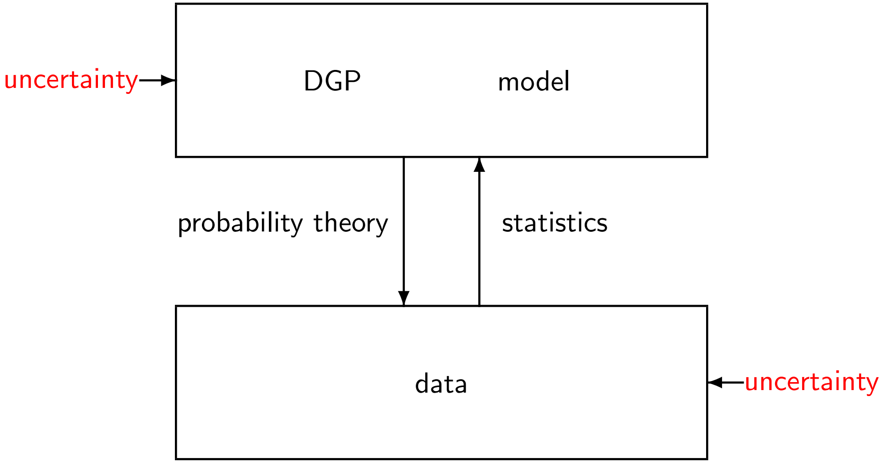


Table of contents

(Epistemic) Data Imprecision

Imprecise observations of something precise

- missing data
  - non-response
  - missingness by treatment design
  - statistical matching, e.g. imprecise imputation: Endres, Fink & Augustin (2020, J. Off. Stat.)
- measurement error
  - beyond classical m.e. modelling: generalized score functions
  - data protection: Fink & Augustin (2017, ISIPTA)
- coarse data
  - data merging with partially overlapping categories
  - secondary data analysis
  - refined responses of primary refusals, typically coarsening/missing not at random: Plass, Cattaneo, Augustin, Schollmeyer & Heumann (2019, Int. Stat. Rev)
  - forecasts derived from set-valued observations (see below)

Election Forecasting with Yet Undecided Voters

- Project with the polling institute Civey, together with Dominik Kreiss
- pre-election polling data for the 2021 German federal election
- new questionnaire design: explicit collection of the consideration sets (Oscarsson & Rosema (2019, Elect.Stud)) of yet undecided voters ("Between which parties are you undecided?")
- valuable information far beyond "don't know":
  - typically indecisiveness only between (very) few parties
  - precise vote for all coalitions containing parties in the voter's consideration set
- Kreiss & Augustin (2021, ArXiv), Kreiss & Augustin (2020, SUM), Kreiss, Nalenz & Augustin (2020, ECML/PKDD-WUML)
- builds on first ideas in Plass, Augustin, Cattaneo, Schollmeyer (2015b, ISIPTA) and Plass (2018, DissLMU)
- Couso & Dubois (2014, IJAR), Couso, Dubois & Sánchez (2014, Springer)

- S set of parties standing for election
- two levels of (generic) response variables
  - $\mathfrak{Y}$: consideration set, set $\mathfrak{l}$ of preferred parties, observable
  - Y: final choice, party $\ell \in \mathfrak{l}$, not observable
  - covariates X, realizations x

- S set of parties standing for election
- two levels of (generic) response variables
  - $\mathfrak{Y}$: consideration set, set $\mathfrak{l}$ of preferred parties, observable
  - Y: final choice, party $\ell \in \mathfrak{l}$, not observable
  - covariates X, realizations x
- point estimator for percentage of votes a set A of parties achieves

$$\widehat{p}(Y \in A) = \sum_{\substack{(\ell, \mathfrak{l}, x) \in \\ A \times \mathcal{P}(S) \times \mathcal{X}}} \underbrace{p(Y = \ell \mid \mathfrak{Y} = \mathfrak{l}, X = x)}_{\text{latent transition model}} \cdot \underbrace{\widehat{p}(\mathfrak{Y} = \mathfrak{l} \mid X = x)}_{\text{from data}} \cdot \underbrace{\widehat{p}(X = x)}_{\text{from data, sampling weights}}$$

- S set of parties standing for election
- two levels of (generic) response variables
  - $\mathfrak{Y}$: consideration set, set $\mathfrak{l}$ of preferred parties, observable
  - Y: final choice, party $\ell \in \mathfrak{l}$, not observable
  - covariates X, realizations x
- point estimator for percentage of votes a set A of parties achieves

$$\widehat{p}(Y \in A) = \sum_{\substack{(\ell, \mathfrak{l}, x) \in \\ A \times \mathcal{P}(S) \times \mathcal{X}}} \underbrace{p(Y = \ell \mid \mathfrak{Y} = \mathfrak{l}, X = x)}_{\text{latent transition model}} \cdot \underbrace{\widehat{p}(\mathfrak{Y} = \mathfrak{l} \mid X = x)}_{\text{from data}} \cdot \underbrace{\widehat{p}(X = x)}_{\text{from data, sampling weights}}$$

- results depend strongly on the unknown transition model

$$\big(p(Y = \ell \mid \mathfrak{Y} = \mathfrak{l}, X = x)\big)_{\ell \in S, \mathfrak{l} \in \mathcal{P}(S)},$$

  which is, however, unidentifiable without further assumptions.
- For every $\mathfrak{l}$ and x the transition model specifies a probability distribution $p_{(\mathfrak{l},x)}$ on $(\mathfrak{l}, \mathcal{P}(\mathfrak{l}))$.

"Modelling"

- For the moment let's argue without the covariates: $p_{(\mathfrak{l},x)} \hookrightarrow p_{(\mathfrak{l})}$
- Thinking of a concrete example may be helpful; consider, e.g., $\mathfrak{l} = \{SPD, Left, Green\}$.
- See above: results depend strongly on the unknown transition model.
- Therefore, think of the forecast as a function of the transition model underlying, i.e. consider

$$\widehat{p}(Y \in A)\left[(p_{\mathfrak{l}})_{(\mathfrak{l} \in \mathcal{P}(S))}.\right]$$

"Precise modelling"

Potential ideas to specify the latent transition model precisely:

- prophetic: give exact numbers for $(p_{\mathfrak{l}})_{(\mathfrak{l} \in \mathcal{P}(S))}$
- transfer knowledge from polls of older elections

- uniform (max ent)

$$p(Y = \ell \mid \mathfrak{Y} = \mathfrak{l}) := \frac{1}{|\mathfrak{l}|}$$

- homogeneous with respect to the decided

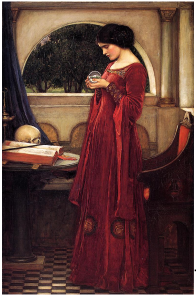
$$p(Y = \ell \mid \mathfrak{Y} = \mathfrak{l}) := \frac{p(\mathfrak{Y} = \{\ell\})}{\sum_{\ell' \in \mathfrak{l}} p(\mathfrak{Y} = \{\ell'\})}$$

- noninformativeness of coarsening (CAR: coarsening at random) (indirect)

$$\forall \mathfrak{l} \in \mathcal{P}(S) : \forall \ell_1, \ell_2 \in \mathfrak{l} : \frac{p(Y = \ell_1 | \mathfrak{Y} = \mathfrak{l})}{p(Y = \ell_2 | \mathfrak{Y} = \mathfrak{l})} = \frac{p(Y = \ell_1)}{p(Y = \ell_2)}$$

Justification of these Assumptions

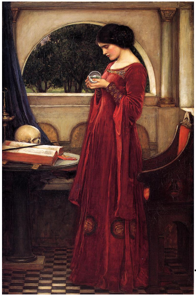
Justification of these Assumptions

[7]

[7] John William Waterhouse: The Crystal Ball (1902)
http://www.wikiart.org/en/john-william-waterhouse/the-crystal-ball-1902,Gemeinfrei, [Nov 27th, 2020]

Justification of these Assumptions

- Assumptions specifying the transition model have to be well-grounded in good subject-matter arguments, derived from the domain knowledge.
- All the assumptions just stated (and many more) are indistinguishable by relying on the data only.
- There CANNOT be any meaningful statistical test to support/reject any of these assumptions.


[7]

- Relying on such assumptions just for the sake of receiving (seemingly) precise solutions is questionable.

[7] John William Waterhouse: The Crystal Ball (1902)
http://www.wikiart.org/en/john-william-waterhouse/the-crystal-ball-1902,Gemeinfrei, [Nov 27th, 2020]

Enveloping all Possible Specifications of the Transition Model

- What do we know "for sure"?
- Consider all possible specifications for

$$\big(p(Y = \ell \mid \mathfrak{Y} = \mathfrak{l}, X = x)\big)_{\ell \in S, \mathfrak{l} \in \mathcal{P}(S)}$$

- That is, consider for each $\mathfrak{l}$, the set of all probabilities on $(\mathfrak{l}, \mathcal{P}(\mathfrak{l}))$.

Enveloping all Possible Specifications of the Transition Model

- What do we know "for sure"?
- Consider all possible specifications for

$$\big(p(Y = \ell \mid \mathfrak{Y} = \mathfrak{l}, X = x)\big)_{\ell \in S, \mathfrak{l} \in \mathcal{P}(S)}$$

- That is, consider for each $\mathfrak{l}$, the set of all probabilities on $(\mathfrak{l}, \mathcal{P}(\mathfrak{l}))$.
- By assuming error-freeness of coarsening

$$p(Y \in A) \mid \mathfrak{Y} = \mathfrak{l}, X = x)\big) = \begin{cases} 0 & \mathfrak{l} \subseteq A^C \\ 1 \quad \text{if} & \mathfrak{l} \subseteq A \\ [0;1] & \mathfrak{l} \cap A \neq \varnothing \wedge \mathfrak{l} \cap A^C \neq \varnothing \end{cases}$$

What has the Theory of Partial Identification to Offer here?

- Enveloping all scenarios: worst- and best case estimates
- When weak, but well- supported information is available, utilize it to increase precision!

Enveloping all Possible Specifications of the Transition Model (continued)

- $$p(Y \in A) \mid \mathfrak{Y} = \mathfrak{l}, X = x)) = \begin{cases} 0 & \mathfrak{l} \subseteq A^C \\ 1 \quad \text{if} & \mathfrak{l} \subseteq A \\ [0;1] & \mathfrak{l} \cap A \neq \varnothing \wedge \mathfrak{l} \cap A^C \neq \varnothing \end{cases}$$
- move probability mass around where not fixed
- **lower bound ("guarantee"):**

$$\underline{P}(Y \in A) = \sum_{\mathfrak{l} \subseteq A} p(\mathfrak{Y} = \mathfrak{l})$$

$$\underline{P}(SPD, Gr, FDP) = p(SPD) + p(Gr) + p(FDP) + p(SPD, Gr) + p(SPD, FDP) + p(Gr, FDP) + p(SPD, Gr, FDP)$$

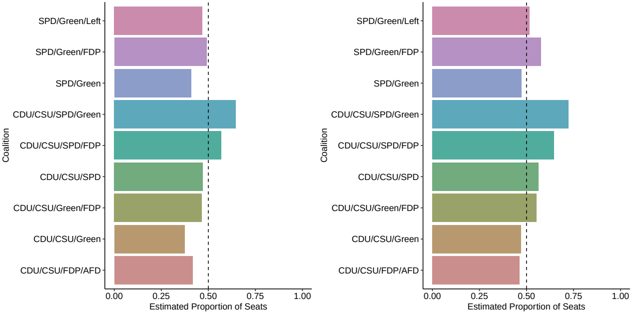
- **upper bound ("potential"):**

$$\overline{P}(Y \in A) = \sum_{\mathfrak{l} \cap A \neq \varnothing} p(\mathfrak{Y} = \mathfrak{l}) .$$

- Construction goes back to Dempster (1967, Ann.Math.Stat) and Shafer (1976, Princeton UP) in the context of fiducial inference and modelling uncertain knowledge, respectively.

Dempster Bounds

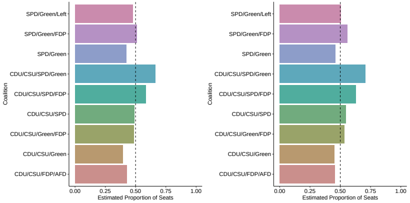
- most cautious analysis:[8] appropriate communication of full uncertainty about transitions
- Considerable increase in precision when coalitions are considered! For instance, being undecided between SPD and Green is a precise vote for any coalition containing these parties!



[8] Figure is taken from Kreiss & Augustin (2021, Arxiv; p. 10)

Exploit Weak Knowledge about Transition Probabilities

- weigh precision and credibility
- communication of the uncertainty present
- work with plausible weak assumptions not exploitable in traditional statistics
- expert opinions, like: "the undecided between Party I and Party II tend as least as much to Party I than to Party II"
- weaken "precise conditions" by considering neighborhood models
- generalized uniform probability: between 50-c% and 50+c% for all parties[9]
- easy technical handling via linear optimization



[9] Figure for $c = 30$ is taken from Kreiss & Augustin (2021, Arxiv; p. 10)

Reliable inference instead of overprecision!!

Consequences to be drawn from the Law of Decreasing Credibility:

- adding untenable assumptions to produce precise solution may destroy credibility of statistical analysis, and therefore its relevance for the subject matter questions.
- make *realistic* assumptions and consider the *set* of *all* models that are compatible with the data and these assumptions (and then add successively additional assumptions, if desirable)
- the results may be imprecise, but are more reliable
- the extent of imprecision is related to data quality!
- as a welcome by-product: clarification of the implication of certain assumptions
- often still sufficient to answer subjective matter question
- "weak information" may be powerful in refining results

Table of contents

Machine Learning

- neighborhood models for stability and interpretability (see above)
- potential of near-ignorance models in Bayes-based methods

- ensembles of dimension reduction-based clustering methods with applications to single-cell data: Fuetterer, Augustin & Fuchs (2020, ADAC), Fuetterer & Augustin (2021, MABM@IEEE BIBM)
- Discriminative power Lasso: Fuetterer, Nalenz & Augustin (2021, TR)

- power of credal classification (set-valued, predict several categories for some hard to classify instances)

- study effects of sample design on ML-methods
- project with `destatis`

Decision Theory under with Imprecise Probabilities

- Think more in terms of actions and their consequences!
- Intensive interest in economics and mathematical philosophy, however mostly data-free problems: Ellsberg paradoxes, bounded rationality
- Optimization theoretic framework: Jansen, Schollmeyer & Augustin (2017, ESQARU)
- Degrees of E-admissibility: Jansen, Schollmeyer & Augustin (2022, Seidenfeld Festschrift)
- Guiding superstructure upon Bayesian inference in psychology Schwaferts (2021, Diss LMU), e.g. Schwaferts & Augustin (2019/21, ISIPTA)
- great opportunities for data-based decision making
- likelihood-based decision theory, e.g. Cattaneo (2013, Elec.J.Stat)
- imprecise sampling models
- explicit incorporation of sequentiality $\hookrightarrow$? robust reinforcement learning

Information Efficient Decision Making with Complexly Structured Preferences

- Beyond real-valued, externally given utility functions
- *Preference systems* $\mathcal{A} = [A, R_1, R_2]$ with $R_1 \subseteq A \times A$ a pre-order on A and $R_2 \subseteq R_1 \times R_1$ a pre-order on R_1.
- Interpretation of preference systems:
  - $(a, b) \in R_1$: "a at least as desirable as b" **(ordinal part)**
  - $((a, b), (c, d)) \in R_2$: "exchanging b by a is at least as desirable as exchanging d by c" **(cardinal part)**
- Jansen, Blocher, Augustin & Schollmeyer (2021, IJAR min rev)
  - (I) Methods for eliciting $\mathcal{A}$ by only asking *ranking questions about* R_1.
  - (II) *Data-driven* guidance of elicitation with *previous user experience*.
  - (III) Utilizing elicitation methods for *information efficient decision making* between acts $X : S \to A$ taking values in A.

Partially Ordered Data and Data Depth

- Partially ordered data, ontic interpretation of order
- rather than on distances better to rely on *data depth*, measuring centrality and outlyingness (e.g. Serfling & Zuo (2000; Ann.Statist.))
- transform the existing properties for spatial depth functions into the more general framework of formal concept analysis (Blocher & Schollmeyer (2022), see below)
- computational aspects (Blocher (2019, MSc.LMU))
  - computing depth functions,
  - sampling from a probability given by such a depth function
  - sampling a set given by a formal context or closure system

Formal Concept Analysis

- Formal concept analysis (Ganter & Wille (2012, Springer)) in statistics: data set consisting of observations and their attribute values.
- produce a hierarchy of concepts by representing the data set as a family of subsets of the observations such that elements sharing the same attributes are grouped together.
- embedding into the frameworks of complete lattices and closure systems.
- powerful representation of non-standard data types, for instance
  - partial ranking data (e.g. preference systems) (see above)
  - observations with multi-type attributes such as numerical and spatial combinations
- VC-dimension and regularization in formal concept analysis (Schollmeyer (2017, DissLMU))
  - descriptive analysis
  - test statistics based on suprema over families of sets

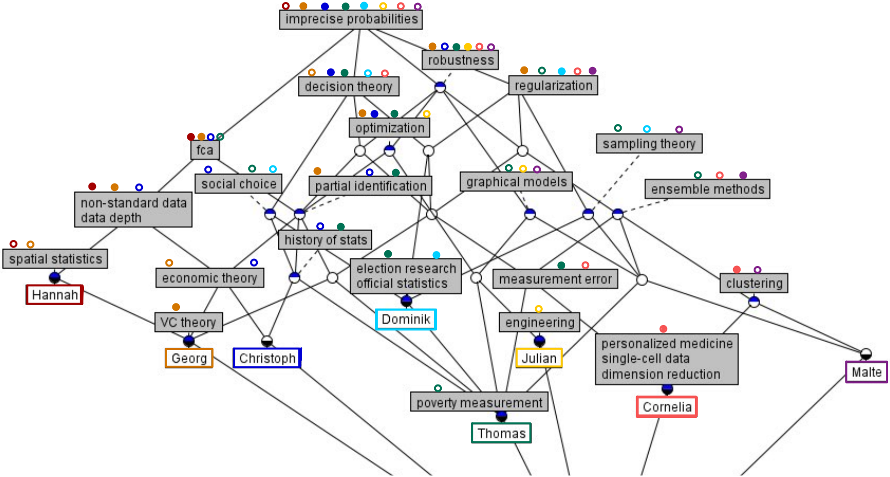

imprecise probabilities
robustness
decision theory
regularization
optimization
fca
sampling theory
social choice
partial identification
graphical models
ensemble methods
non-standard data
data depth
history of stats
spatial statistics
economic theory
election research
official statistics
measurement error
clustering
Hannah
VC theory
Dominik
engineering
personalized medicine
single-cell data
dimension reduction
Georg
Christoph
Julian
Malte
poverty measurement
Cornelia
Thomas

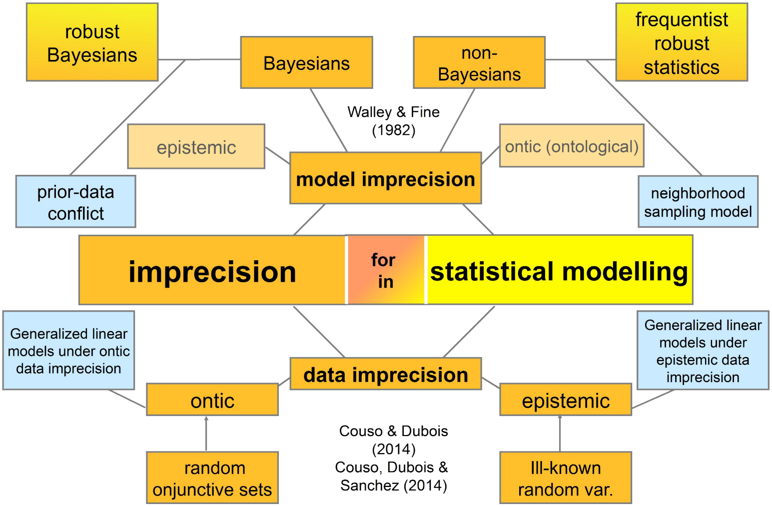
Table of contents

Concluding Remarks

Foundations of Statistics
and
Their Applications

- learning from weakly structured information
- generalized view of complex uncertainty
  - defensive handling of imprecision: robustification
  - active handling of imprecision opens new avenues
- power of decision theoretic and order-theoretic concepts

- many cross-sectoral topics and potential links
- looking forward to discussions and co-operations


robust Bayesians
Bayesians
non-Bayesians
frequentist robust statistics
Walley & Fine (1982)
epistemic
ontic (ontological)
prior-data conflict
model imprecision
neighborhood sampling model
imprecision
for
in
statistical modelling
Generalized linear models under ontic data imprecision
Generalized linear models under epistemic data imprecision
data imprecision
ontic
epistemic
Couso & Dubois (2014)
Couso, Dubois & Sanchez (2014)
random onjunctive sets
Ill-known random var.