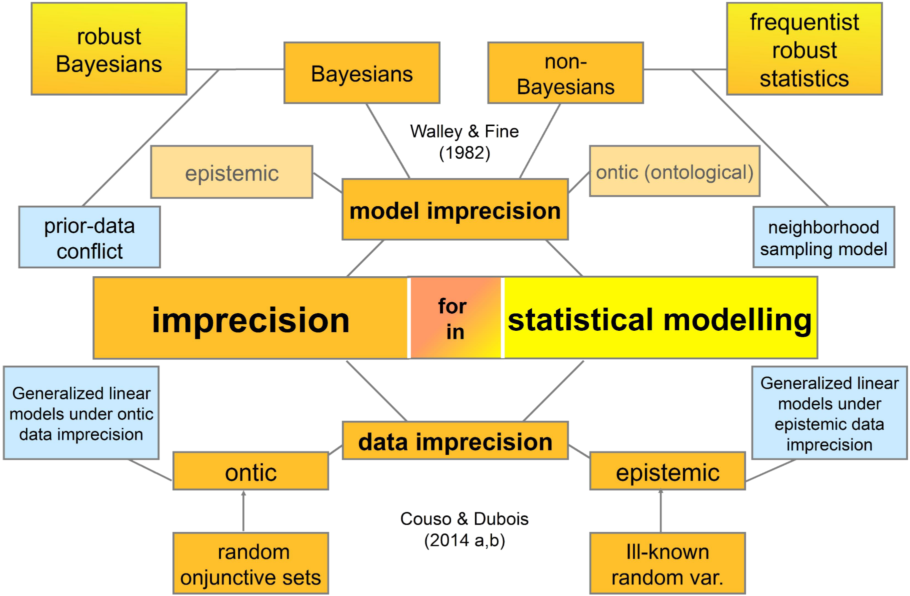

robust Bayesians
Bayesians
non-Bayesians
frequentist robust statistics
Walley & Fine (1982)
epistemic
ontic (ontological)
model imprecision
prior-data conflict
neighborhood sampling model
imprecision
for
in
statistical modelling
Generalized linear models under ontic data imprecision
Generalized linear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

# Imprecise Probability in Statistical Modelling: A Critical Review

T. Augustin M. Cattaneo P. Fink J. Plaß G. Schollmeyer
G. Walter A. Wiencierz F. Coolen U. Pötter M. Seitz

University of Munich (LMU)

Toulouse
May 27th, 2015

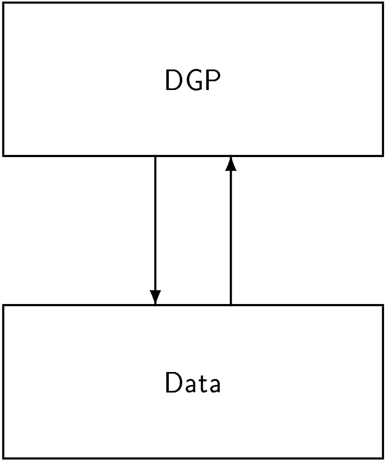
# Imprecision in Statistics

– hide/neglect imprecision!
– model imprecision away!

!! take imprecision into account in a reliable way!

!! imprecision as a modelling tool

# Imprecision



# Two kinds of imprecision

- **data imprecision:** imprecise observations, data are subsets of the intended sample space
  - * imprecisely observed precise observations $\rightarrow$ epistemic
  - * precisely observed imprecise observations $\overset{\approx}{\rightarrow}$
- **model imprecision:** imprecise probability models

$$P(Data||Parameter),$$

maybe also $P(Parameter)$

set-valued approaches: take **sets** of values/probability distributions as the basic entity

Couso & Dubois (2014, IJAR), Couso, Dubois & Sánchez (2014, Springer)

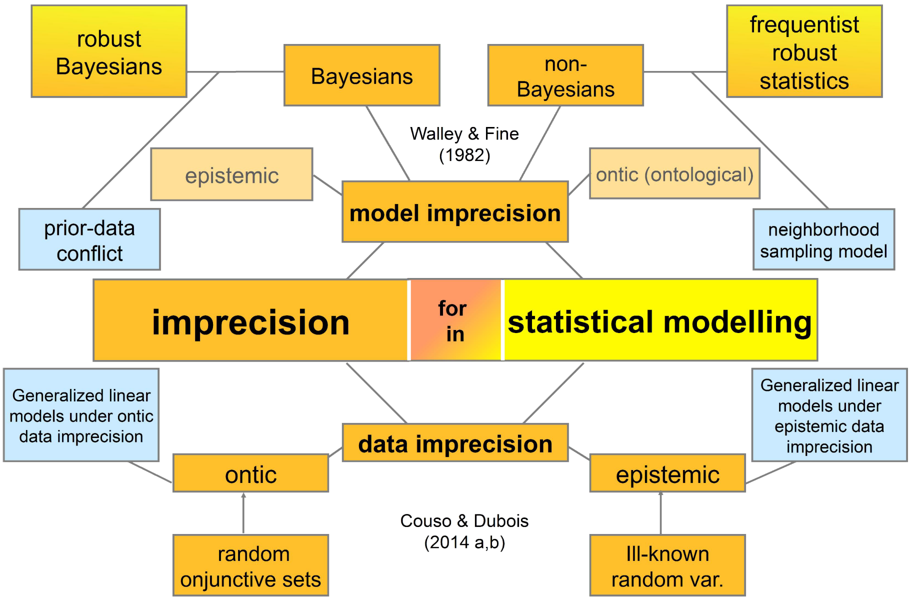
# On the power of IP in statistical modelling

- defensive point-of-view
  - IP protects against the potential disastrous behaviour of standard procedures under violated assumptions $\rightarrow$ robustness in:
  - frequentist and
  - Bayesian settings

# On the power of IP in statistical modelling

- offensive point of view.
  IP is a most powerful methodology, allowing for
    - separation of variability (variance) from indeterminism
    - active modelling of ignorance
    - active modelling of conflicting/surprising information
    - active use of weak knowledge that can not be used in the traditional setting

# Table of Contents


robust Bayesians
Bayesians
non-Bayesians
frequentist robust statistics
Walley & Fine (1982)
epistemic
ontic (ontological)
model imprecision
prior-data conflict
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 a,b)
random onjunctive sets
Ill-known random var.

# Table of Contents

# 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! Robustness

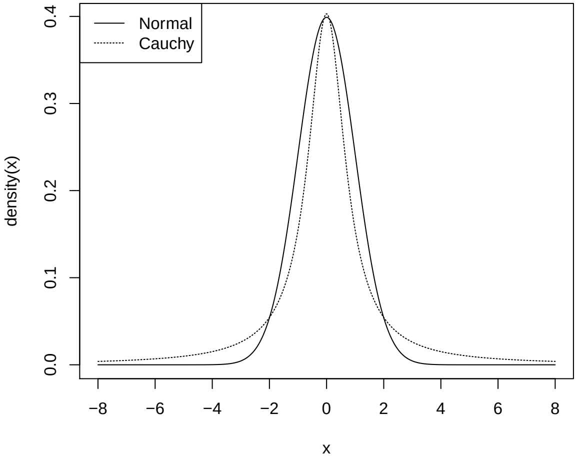



Figure: Densities of the Normal$(0, 1)$ and the Cauchy$(0, 0.79)$ distribution.

# 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

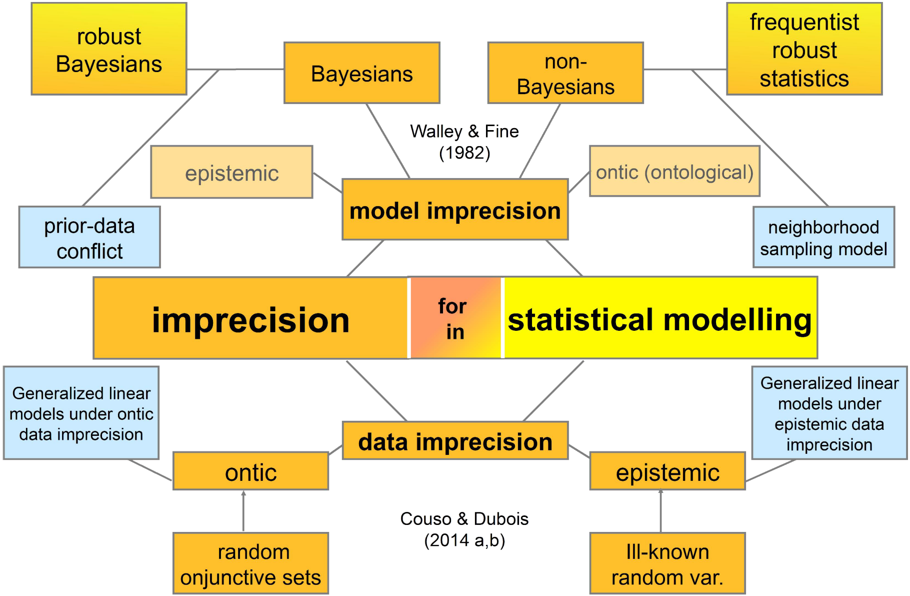
# Assumptions may matter!

- 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 distribution "close to Nähe von $f(x||\vartheta)$ " do

# Huber & Strassen Approach

Surveyed in Augustin, Walter & Coolen (2014, Intro IP, Wiley)

- Applicable to most neighborhood models of precise probabilities
- Extension to neighborhood models of many IP models
- Construction procedures
- Going beyond two-monotonicity
    - parametrically constructed models
    - locally least favorable pairs


robust Bayesians
Bayesians
non-Bayesians
frequentist robust statistics
Walley & Fine (1982)
epistemic
ontic (ontological)
model imprecision
prior-data conflict
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 a,b)
random onjunctive sets
Ill-known random var.

# Table of Contents

# Bayesian inference with sets of priors I: Ignorance

- So-called 'noninformative priors' do contain information
- consider set of all (non-degenerated distributions) instead, e.g. Walley, 1996, JRSSB, Benavoli & Zaffalon, 2012, JSPI
  end $\rightarrow$ proper modelling of prior data-conflict

# Bayesian inference with sets of priors II: Prior-data conflict

- Bayesian models are understood to express prior knowledge (or to "borrow strength")
- What happens when this prior konwledge is wrong?
- Example: $X_1, \ldots, X_n$ i.i.d data, $X_i \sim \mathcal{N}(\mu, \sigma_0^2)$
  conjugated prior: $\mu \sim \mathcal{N}(\nu, \varrho^2)$ then

$$\nu' = \frac{\bar{x}\rho^2 + \nu \cdot \frac{\sigma^2}{n}}{\rho^2 + \frac{\sigma^2}{n}}$$

$$\rho^{2\prime} = \frac{\rho^2 \cdot \frac{\sigma^2}{n}}{\rho^2 + \frac{\sigma^2}{n}}$$

# Bayesian inference with sets of priors II: Prior-data conflict

- Let, for sake of simplicity, $\varrho^2 = \frac{\sigma^2}{n}$, then

$$\hat{\mu} = \nu' = \frac{\bar{x} + \nu}{2}$$

  and

$$\varrho^{2'} = \frac{\varrho^4}{2\varrho^2} = \frac{\varrho^2}{2}.$$

- Then

$$\bar{x} = 0.9 \text{ and } \nu = 1.1$$

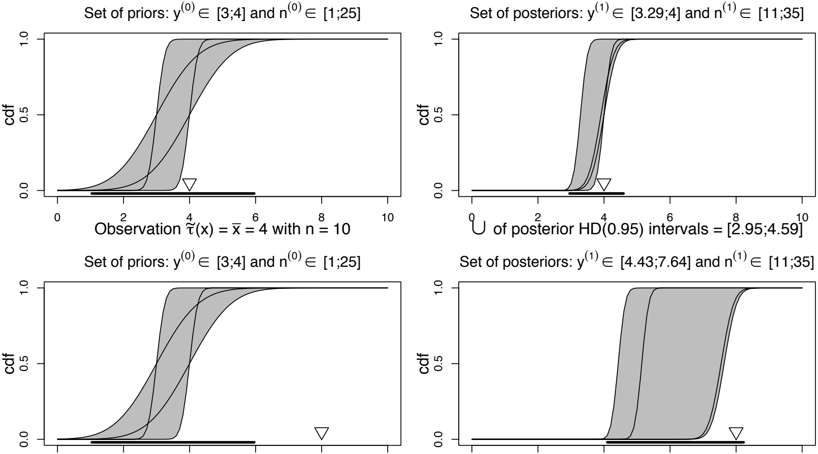
  and

$$\bar{x} = -100 \text{ and } \nu = 102$$

  lead to the same distribution (equal mean and variance )
- General effect for canonical exponential families
- Much more intuitive behaviour when prior parameters are imprecise, e.g. are interval-valued

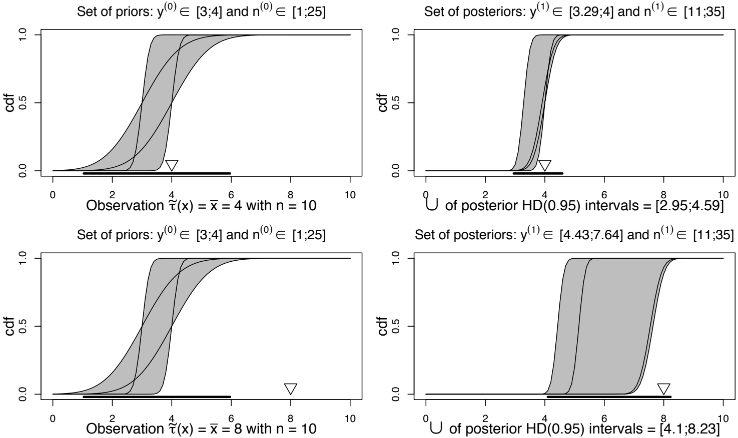
# Bayesian inference with sets of priors II: Prior-data conflict

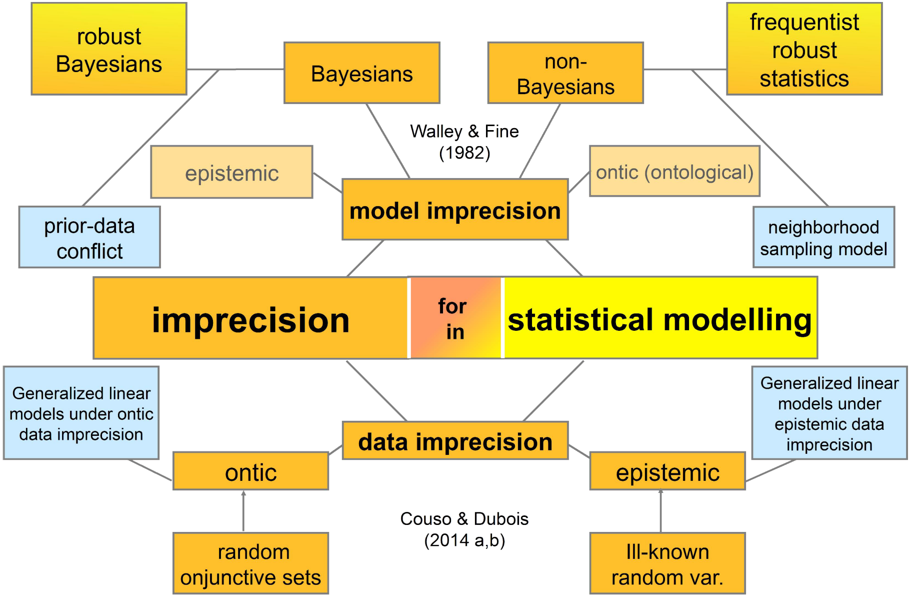
Source: Walter & Augustin (2009, JStTheorPract, p. 268)



# Bayesian inference with sets of priors II: Prior-data conflict

Source: Walter & Augustin (2009, JStTheorPract, p. 268)




robust Bayesians
Bayesians
non-Bayesians
frequentist robust statistics
Walley & Fine (1982)
epistemic
ontic (ontological)
model imprecision
prior-data conflict
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 a,b)
random onjunctive sets
Ill-known random var.

# Table of Contents

# Ontic imprecision: example

Plass, Fink, Schöning & Augustin (2015, ISIPTA)

- Pre-election study (GLES 2013: German Longitudinal Election Study)
- A considerable amount of voters is still undecided, but mainly only between two or three parties
- These voters constitute different subgroups of there own with specific characteristics (, which have to be neglected in the traditional analysis)
- Here NO forecast aimed at, instead analysis of individual preferences as they are in the moment

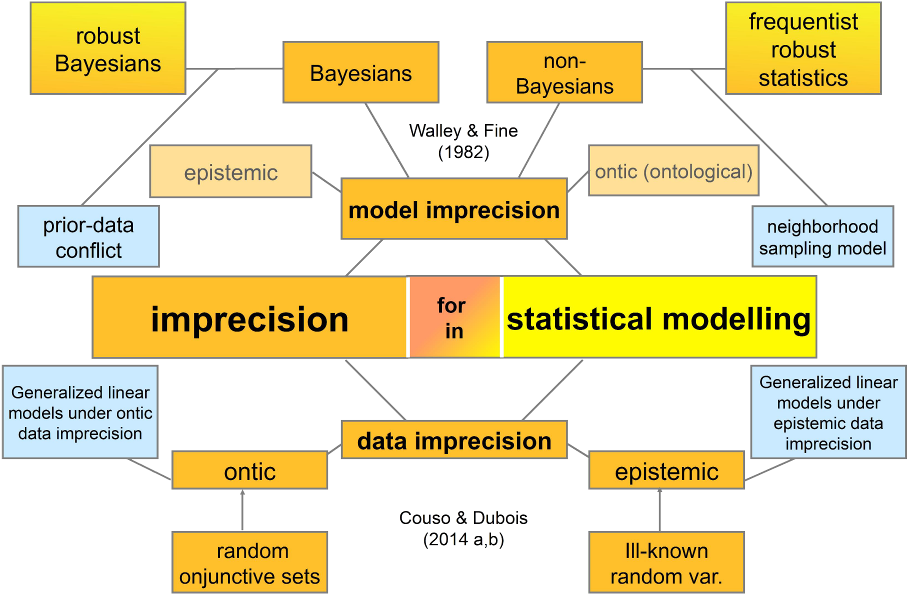
# Ontic imprecision: modelling idea

- Modelled by random conjunctive sets
- Change sample space $\mathcal{S} = \{CD, SPD, Green, Left, \ldots\}$ into $\mathcal{S}^* \subset \mathcal{P}(\mathcal{S})$
- Oberservations are precise observations in $\mathcal{S}^*$ and can be treated as like tradtional categorical data
- Whole statistical modelling framework can be applied, here logistic regression
- For each non-empty element of $\mathcal{S}^*$ vector of regression coefficients

# Ontic imprecision: example, Plass et al (2015, Table 4)

| Coefficient | ontic | | classical |
|---|---|---|---|
| | CD | G:S | CD |
| intercept | 0.33 | −1.41 ** | −0.12 |
| rel.christ | 0.37 ** | −0.25 | 0.52 *** |
| info.tv | −0.02 | −0.32 | 0.25 |
| info.np | −0.12 | −1.69 ** | 0.13 |

Table 4: Comparison of results (first vote).


robust Bayesians
Bayesians
non-Bayesians
frequentist robust statistics
Walley & Fine (1982)
epistemic
ontic (ontological)
model imprecision
prior-data conflict
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 a,b)
random onjunctive sets
Ill-known random var.

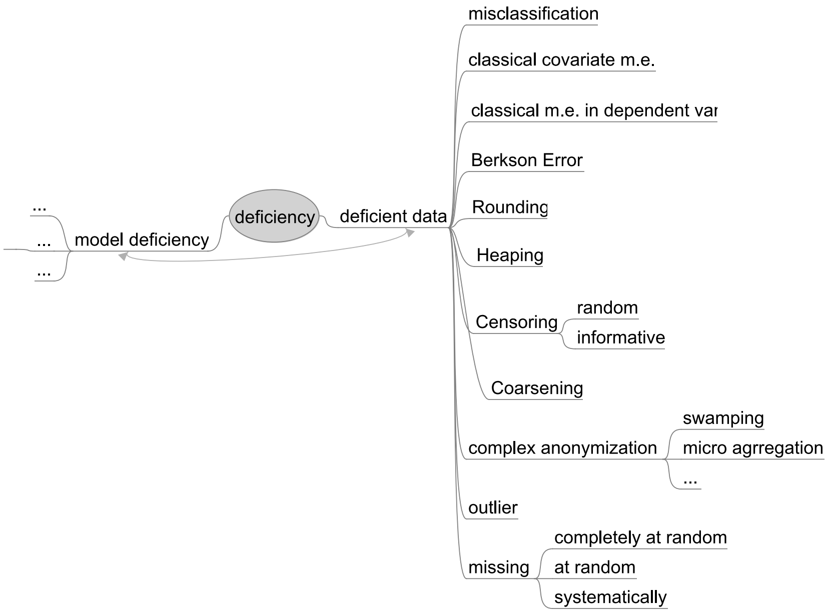
# Table of Contents

# Epistemic data imprecision

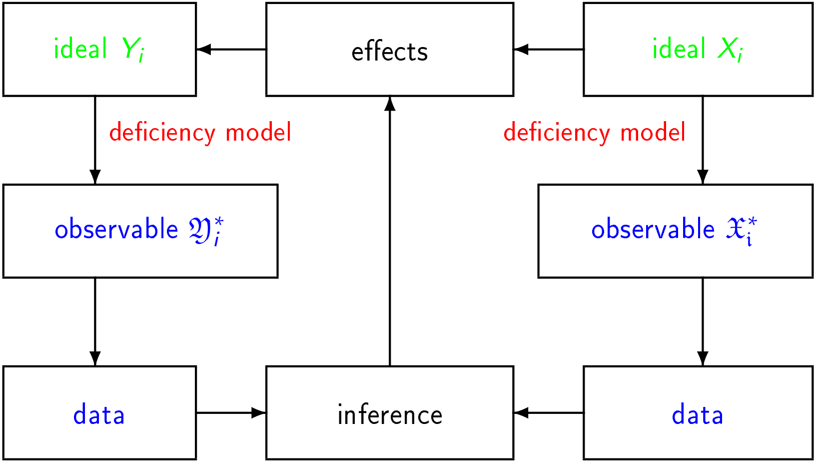
imprecise observation of something precise

- missing data (refusals, treatment design)
- data protection
- data merging with partially overlapping categories
- secondary analysis
- forecasts derived from set-valued (ontic) observations
- primary refusals, typically coarsening/missing not at random

# Spinney of Deficiencies



# The two-layers perspective



# Traditional treatment of deficiencies

- Model the deficiency process!
- Characterize situations where the deficiency may be ignored or when one can correct for it!
- But typically very restrictive – often untestable – asumptions needed to ensure identifiability = precise solution

# Traditional treatment of deficiencies

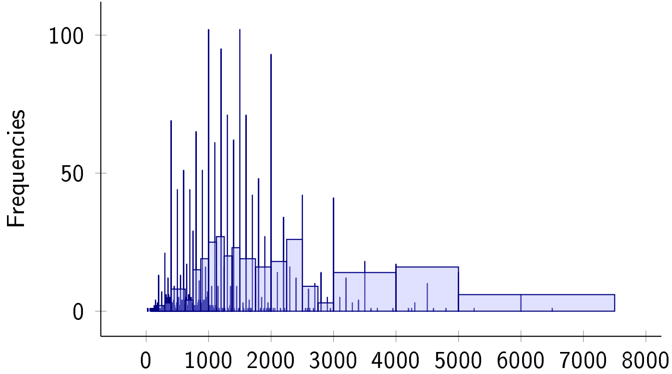
For instance, in measurement error models ("classical model of testing theory"):

measurement error model must be known precisely

- type of error, especially assumptions on (conditional) independence
  - independence of true value
  - independence of other covariates
  - independence of other measurements
- type of error distribution
- moments of error distribution

validation studies typically not available

German General Social Survey (ALLBUS) 2010:
2827 observations in total, approx. 2000 report personal income (30% missing). An additional 10% report only income brackets.



# Interval Data: Example

1. We see *heaping* at 1000 €, 2000 €, ..., less so at 500 €, 1500 €, ...
2. Both heaping and grouping depend on the amount of income reported.
3. Missingness (some 20% of the data) might as well depend on the amount of income.

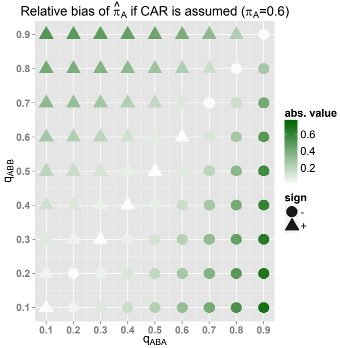
# Interval Data: Example

1. We see *heaping* at 1000 €, 2000 €, ..., less so at 500 €, 1500 €, ...
2. Both heaping and grouping depend on the amount of income reported.
3. Missingness (some 20% of the data) might as well depend on the amount of income.

*Consequences:*

1. Missingness, grouping, and heaping can often be represented by intervals.
2. Missingness, grouping, and heaping will rarely conform to the assumption of "coarsening at random" (CAR).
3. Missingness, grouping, and heaping add an additional type of uncertainty apart from classical statistical uncertainty. This uncertainty can't be decreased by sampling more data.

# Wrongly assuming CAR



# Manski's Law of Decreasing Credibility

**Reliability !? Credibility ?**

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

# Reliable Inference Instead of Overprecision!!

Consequences from the Law of Decreasing Credibility:

- Make *realistic* assumptions and let the data speak for themselves!
- The results may be imprecise, but are more reliable
- The extent of imprecision is related to the data quality!
- Often still sufficient to answer subjective matter question

# Much IP work on epistemic date impprecision, e.g.

- De Cooman & Zaffalon (2004, AI), Zaffalon & Miranda (2009, JAIR)
- Utkin & Augustin (2007, IJAR), Troffaes & Coolen (2009, IJAR)
- Utkin & Coolen (2011, ISIPTA).
- Cattaneo & Wiencierz (2012, IJAR)
- Schollmeyer & Augustin (2015, IJAR)
- Denoeux (2014, IJAR)

# Relation to work in econometrics, biometrics and engineering

- Partial identification: Manski (2003, Springer)
- Systematic sensitivity analysis: Vansteelandt, Goetghebeur, Kenword, Molenberghs (2006, Stat. Sinica)
- Ferson; reliable computing community, interval: Kreinovich

# Recent likelihood approach

Plass, Augustin, Schollmeyer (2015, ISIPTA)

- Utilize invariance of likelihhod under paramtertrans formation
- observable part: set-valued observations, parameter $\vartheta$, maximum likelihood estimator $\vartheta$
- latent part: parameter of interest $\gamma$
- related via observation model: expressed by mapping $\Phi$
- set-valued maximum likelihood estimator $\hat{\Gamma} = \left\{\gamma | \Phi(\gamma) = \hat{\vartheta}\right\}$
- application also to some basic logistic regression models

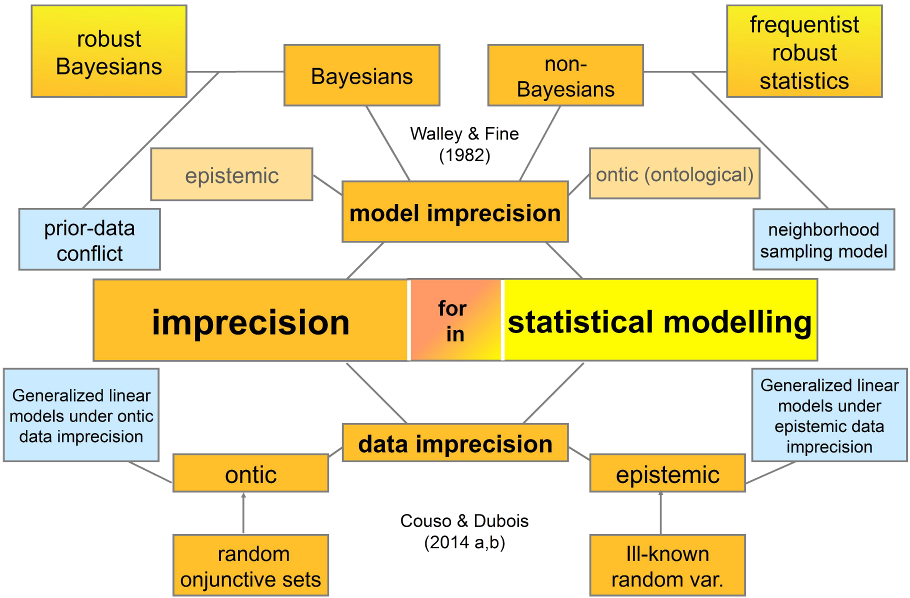
# Estimating Equations

Generalizing from the linear case, suppose there is a consistent (score-) estimating equation for the ideal model $\{\mathcal{P}_\vartheta \,|\, \vartheta \in \Theta\}$, i.e.:

$$\forall \vartheta \in \Theta : \;\; \mathbb{E}_\vartheta(\psi(X, Y; \vartheta)) = 0$$

With interval data, one gets a set of estimating equations, one for each random vector (selection) $(X, Y) \in (\mathfrak{X}, \mathfrak{Y})$:

$$\Psi(\mathfrak{X}, \mathfrak{Y}; \vartheta) := \{\psi(X, Y; \vartheta) \,|\, X \in \mathfrak{X}, Y \in \mathfrak{Y}\}$$


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 a,b)
random onjunctive sets
Ill-known random var.

# Table of Contents

# Concluding Remarks

- Law of decreasing credibility !
- Reliable use of information
- Sets-valued analysis: imprecise data, imprecise models
- Imprecise but reliable results; often sufficient!
- Natural behaviour of imprecision!
- Use this actively in modelling
- Towards a general framework for reliable analysis of non-idealized data

# On the power of IP in statistical modelling

- defensive point-of-view
  - IP protects against the potential disastrous behaviour of standard procedures under violated assumptions → robustness in:
  - frequentist and
  - Bayesian settings

# On the power of IP in statistical modelling

- offensive point of view.
  IP is a most powerful methodology, allowing for
  - separation of variability (variance) from indeterminism
  - active modelling of ignorance
  - active modelling of conflicting/surprising information
  - active use of weak knowledge that can not be used in the traditional setting

# Future directions

Popularize the defensive point of view

- case studies, illustrating the power of
- robust procedures for generalized linear models etc.
- cautious data completion for generalized linear models etc.
- (disc. with H. Rieder): for each result complement p-value routinely by stability level: smallest level of contamination where the result is no longer significant

Propagate the offensive view

- case studies, illustrating the power of
- separation of variability (variance) from indeterminism
- active modelling of ignorance
- active modelling of conflicting/surprising information
- active use of weak knowledge that can not be used in the traditional setting

# Future directions

- Statisticians start to think from data
  $\rightarrow$ improve understanding of imprecise *sampling* models
  - imprecise probabilities for the *observables*!
  - generalized sampling theory: imprecise selection probabilities
  - utilize variety of independence concepts (model slight dependence)
  - develop methodology of estimation from imprecise sampling models
- develop simulation techniques for imprecise probabilities
- how to handle regression models?

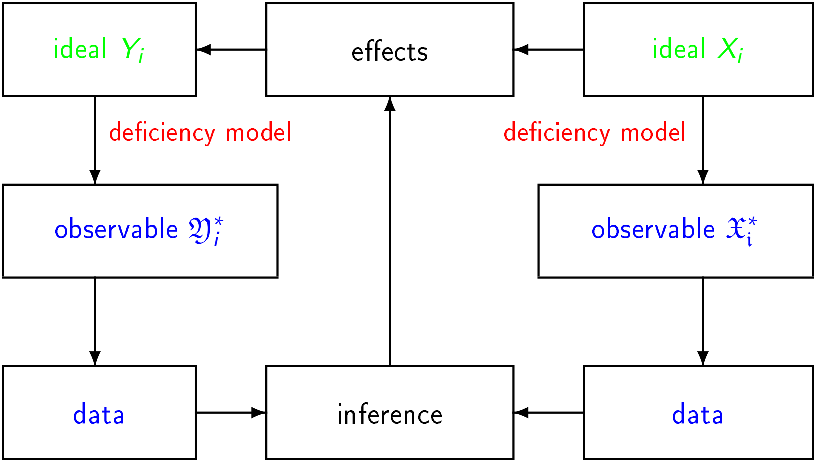
# Future directions

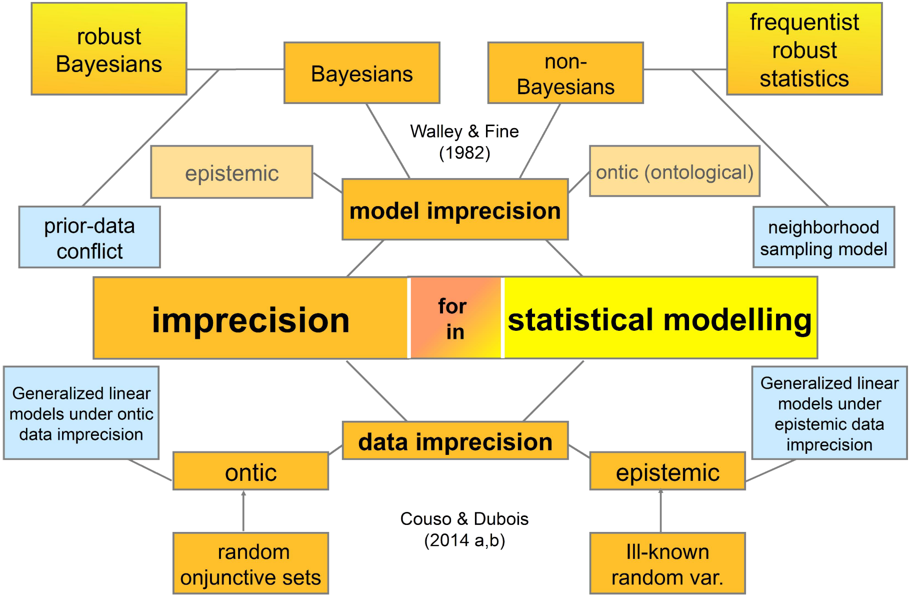
- Develop heuristics, "semi imprecise" methods
  "IP should make life better or easier (or both)" (Frank Coolen)
- Develop direct methods
  - leave the necessarily more complicated "set-of traditional model views"
  - direct processing of information (e.g., statistics with desirable gambles?)

# Future directions

- Develop a methodology for statistical modelling with sets of models
  - generalized linear models
  - nonparametric regression models → smoothing
  - variable selection
  - realistic measurement error and random effect models
  - importance of unbiased estimation equations

# The two-layers perspective




robust Bayesians
Bayesians
non-Bayesians
frequentist robust statistics
Walley & Fine (1982)
epistemic
ontic (ontological)
model imprecision
prior-data conflict
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
random onjunctive sets
Couso & Dubois (2014 a,b)
Ill-known random var.