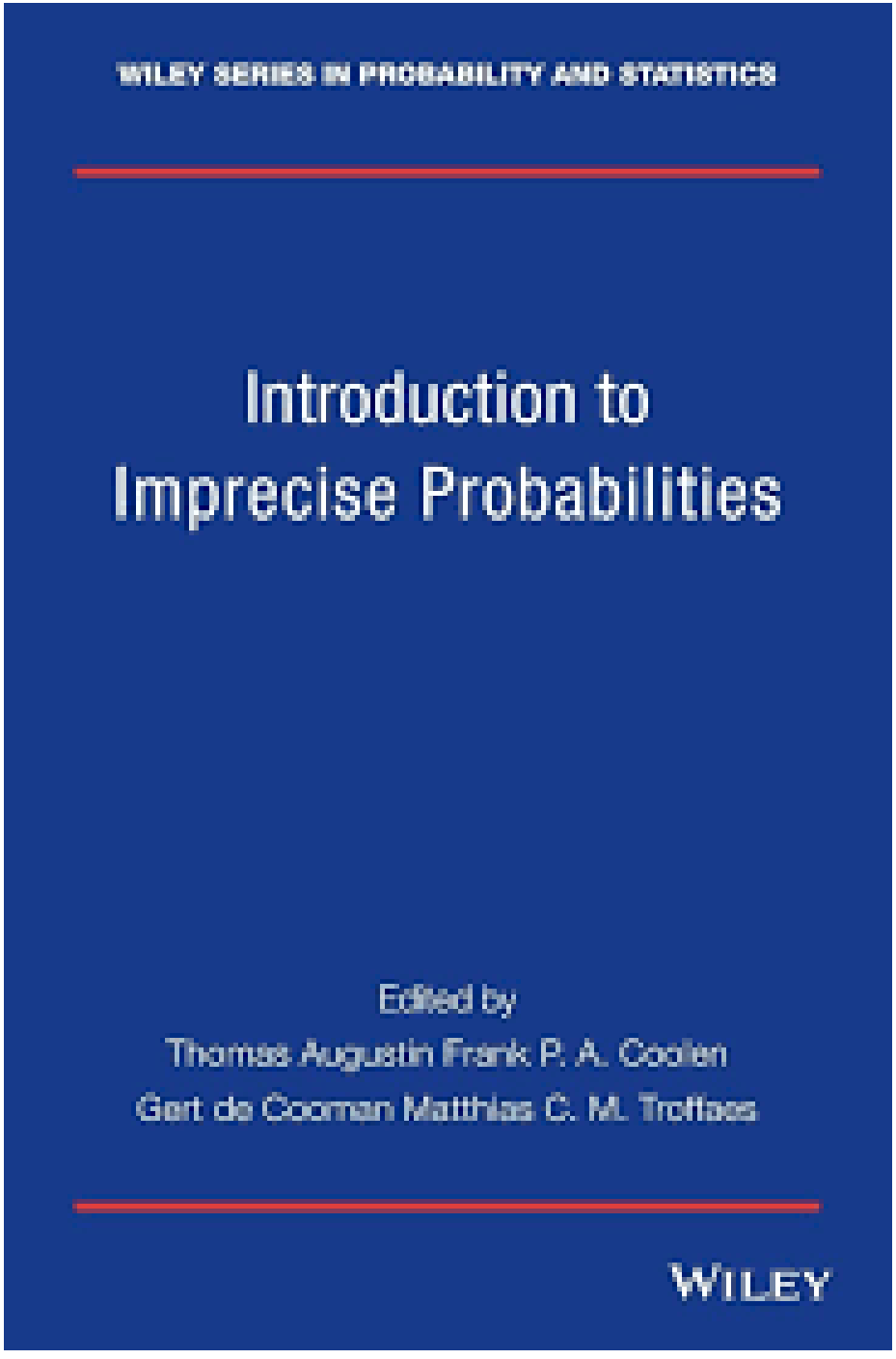
Statistics: Partial Identification and (a first look into) Robust Statistics

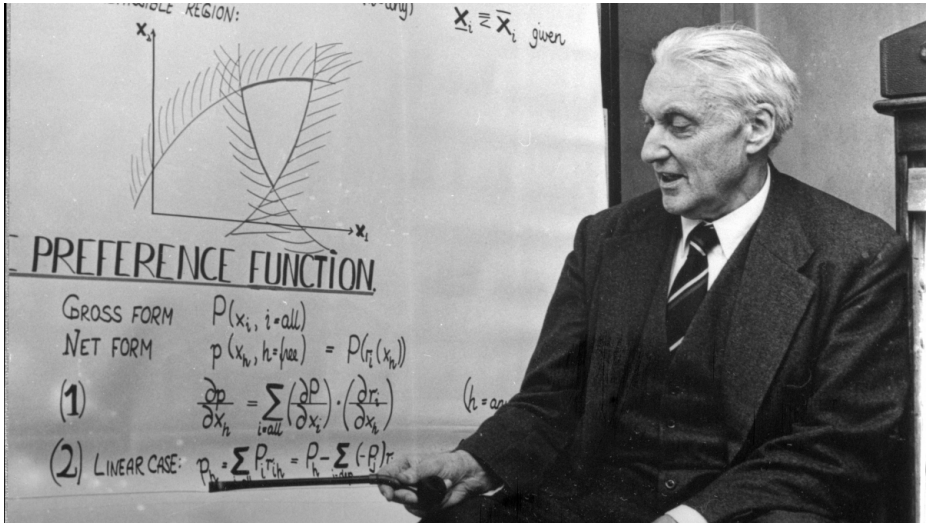
Georg Schollmeyer (georg.schollmeyer@stat.uni-muenchen.de)

30.07.2026


WILEY SERIES IN PROBABILITY AND STATISTICS
Introduction to
Imprecise Probabilities
Edited by
Thomas Augustin Frank P. A. Coolen
Gert de Cooman Matthias C. M. Troffaes
WILEY

(Sources of Partial) Identification

An introductory example: Frisch's (1885 - 1973) true regression



(first Nobel Prize in economics (together with Jan Tinbergen) in 1969)

Introductory example: (Ragnar) Frisch's true regression [Frisch, 1934]

Model of **true** variables of interest (not observable):

- $\tilde{Y} = \beta_0 + \beta_1 \tilde{X} + \varepsilon$
- $\varepsilon \perp \tilde{X}$
- $\mathbb{E}(\varepsilon) = 0$
- $Var(\varepsilon) < \infty$

Model of actually observed variables (think of e.g. measurement error):

- $X = \tilde{X} + \delta_1$
- $Y = \tilde{Y} + \delta_2$
- $Var(\delta_1) < \infty, Var(\delta_2) < \infty$
- $\varepsilon, \delta_1, \delta_2$ indpendent of each other and indpendent of $\tilde{X}, \tilde{Y}$

Some things to remember / taking note of

- The covariance of two random variables (or samples thereof) is a (positive semidefinite symmetric) bilinear form, denoted here with

$$\langle \cdot, \cdot \rangle.$$

- For $X \perp Y$ we have $\langle X, Y \rangle = 0$ (if the covariance exists)
- The slope of a simple linear regression (of y on x) can be computed as

$$\hat{\beta} = \frac{\langle x, y \rangle}{\langle x, x \rangle}.$$

- Analogously, for the population version we have

$$\beta = \frac{\langle X, Y \rangle}{\langle X, X \rangle} \text{ (if (co-)variances exist).}$$

- For multivariate linear regression we can estimate the model via ordinary least squares (OLS):

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T Y$$

with $X = \begin{pmatrix} 1, & X_{11}, \ldots, X_{1p} \\ 1, & X_{21}, \ldots, X_{2p} \\ \vdots & \\ 1, & X_{n1}, \ldots, X_{np} \end{pmatrix}$ the design matrix and $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$ the response vector.

$$\hat{\beta} = \underbrace{\left(X^T X\right)^{-1} X^T}_{=:M} Y$$

- For fixed X this is a linear function in Y
- For fixed Y, is this a linear function in X?
- Is $\left(X^T X\right)^{-1}$ always well-defined?

Slope of the direct regression of Y on X

$$
\begin{aligned}
\beta_{YX} &= \frac{\langle X, Y\rangle}{\langle X, X\rangle} \\
&= \frac{\langle \tilde{X} + \delta_1, \tilde{Y} + \delta_2\rangle}{\langle \tilde{X} + \delta_1, \tilde{X} + \delta_1\rangle} \\
&= \frac{\langle \tilde{X} + \delta_1, \beta_0 + \beta_1 \tilde{X} + \varepsilon + \delta_2\rangle}{\langle \tilde{X}, \tilde{X}\rangle + \langle \delta_1, \delta_1\rangle} \\
&= \frac{\beta_1 \langle \tilde{X}, \tilde{X}\rangle}{\langle \tilde{X}, \tilde{X}\rangle + \langle \delta_1, \delta_1\rangle} \\
&= \beta_1 \cdot \frac{\langle \tilde{X}, \tilde{X}\rangle}{\langle \tilde{X}, \tilde{X}\rangle + \langle \delta_1, \delta_1\rangle} \\
&\leq \beta_1
\end{aligned}
$$

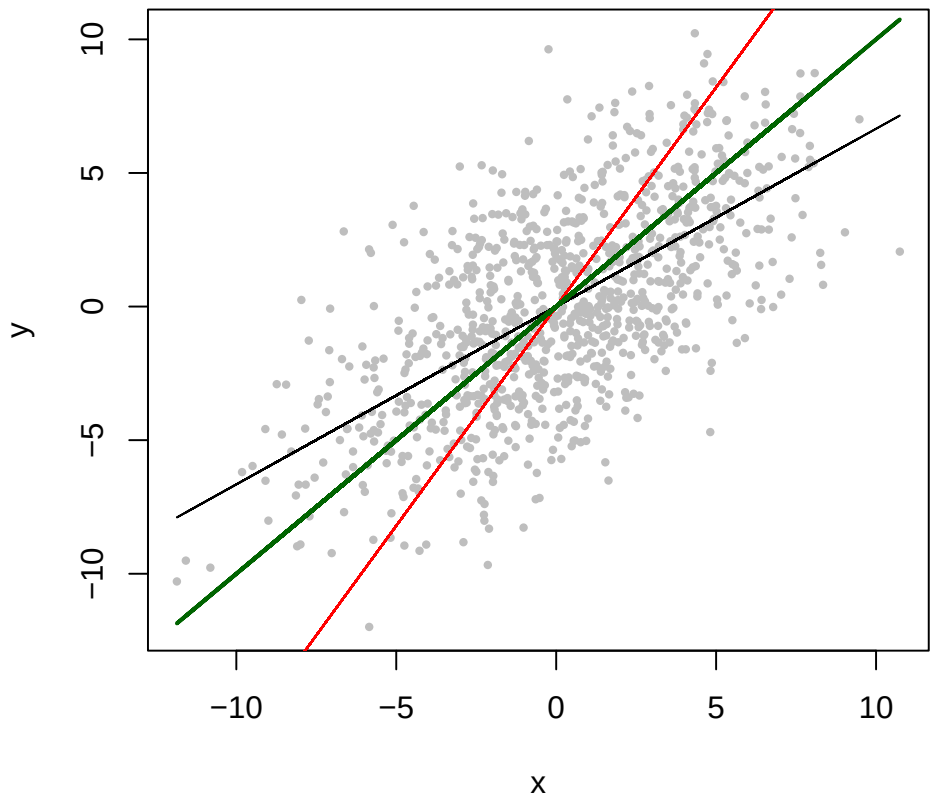
with equality iff $\langle \delta_1, \delta_1\rangle = 0$, i.e., iff $\delta_1 =$ constant almost surely.

Slope of the inverse regression of X on Y

$$
\begin{aligned}
\beta_{XY} &= \frac{\langle X, Y\rangle}{\langle Y, Y\rangle} \\
&= \frac{\langle \tilde{X} + \delta_1, \beta_0 + \beta_1 \tilde{X} + \varepsilon + \delta_2\rangle}{\langle \beta_0 + \beta_1 \tilde{X} + \varepsilon + \delta_2, \beta_0 + \beta_1 \tilde{X} + \varepsilon + \delta_2\rangle} \\
&= \frac{\beta_1 \langle \tilde{X}, \tilde{X}\rangle}{\beta_1^2 \langle \tilde{X}, \tilde{X}\rangle + \langle \varepsilon, \varepsilon\rangle + \langle \delta_2, \delta_2\rangle} \\
&\leq \frac{\beta_1 \langle \tilde{X}, \tilde{X}\rangle}{\beta_1^2 \langle \tilde{X}, \tilde{X}\rangle} \\
&= \frac{1}{\beta_1} \\
&\implies \beta_1 \geq \frac{1}{\beta_{XY}}
\end{aligned}
$$

with equality iff ε and δ_2 are constant almost surely.

$$\beta_1 \in \left[\beta_{YX}, \frac{1}{\beta_{XY}}\right]$$



We can only deduce these bounds!

Therefore the model is only partially identified!

Partial Identifiction

Definition (Identification)
A statistcial Model $\{P_\vartheta \mid \vartheta \in \Theta\}$ is identified if the joint **distribution** of all **observable** variables uniquely identifies the underlying parameter(s), i.e. if the mapping

$$\Phi : \vartheta \mapsto \text{ distribution of all observables}$$

is injective, i.e., if:

$$P_\vartheta = P_{\tilde{\vartheta}} \Longrightarrow \vartheta = \tilde{\vartheta}.$$

Otherwise it is called not identified or partially identified.

Exercise: The simple linear model with one pair of observations

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon$$

X, Y : observable random variables

ε : noise

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \in \Theta = \mathbb{R}^2$$

$$0 < Var(X) < \infty$$

$$\mathbb{E}(\varepsilon) = 0$$

$$Var(\varepsilon) < \infty$$

$$\varepsilon \perp X$$

Assume, we observe only a sample of size $n = 1$, i.e. one observation for X and one for Y. Is this model identified? (And is this model "reasonably estimable"?)

Solution

$$\begin{aligned}\langle X, Y\rangle &= \langle X, \beta_0 + \beta_1 X + \varepsilon\rangle \\ &= \langle X, \beta_1 X\rangle \\ &= \beta_1 \cdot \underbrace{Var(X)}_{>0}\end{aligned}$$

$$\begin{aligned}\mathbb{E}(Y) &= \mathbb{E}(\beta_0 + \beta_1 X + \varepsilon) \\ &= \beta_0 + \beta_1 \mathbb{E}(X)\end{aligned}$$

$$\beta_1 = \frac{\langle X, Y \rangle}{Var(X)}$$

$$\beta_0 = \mathbb{E}(Y) - \beta_1 \cdot \mathbb{E}(X)$$

What does this mean for identifiability?

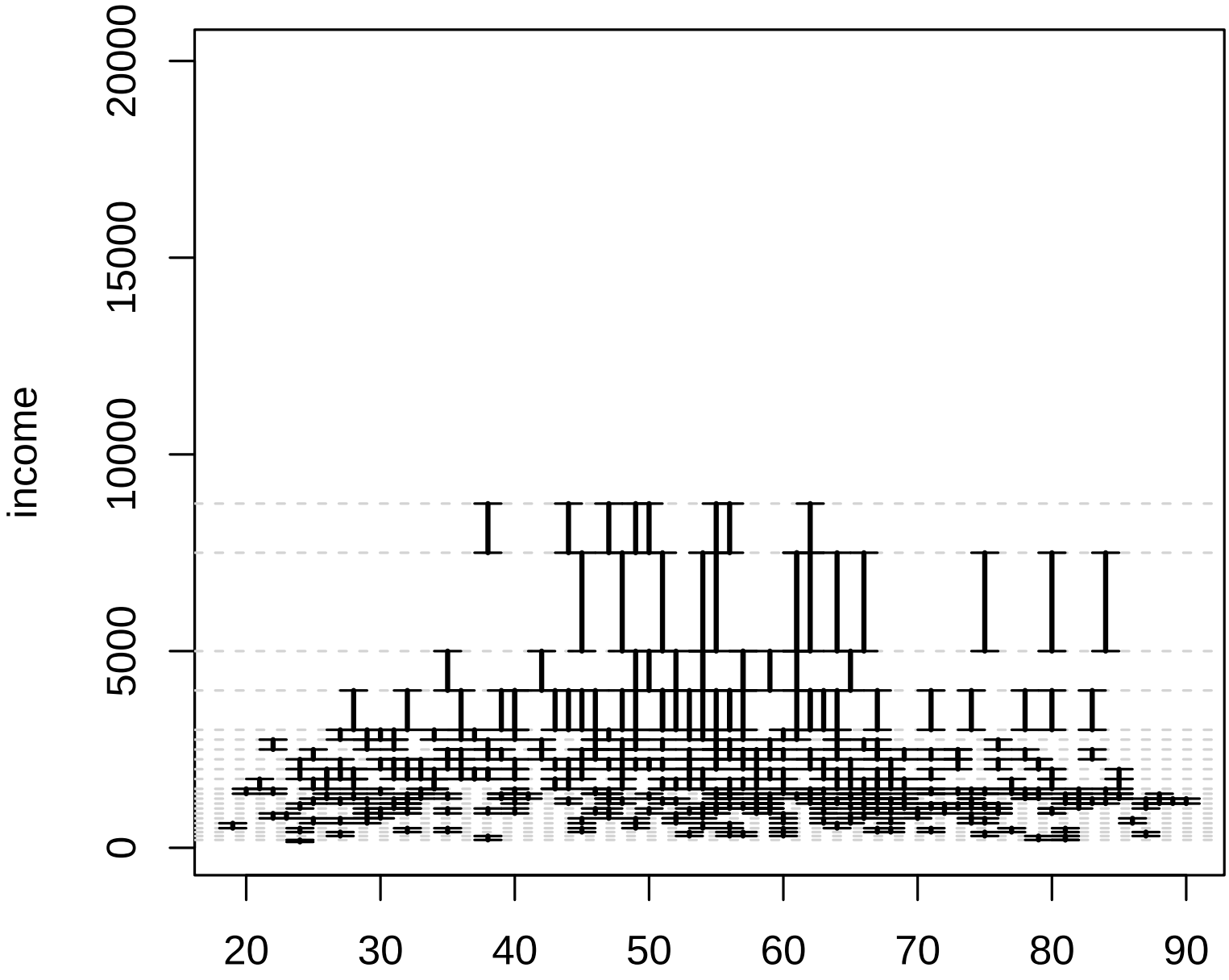
What about estimability for $n = 1$?

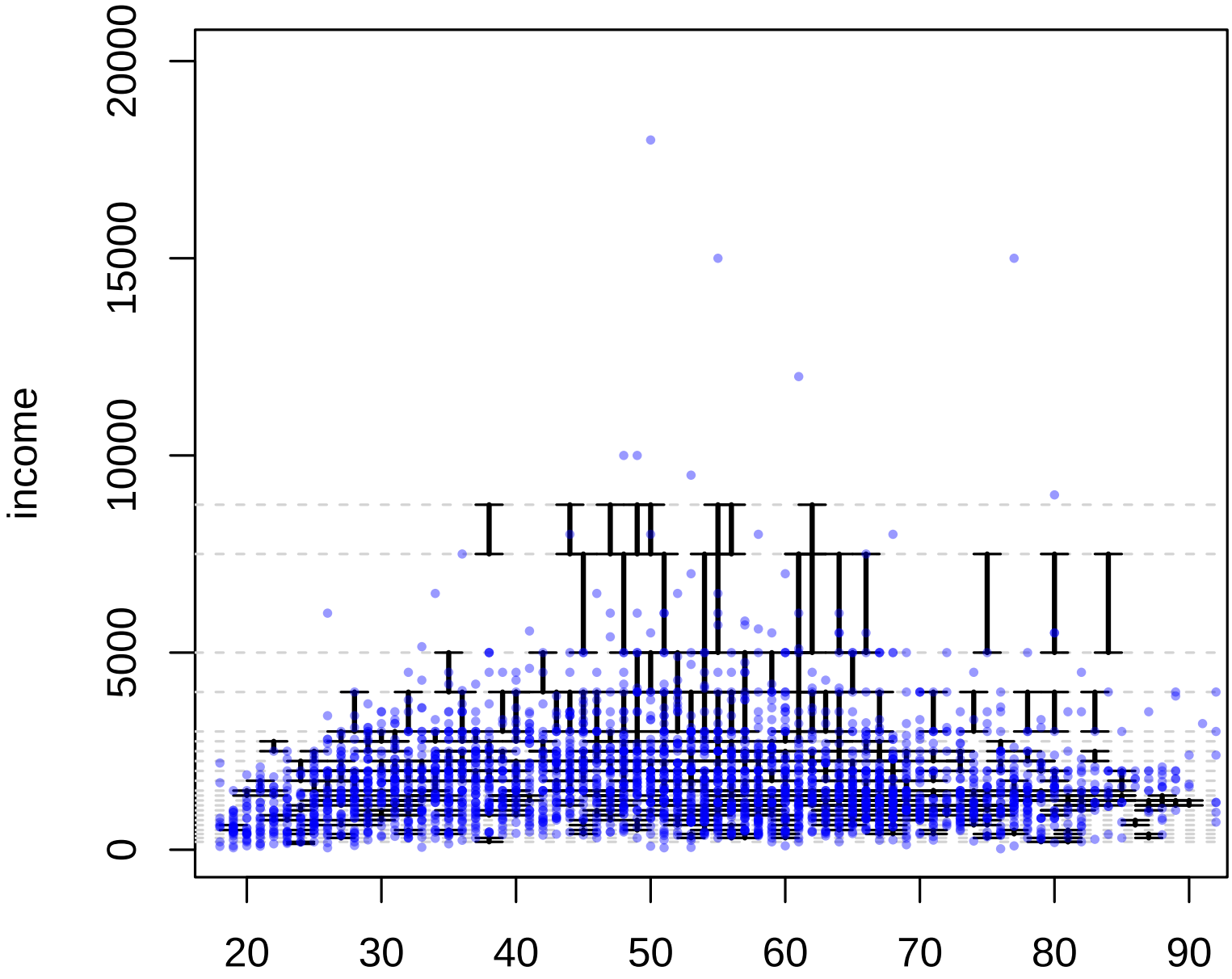
Another prototypic example of partial Identification: Interval data in linear regression

- Dataset: Allbus (German General Social Survey) 2018
- 3477 respondents
- Consider question about age and income
- Only 2644 respondents answered the open question about income
- But 3087 respondents answered either the open question or the list query about income (e.g., income $\in$ [1500€ – 1749€])
- Model:

$$income = \beta_0 + \beta_1 \cdot age + \varepsilon$$

(inadequate, only for illustration!)


income
0
5000
10000
15000
20000
20
30
40
50
60
70
80
90


income
0
5000
10000
15000
20000
20
30
40
50
60
70
80
90

How to estimate a partially identified model

- In the first place: not (consistently) possible
- Instead: estimate the identified set, i.e. estimate the inverse image of Φ
- Generally, developing reasonable (in a certain sense optimal) estimators is a very hard problem (not yet solved?)

One “simple” heuristic approach for deficient data

Definition (Cautious Data Completion, Augustin et al. [2014])
Let $\mathcal{X}$ be the sample space for the data of actual interest (think of $\mathcal{X} = \mathbb{R}$ or $\mathcal{X} = \mathbb{R}^d$). Let $(X_1, \ldots, X_n)$ be a sample thereof (think of an i.i.d. sample). Let $(X_1, \ldots, X_n)$ be not observable. Let $(\mathfrak{X}_1, \ldots, \mathfrak{X}_n)$ be observable and all we know is

$$X_i \in \mathfrak{X}_i \subseteq \mathcal{X}$$

for $i = 1, \ldots, n$. Let T be a classical estimator for the parameter of interest $\vartheta \in \Theta$. Then, the **cautious data completion (CDC)** is defined as

$$\mathfrak{T}(\mathfrak{x}_1, \ldots, \mathfrak{x}_n) := \{ T(x_1, \ldots, x_n) \mid x_i \in \mathfrak{x}_i, i = 1, \ldots, n \}.$$

Exercise

- $X_1, \ldots, X_n \overset{i.i.d.}{\sim} P_\vartheta$
- Parameter of interest: $\vartheta := \mathbb{E}(X_1) < \infty$
- What is a reasonable estimator?
- average: $ave(X_1, \ldots, X_n) := \frac{1}{n} \sum_{i=1}^{n} X_i$
- Now, let $(X_1, \ldots X_n)$ be only observable in intervals, i.e.

$$X_i \in \mathfrak{x}_i = \left[\underline{X}_i, \overline{X}_i\right]$$

- What do you think the cautious data completion (CDC) would look like?

Solution

$$\mathfrak{ave}(\mathfrak{x}_1, \ldots \mathfrak{x}_n) = [ave(\underline{x}_1, \ldots \underline{x}_n), ave(\overline{x}_1, \ldots \overline{x}_n]$$

Exercise

- $X_1, \ldots, X_n \overset{i.i.d.}{\sim} P_\vartheta$
- Parameter of interest: $\vartheta := Var(X_1) < \infty$
- What is a reasonable estimator?
- (corrected) variance:

For $i \neq j$ we have ("location-free representation of the variance"):

$$\begin{aligned}
\mathbb{E}((X_i - X_j)^2) &= \mathbb{E}(X_i^2 - 2X_iX_j + X_j^2) \\
&= \mathbb{E}(X_i^2) - 2\mathbb{E}(X_i)\mathbb{E}(X_j) + \mathbb{E}(X_j^2) \\
&= \mathbb{E}(X_i^2) - 2[\mathbb{E}(X_i)]^2 + \mathbb{E}(X_i^2) \\
&= 2\left[\mathbb{E}(X_i^2) - [\mathbb{E}(X_i)]^2\right] \\
&= 2Var(X_1).
\end{aligned}$$

- Therefore, $1/2 \cdot (X_i - X_j)^2$ would be an unbiased estimator of $Var(X_1)$.
- But a better estimator uses all data, i.e.:

$$\hat{Var}(X_1, \ldots, X_n) := \frac{1}{2 \cdot \binom{n}{2}} \sum_{i \neq j} (X_i - X_j)^2$$

$$\left(= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - ave(X_1, \ldots, X_n))^2 \right)$$

- Now, let $(X_1, \ldots X_n)$ be only observable in intervals, i.e. $X_i \in \mathfrak{x}_i := [\underline{X}_i, \overline{X}_i]$
- What do you think the cautious data completion would look like?

Computing Variance for Interval Data is NP-Hard*

Scott Ferson[1], Lev Ginzburg[1],
Vladik Kreinovich[2], Luc Longpré[2], and Monica Aviles[2]

[1]Applied Biomathematics, 100 North Country Road,
Setauket, NY 11733, USA, {scott,lev}@ramas.com
[2]Computer Science Department, University of Texas at El Paso
El Paso, TX 79968, USA, {maviles,longpre,vladik}@cs.utep.edu

Abstract

When we have only interval ranges $[\underline{x}_i, \overline{x}_i]$ of sample values $x_1, \ldots, x_n$, what is the interval $[\underline{V}, \overline{V}]$ of possible values for the variance V of these values? We prove that the problem of computing the upper bound $\overline{V}$ is NP-hard. We provide a feasible (quadratic time) algorithm for computing the lower bound $\underline{V}$ on the variance of interval data. We also provide a feasible algorithm that computes $\overline{V}$ under reasonable easily verifiable conditions.

1 Introduction

When we have n measurement results $x_1, \ldots, x_n$, traditional statistical approach usually starts with computing their (sample) average $E = \bar{x} = \frac{x_1 + \ldots + x_n}{n}$ and their (sample) variance $V = \frac{(x_1 - E)^2 + \ldots + (x_n - E)^2}{n - 1}$.

In some practical situations, we only have intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ of possible values of x_i. This happens, for example, if instead of observing the actual value x_i of the random

Approximate Solution via Interval Analysis

Define

$$\underline{\Delta}_{ij}^{x} := \min_{x_i \in \mathfrak{x}_i, x_j \in \mathfrak{x}_j} (x_i - x_j)^2 = \begin{cases} 0 & \text{if } \mathfrak{x}_i \cap \mathfrak{x}_j \neq \emptyset \\ (\underline{x}_i - \overline{x}_j)^2 & \text{if } \underline{x}_i > \overline{x}_j \\ (\underline{x}_j - \overline{x}_i)^2 & \text{if } \underline{x}_j > \overline{x}_i \end{cases}$$

$$\overline{\Delta}_{ij}^{x} := \max_{x_i \in \mathfrak{x}_i, x_j \in \mathfrak{x}_j} (x_i - x_j)^2 = max\left((\overline{x}_i - \underline{x}_j)^2, (\overline{x}_j - \underline{x}_i)^2 \right).$$

Then, an outer approximation of the CDC is given by

$$\underline{Var}(\mathfrak{x}_1, \ldots, \mathfrak{x}_n) = \frac{1}{2\binom{n}{2}} \sum_{i \neq j} \underline{\Delta}_{ij}^{x} \quad \text{and}$$

$$\overline{Var}(\mathfrak{x}_1, \ldots, \mathfrak{x}_n) = \frac{1}{2\binom{n}{2}} \sum_{i \neq j} \overline{\Delta}_{ij}^{x}$$

Why is this generally only an outer approximation?

What about linear regression for interval-valued y?

- Remember: For fixed (precise) X, the OLS estimator for β is linear in Y:

$$\hat{\beta}_i = \sum_{j=1}^{n} M_{ij} \cdot Y_j.$$

Solution

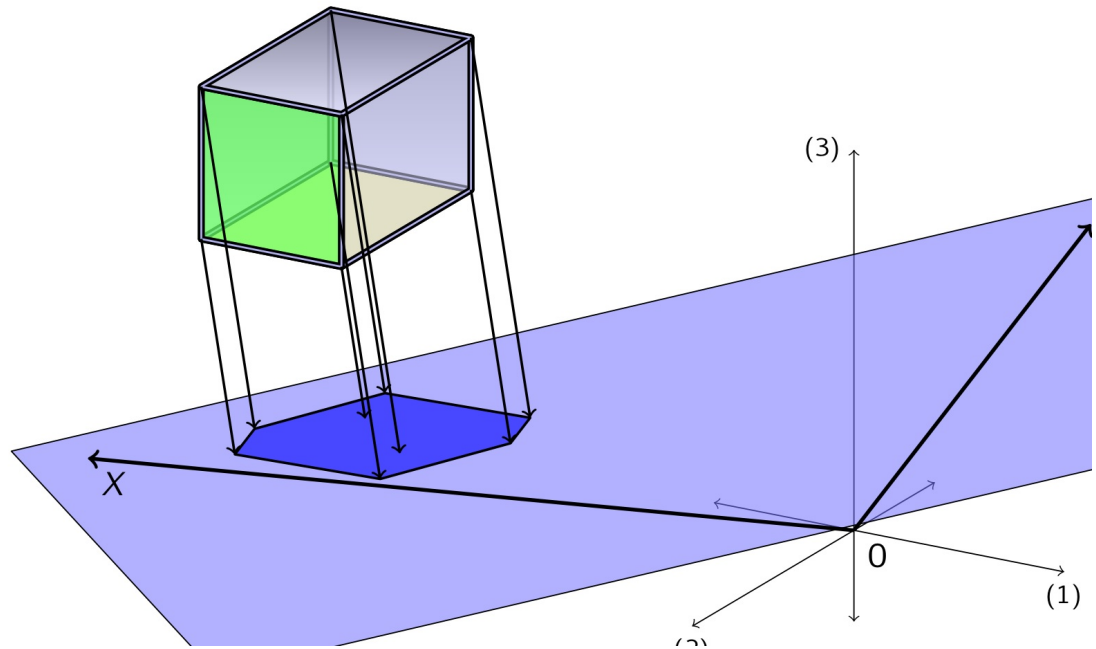
$$\underline{\hat{\beta}}_i = \sum_{j=1}^{n} M_{ij} \cdot L_j$$

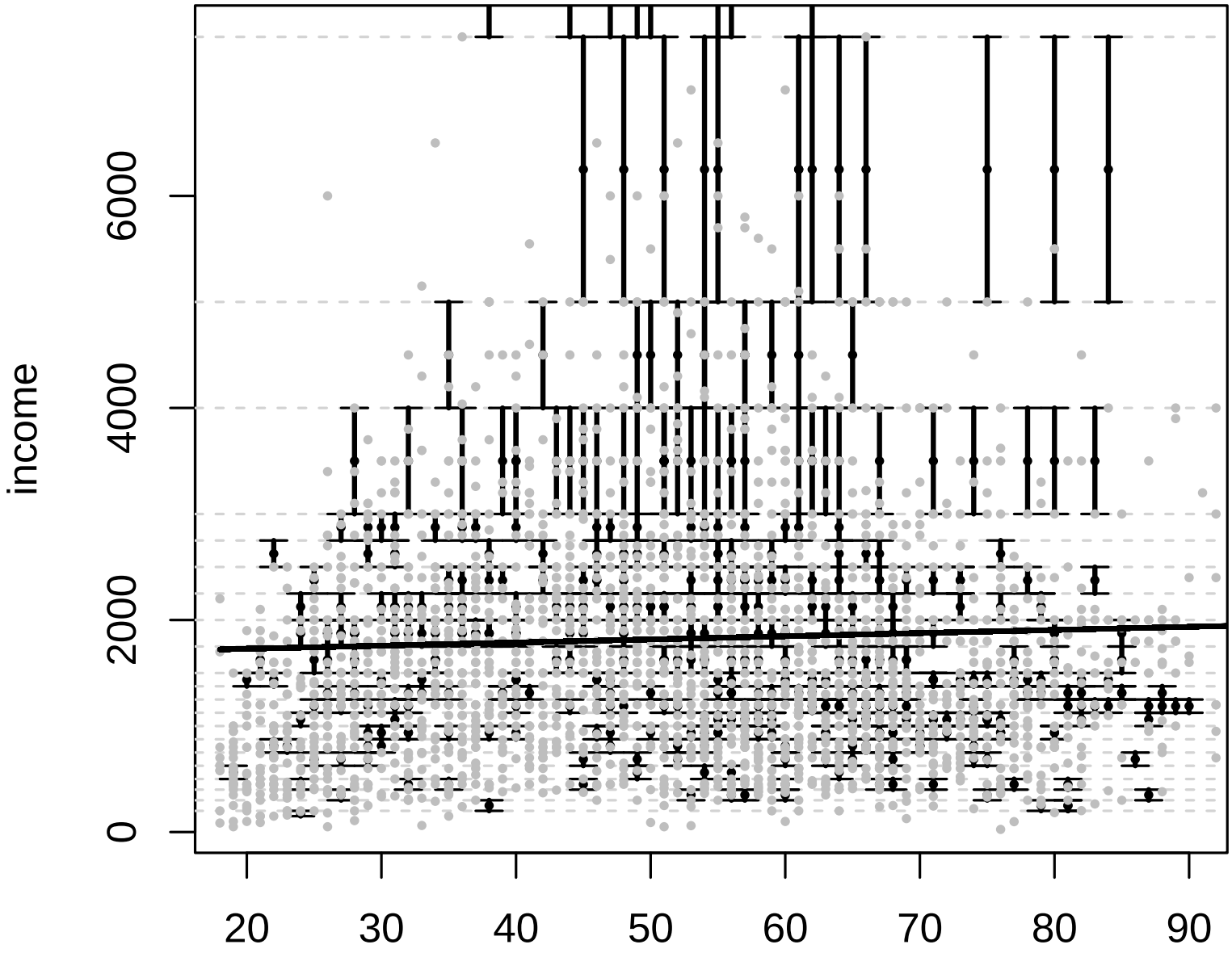
$$L_j = \begin{cases} Y_j & \text{if } M_{ij} > 0 \\ \overline{Y_j} & \text{else} \end{cases}$$

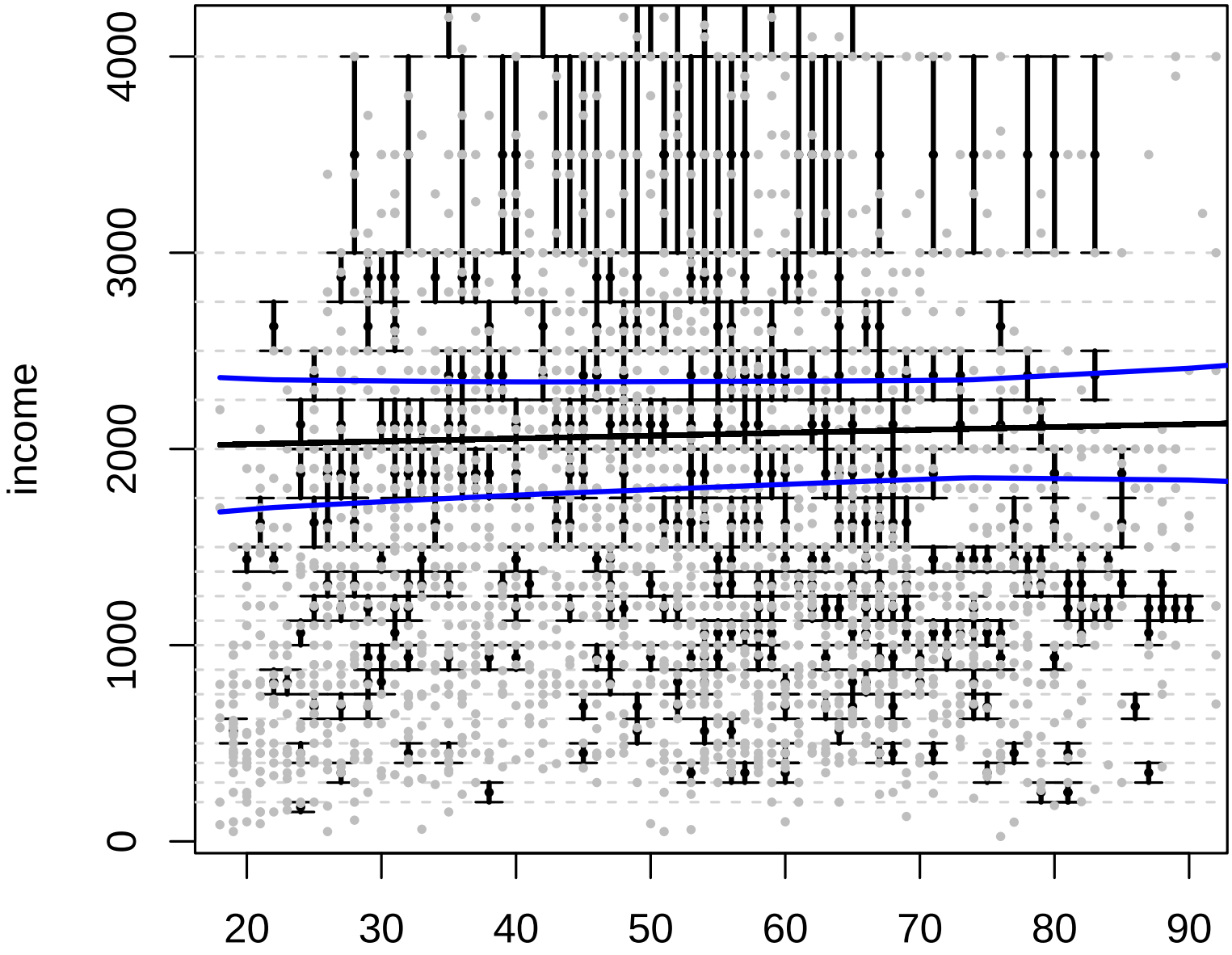
$$\overline{\hat{\beta}}_i = \sum_{j=1}^{n} M_{ij} \cdot U_j$$

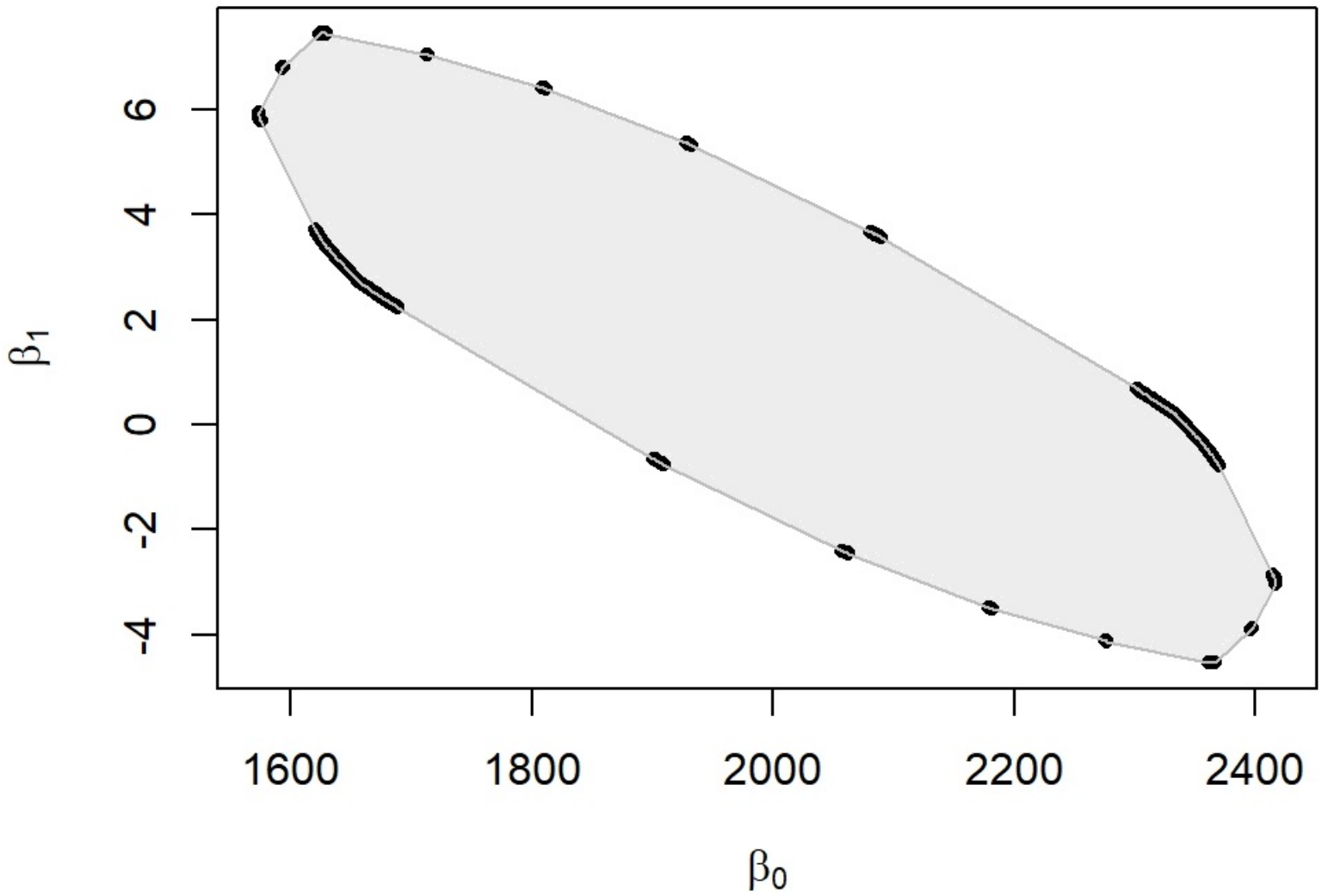
$$U_j = \begin{cases} Y_j & \text{if } M_{ij} < 0 \\ \overline{Y_j} & \text{else} \end{cases}$$

- The exact identified set/CDC is the linear image of the cuboid $[\underline{Y}, \overline{Y}]$ under the linear map induced by M
- It can be computed for example via so-called support functions
- Geometrically it is a so-called zonotope/zonoid known from computational geometry




income
0
2000
4000
6000
20
30
40
50
60
70
80
90


income
0
1000
2000
3000
4000
20
30
40
50
60
70
80
90


β1
6
4
2
0
-2
-4
1600
1800
2000
2200
2400
β0

What about (the slope of a simple) linear regression for interval-valued x and y?

What about (the slope of a simple) linear regression for interval-valued x and y?

For $i \neq j$ look at $\hat{\beta}_{ij} := \frac{Y_i - Y_j}{X_i - X_j}$:

$$
\begin{aligned}
\mathbb{E}(\hat{\beta}_{ij}) &= \mathbb{E}\left(\frac{Y_i - Y_j}{X_i - X_j}\right) \\
&= \mathbb{E}\left(\frac{\beta_0 + \beta_1 X_i + \varepsilon_i - \beta_0 - \beta_1 X_j - \varepsilon_j}{X_i - X_j}\right) \\
&= \mathbb{E}\left(\frac{\beta_1 (X_i - X_j) + \varepsilon_i - \varepsilon_j}{X_i - X_j}\right) \\
&= \beta_1 \mathbb{E}\left(\frac{X_i - X_j}{X_i - X_j}\right) \\
&= \beta_1 \quad \text{(modulo divion by 0)}
\end{aligned}
$$

$$
\begin{aligned}
Var(\hat{\beta}_{ij}) &= Var\left(\frac{Y_i - Y_j}{X_i - X_j}\right) \\
&= Var\left(\frac{\beta_1 (X_i - X_j) + \varepsilon_i - \varepsilon_j}{X_i - X_j}\right) \\
&= \beta_1^2 \cdot \frac{[Var(\varepsilon_i) + Var(\varepsilon_j)]}{(X_i - X_j)^2}
\end{aligned}
$$

- Therefore, $\hat{\beta}_{ij}$ is an unbiased estimator of β_1 with variance essentially proportional to $\frac{1}{(x_i - x_j)^2}$
- A reasonable estimator is then

$$\hat{\beta}_1 := \sum_{i,j: x_i \neq x_j} \lambda_{ij} \hat{\beta}_{ij}$$

  with non-negative coefficients λ_{ij} summing up to 1.
- If the coeffiecients λ_{ij} are taken such that the variance of the estimator is minimized, we obtain the OLS estimator (similarly also for multivariate regression)

Approximate Solution via Interval Analysis

Therefore, the following interval analysis can be applied for an outer approximation of the CDC for the OLS estimator:

$$\underline{\hat{\beta}} := \sum_{i,j:\overline{\Delta_{ij}^x}\neq 0} \lambda_{ij} \frac{\underline{\Delta_{ij}^y}}{\overline{\Delta_{ij}^x}}$$

$$\overline{\hat{\beta}} := \sum_{i,j:\underline{\Delta_{ij}^x}\neq 0} \lambda_{ij} \frac{\overline{\Delta_{ij}^y}}{\underline{\Delta_{ij}^x}}$$

How to choose the λ_{ij} ?

Discussion

- What abot confidence intervals/confidence regions?
- Is the cautious data completion reasonable?
- Is the cautious data completion too cautious?
- Are there other methdos?

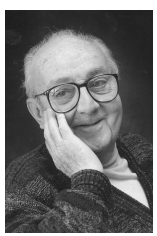
Discussion

- What abot confidence intervals/confidence regions?
- Is the cautious data completion reasonable?
- Is the cautious data completion too cautious?
- Are there other methdos?
  Yes, of course, cf., e.g., Hüllermeier [2014]

(A first look into) Robust Statistics

Do small differences in models matter at all?

- Naturally, every abstraction yields some kind of imprecision.
- Do small differences in models matter at all?
- The mantra of statistical modelling:
  "*Essentially, all models are wrong, but some of them are useful*"
  [Box and Draper, 1987, p. 424]




Are there probability models with
Model 1 "*very similar*" Model 2
BUT
Conclusions (Model 1) "*quite different*" Conclusions (Model 2)?

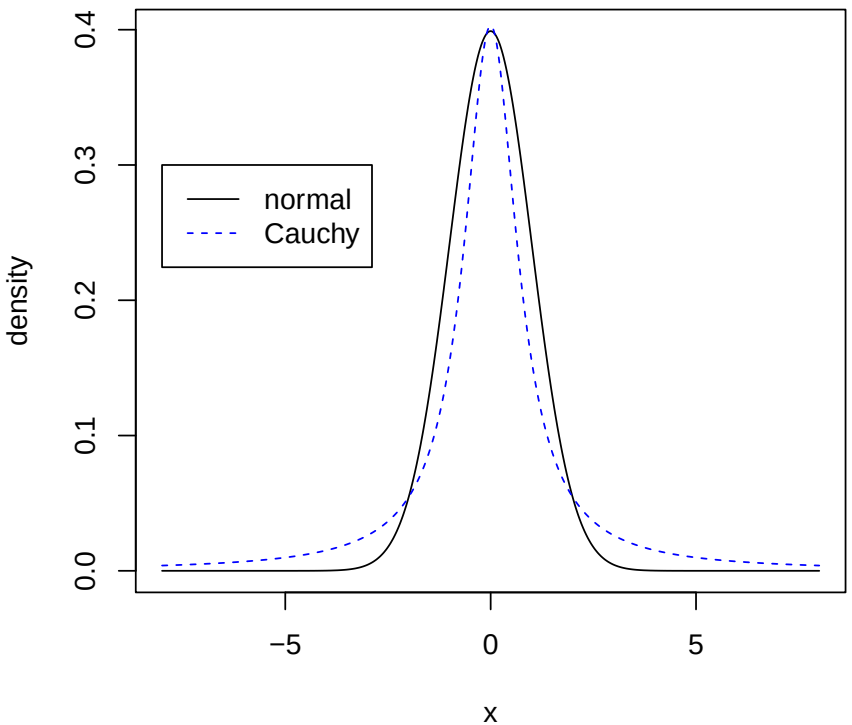
The nonrobustness of classical frequentist methods

Motivating example:

- $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, 1)$

  versus

- $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{C}(\mu, 0.79)$



Estimator for the location parameter?

Estimator for the location parameter: mean?

- $ave(X_1, \ldots, X_n) \sim \mathcal{N}(\mu, \frac{1}{n})$

Estimator for the location parameter: mean?

- $ave(X_1, \ldots, X_n) \sim \mathcal{N}(\mu, \frac{1}{n})$

  versus

- $ave(X_1, \ldots, X_n) \sim$

Estimator for the location parameter: mean?

- $ave(X_1, \ldots, X_n) \sim \mathcal{N}(\mu, \frac{1}{n})$

  versus

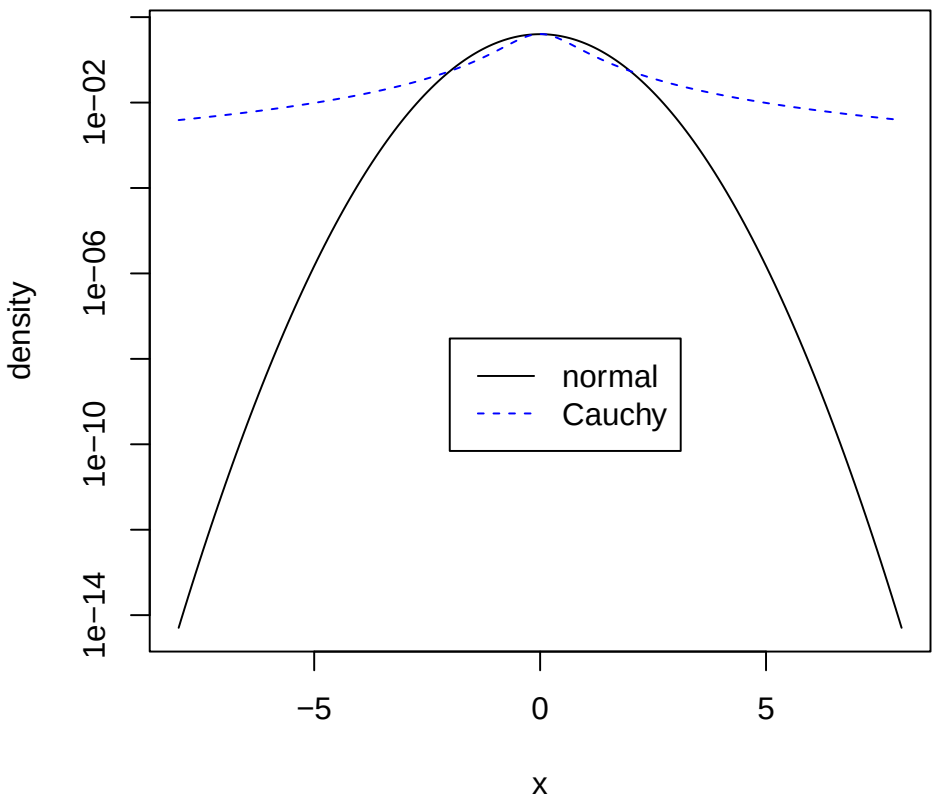
- $ave(X_1, \ldots, X_n) \sim \mathcal{C}(\mu, 0.79)$
- What does this mean?

Estimator for the location parameter: mean?

- $ave(X_1, \ldots, X_n) \sim \mathcal{N}(\mu, \frac{1}{n})$

  versus

- $ave(X_1, \ldots, X_n) \sim \mathcal{C}(\mu, 0.79)$
- What does this mean?
- Cauchy has heavier tails ($\Theta(exp(-x^2)$ vs $\Theta(1/x^2)$)



Robust statistics

Fundamental to the development of robust statistics was the realization that, in parametric models, optimal statistical procedures behave potentially disastrously under minimal deviations from the distributional assumption.

"Insurance interpretation": Therefore, one tries to insure against such effects by developing procedures that are not quite as powerful in the "ideal model", but do not break down under small deviations from the ideal model.

Another look: Jacques Hadamard and well-posed problems

A well-posed problem is a (mathematical) problem satisfying:

- A solution exists
- The solution is unique
- Its behavior changes continuously with the auxiliary conditions, such as initial or boundary values, the input data etc.
- The contrary is called an ill-posed problem

These criteria were first introduced by Jacques Hadamard in 1902

Main theorem of statistics/Glivenko-Cantelli theorem

Let $X_1, \ldots, X_n \overset{i.i.d}{\sim} F$ be a sequence of real-valued random variables with distribution function F. let F_n be the empirical distribution function associated to the sample $(X_1, \ldots X_n)$
(i.e. $F_n(c) := P_n((-\infty, c]) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq c\}}$).

Then

$$\sup_{c \in \mathbb{R}} |F(c) - F_n(c)| \xrightarrow{a.s.} 0.$$

- There are multivariate generalizations.
- There is a whole theory devoted to generalizations: VC theory:

THEORY OF PROBABILITY
AND ITS APPLICATIONS
Volume XVI *Number 2*
1971

ON THE UNIFORM CONVERGENCE OF RELATIVE FREQUENCIES OF EVENTS TO THEIR PROBABILITIES

V. N. VAPNIK AND A. YA. CHERVONENKIS

(*Translated by B. Seckler*)

Introduction

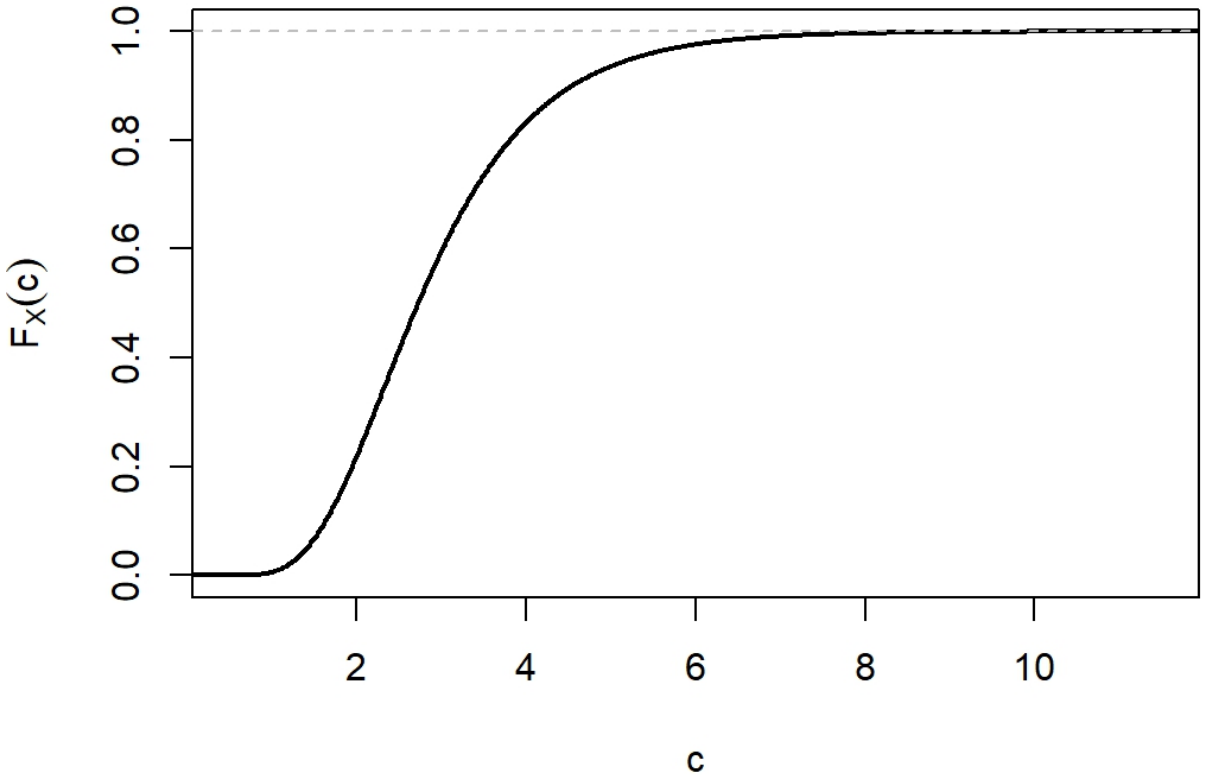
According to the classical Bernoulli theorem, the relative frequency of an event A in a sequence of independent trials converges (in probability) to the probability of that event. In many applications, however, the need arises to judge simultaneously the probabilities of events of an entire class S from one and the same sample. Moreover, it is required that the relative frequency of the events converge to the probability uniformly over the entire class of events S. More precisely, it is required that the probability that the

- The basic assumption is a finite VC dimension of the underlying class of events replacing the events in $P((-\infty, c])$ and $P_n((-\infty, c])$ in the main theorem.

Why the name main theorem?

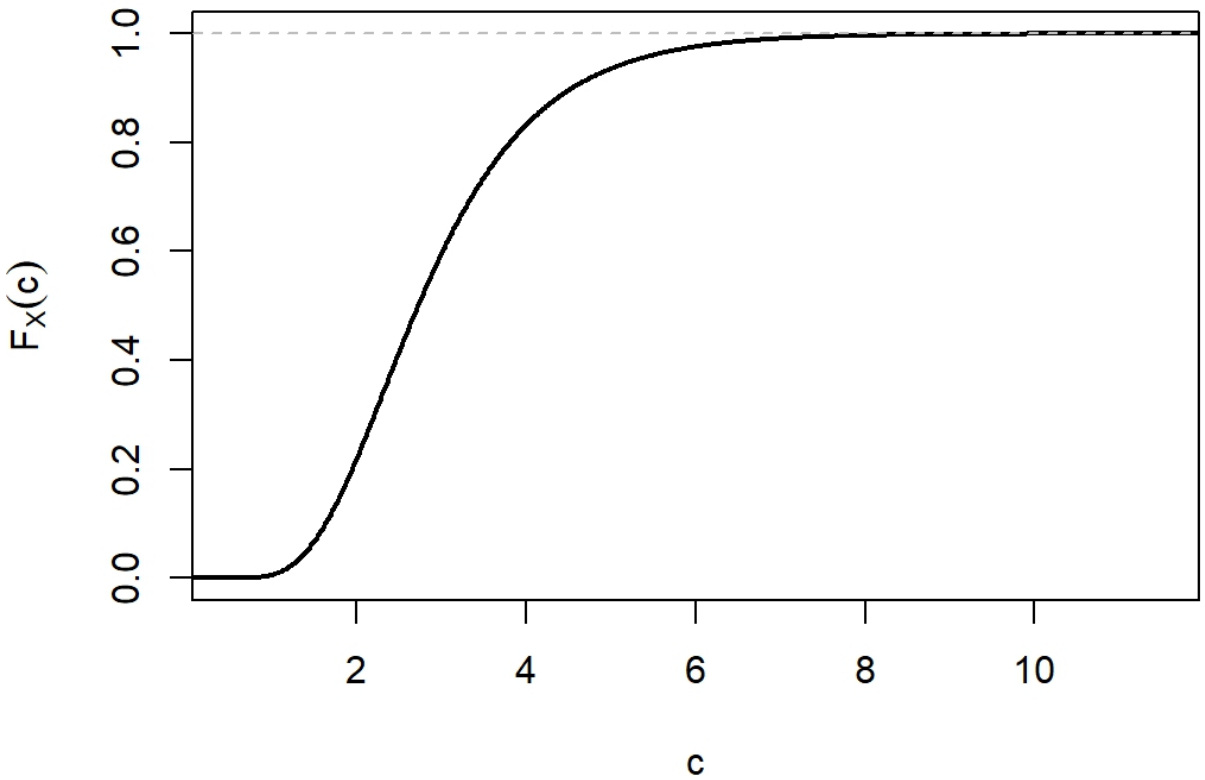
Continuity of some functionals 1: expectation

Let $X \geq 0$. How can we illustrate $\mathbb{E}(X)$ with the help of F_X?



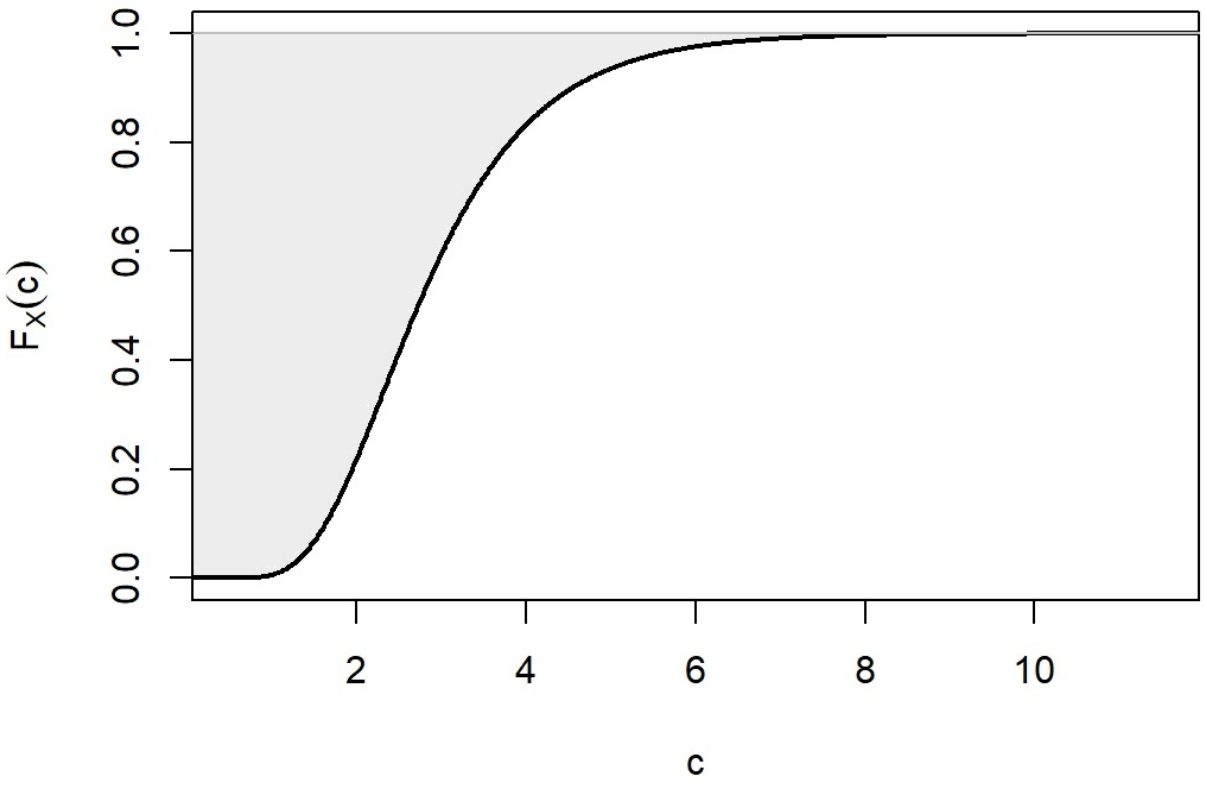
Continuity of some functionals 1: expectation

$$\mathbb{E}(X) = \int_0^\infty 1 - F_X(t)dt$$



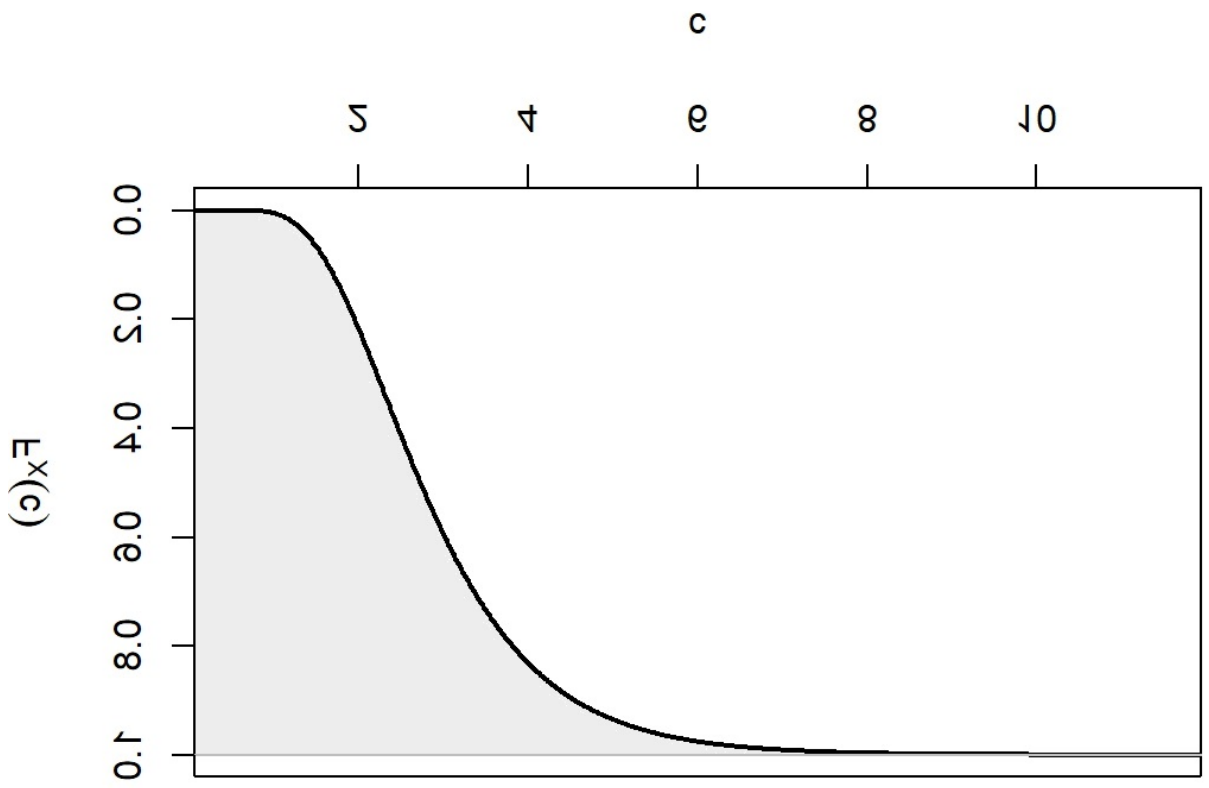
Continuity of some functionals 1: expectation

$$\mathbb{E}(X) = \int_0^\infty 1 - F_X(t)dt$$



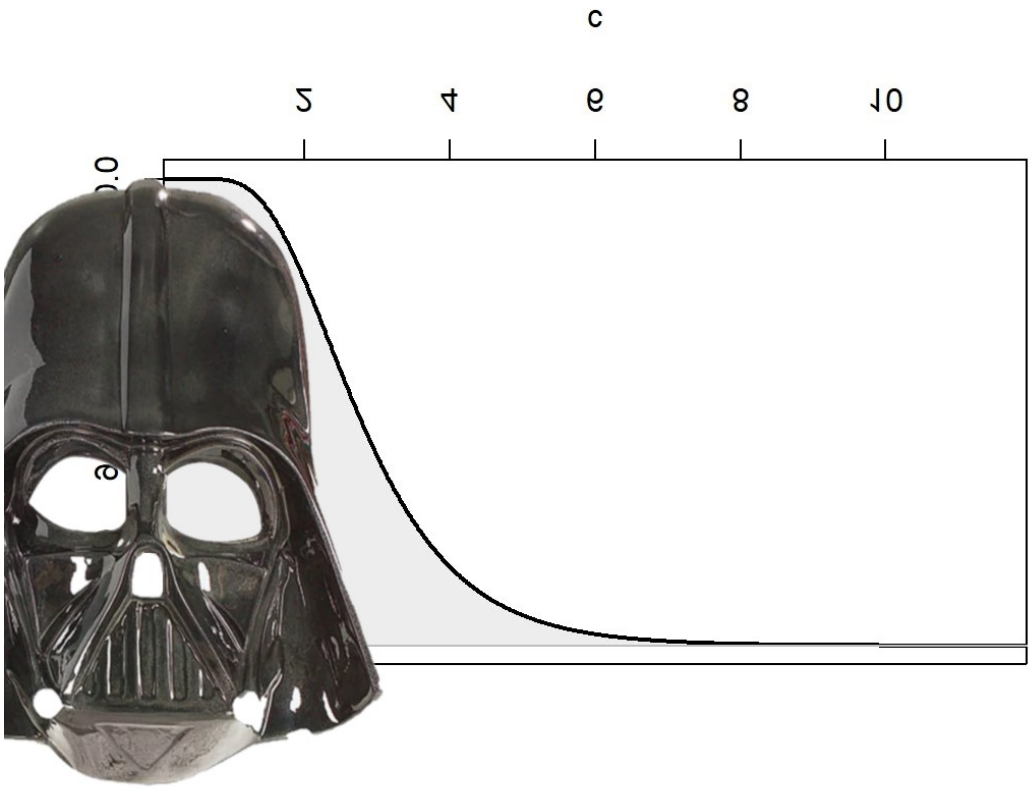
Continuity of some functionals 1: expectation

$$\mathbb{E}(X) = \int_0^\infty 1 - F_X(t)dt = \int_0^\infty S_X(t)dt$$



Continuity of some functionals 1: expectation

$$\mathbb{E}(X) = \int_0^\infty S_X(t)dt$$



"Darth Vader rule"

THE DARTH VADER RULE

PAT MULDOWNEY — KRZYSZTOF OSTASZEWSKI — WOJCIECH WOJDOWSKI

Dedicated to Professor Władysław Wilczyński on his 65th birthday

ABSTRACT. Using Henstock's generalized Riemann integral, we show that, for any almost surely non-negative random variable X with probability density function f_X and survival function $s_X(x) := \int_x^\infty f_X(t)\,dt$, the expected value of X is given by $\mathbf{E}(X) = \int_0^\infty s_X(x)\,dx$.

[1]This is not a reference to a discoverer. But the designation may capture the somewhat counter-intuitive—if not slightly unsettling and surreal—impression which the result can evoke on first encounter.

Exercise

Is the expectation functional

$$F \mapsto \mathbb{E}_F := \int_0^\infty 1 - F(t)dt$$

continuous (w.r.t. the $||\cdot||_\infty$-norm in the domain and the Euclidean norm in the codomain) (think about the grey area)

Exercise: Continuity of some functionals 2: Median

Is the median functional

$$F \mapsto median(F)$$

continuous? (think in terms of the distribution function and statistics I)

Robust Statistics and Imprecise Probabilities

- Replace an ideal statistical model $\{P_\vartheta \mid \vartheta \in \Theta\}$ with precise probabilities P_ϑ with corresponding credal sets $\mathcal{M}_\theta$ consisting of all distributions that are in some sense close to P_θ
- The credal set is described, for example, over probability metrics or by bounds on the densities or the distribution functions.
- A particularly popular model is the ε-contamination model, where one considers for some small $\varepsilon > 0$:

$$\mathcal{M}_{\vartheta,\varepsilon} := \{(1-\varepsilon)P_\vartheta + \varepsilon Q \mid Q \in \mathscr{P}(\Omega, \mathcal{A})\},$$

where $\mathscr{P}(\Omega, \mathcal{A})$ is the set of all probability measures on $(\Omega, \mathcal{A})$.

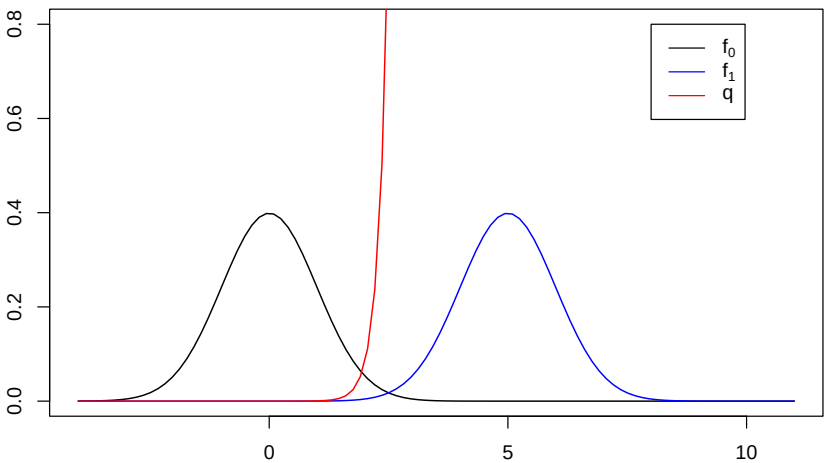

Optimale Tests
bei Intervall-
wahrscheinlichkeit
von
Thomas Augustin
Vandenhoeck & Ruprecht

Background: Neyman-Pearson-Lemma

For testing two statistical hypotheses

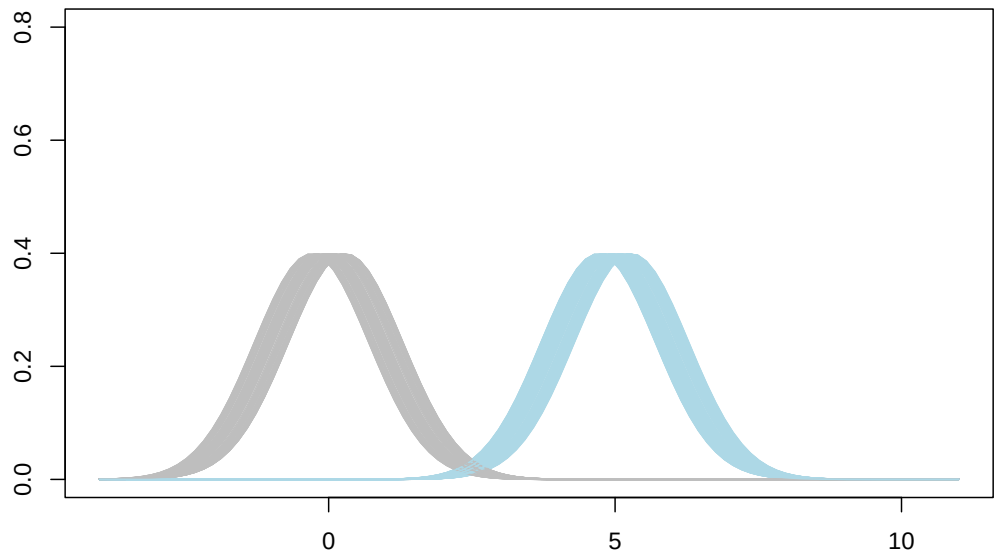
$H_0 : X \sim f_0 \ (\widehat{=} \vartheta_0)$ **versus** $H_1 : X \sim f_1 \ (\widehat{=} \vartheta_1)$

the likelihood-quotient[a] $q(x) = \frac{f_1(x)}{f_0(x)}$ gives an ptimal
Niveau-α-test: reject H_0 if $q(x) > c$ (with $c > 0$ appropriately choosen)

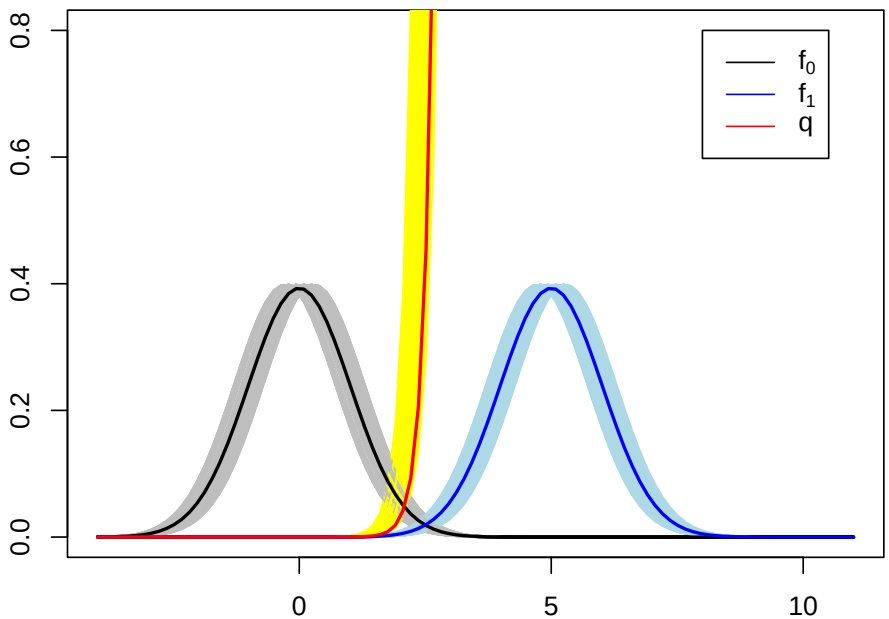


[a] or $(q(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{f_1(x_i)}{f_0(x_i)}$ for i.i.d. sampling)

- What if the situation is too "vague" for specifying two precise probability measures?
- e.g.: test
  H_0 : approximately f_0 **versus** H_1 : approximately
- This is particularly important for adequate statistical modeling (Do not pretend precision that is actually not there)



- One answer for C-probability: Huber - Strassen theory: Idea: instead of testing $\mathcal{M}_0$ against $\mathcal{M}_1$ look at that pair of classical probabilities $p_0 \in \mathcal{M}_0$ and $p_1 \in \mathcal{M}_1$, for which a classical test would be least discriminating
- Such pairs are called **least favorable pairs**




Optimale Tests
bei Intervall-
wahrscheinlichkeit
von
Thomas Augustin
Vandenhoeck & Ruprecht

Summary

- Local methods à la Hadamrd (particularly continuity of functionals)
- Near-local methods (e.g., ε-contamination with ε small)
- Intermediate methods (e.g., ε-contamination with ε medium)
- More "global" methods: Breakdown point Analysis (Hampel (functional analysis style), ...) Example: finite sample breakdown point [Donoho and Gasko, 1992]

$$\varepsilon(T, X^{(n)}) := \min \left\{ \frac{m}{n+m} \;\middle|\; \sup_{Y^{(m)}} ||T(X^{(n)} \cup Y^{(m)}) - T(X^{(n)})|| = \infty \right\}$$

- What next? Non-metric methods ...

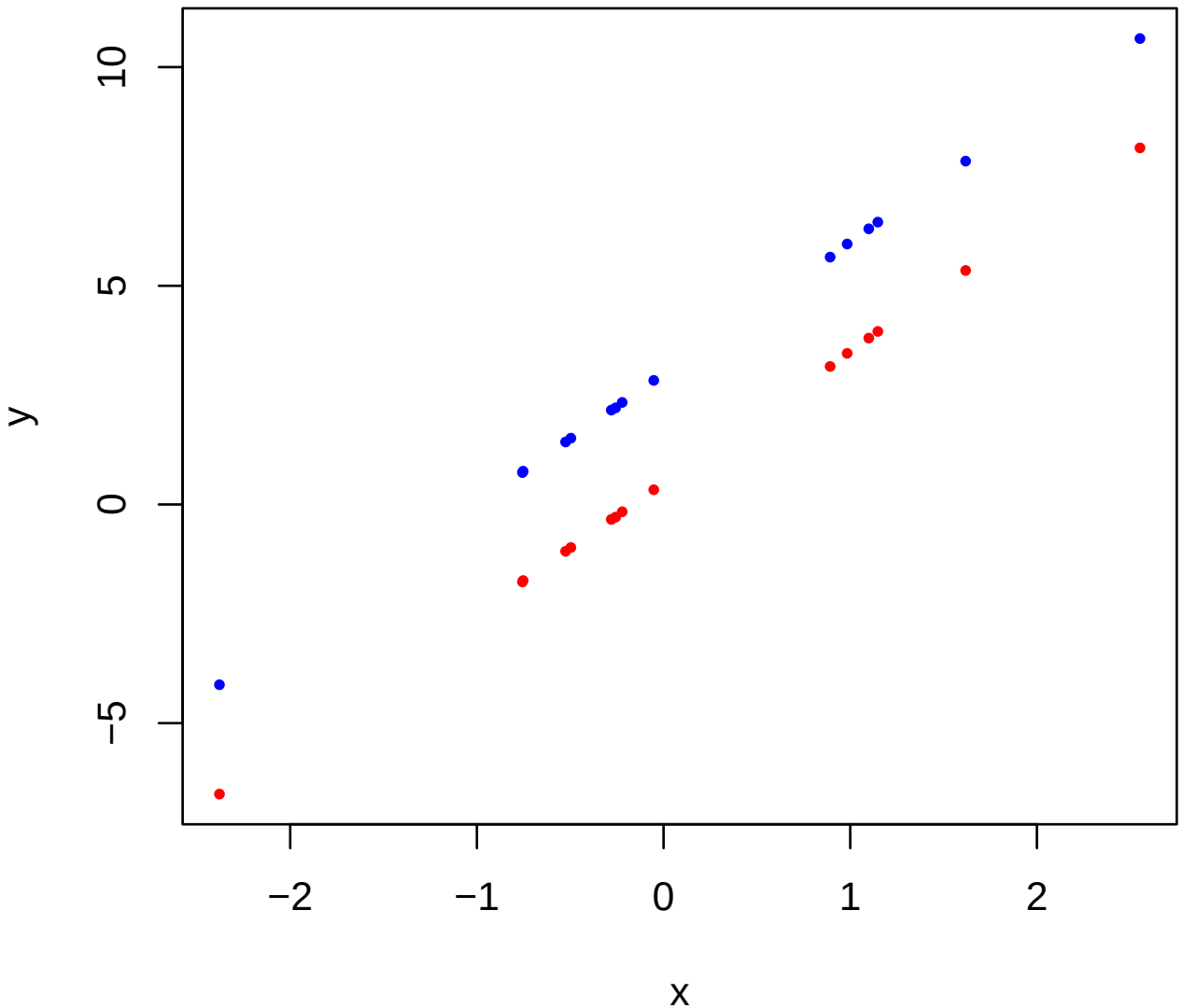
ARTICLE TEMPLATE

On a Notion of Robustness in Formal Concept Analysis: Defining a Contingent Breakdown Point in Non-metric Data Spaces

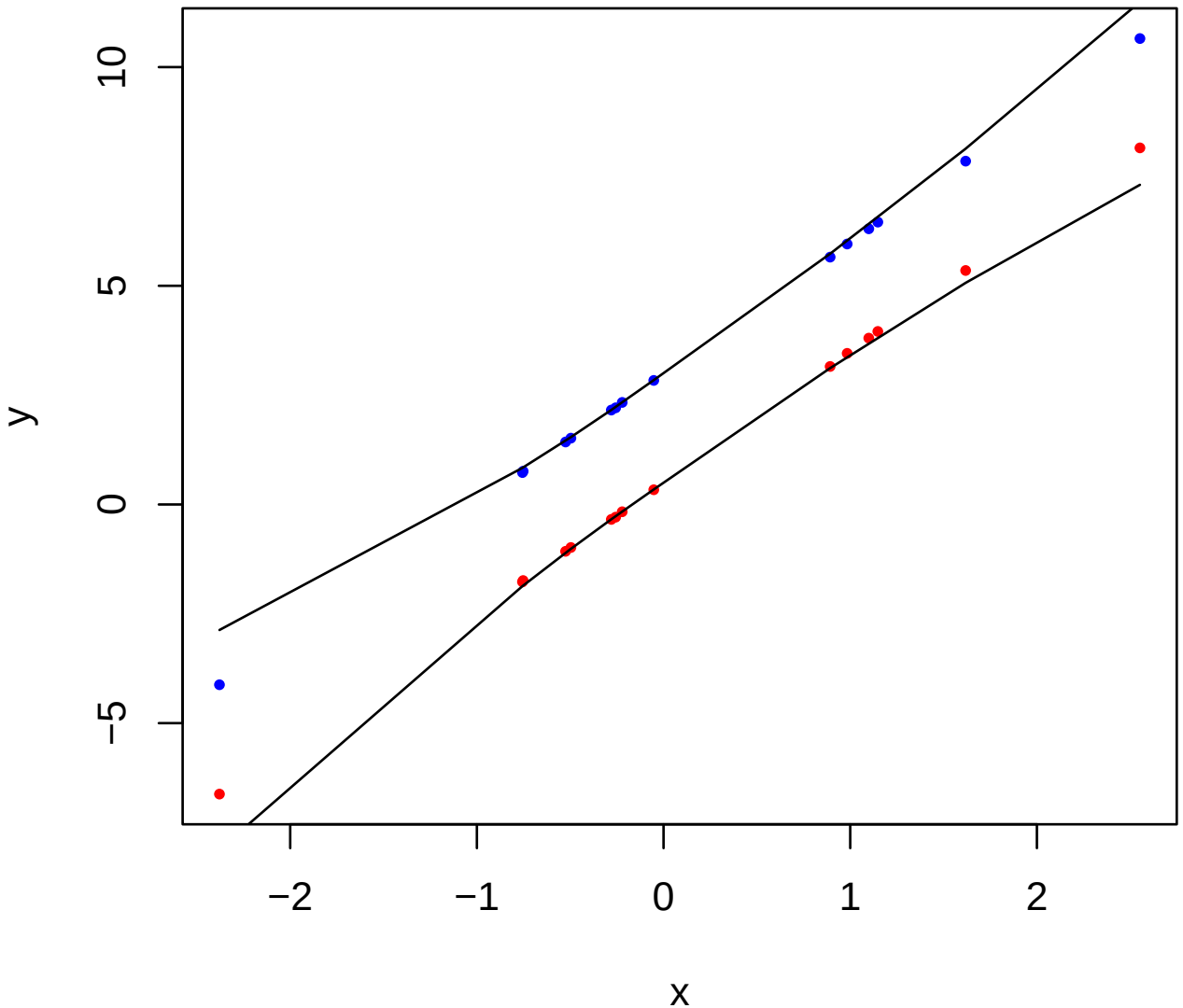
Georg Schollmeyer[a]

[a]Ludwig-Maximilians-Universität München

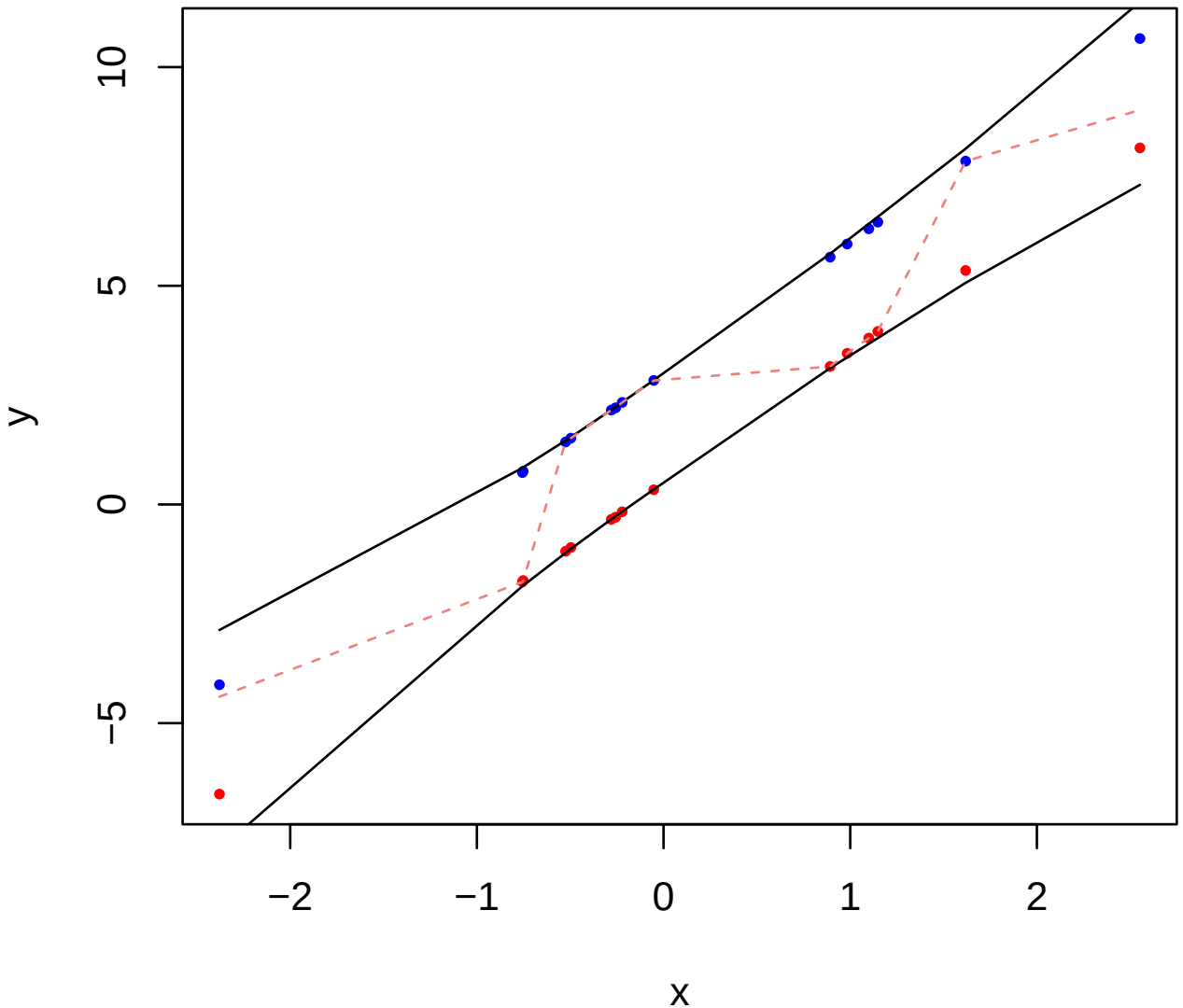
Robustness and Cautious Data Completion



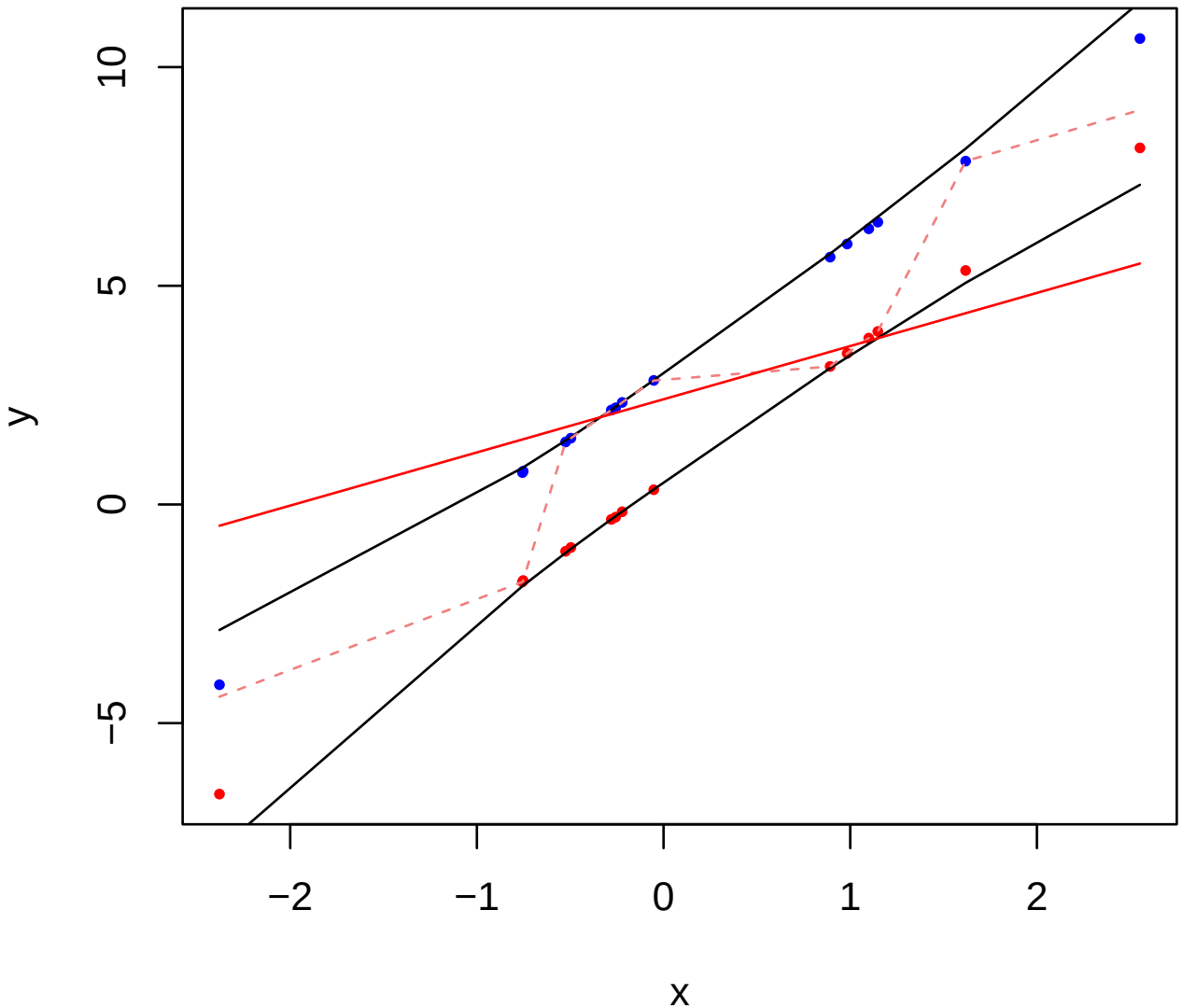
Robustness and Cautious Data Completion



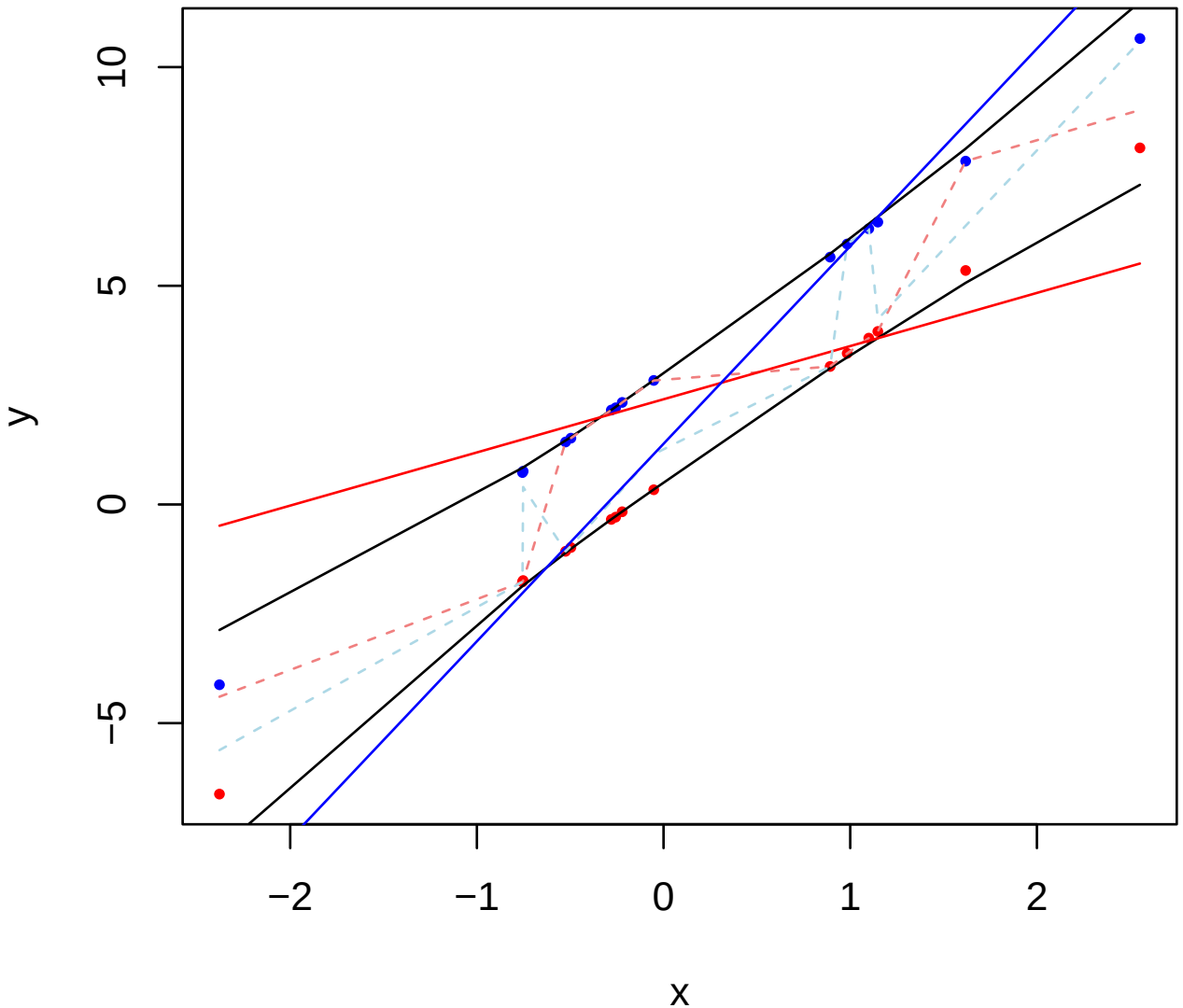
Robustness and Cautious Data Completion



Robustness and Cautious Data Completion



Robustness and Cautious Data Completion



References

T. Augustin, F. P. Coolen, G. De Cooman, and M. C. Troffaes. Introduction to imprecise probabilities. John Wiley & Sons, 2014.

G. E. Box and N. R. Draper. Empirical model-building and response surfaces. John Wiley & Sons, 1987.

D. L. Donoho and M. Gasko. Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness. The Annals of Statistics, 20(4):1803 – 1827, 1992. doi: 10.1214/aos/1176348890. URL https://doi.org/10.1214/aos/1176348890.

R. Frisch. Statistical confluence analysis by means of complete regression systems (1934). The Foundations of Econometric Analysis, 271, 1934.

E. Hüllermeier. Learning from imprecise and fuzzy observations: Data disambiguation through generalized loss minimization. International Journal of Approximate Reasoning, 55(7):1519–1534, 2014. ISSN 0888-613X. doi: https://doi.org/10.1016/j.ijar.2013.