Learning about a Categorical Latent Variable under Prior Near-Ignorance

Alberto Piatti, IDSIA (CH)
Marco Zaffalon, IDSIA (CH)
Fabio Trojani, U.St.Gallen (CH)
Marcus Hutter, ANU (AU)

(Me here)

About IDSIA and myself

Lugano

Università della Svizzera italiana

Scuola universitaria professionale della Svizzera italiana

IDSIA
Istituto Dalle Molle di studi sull'intelligenza artificiale

IDSIA

- Research Institute for AI
- Established in Lugano since 1988
- Since 2000 part of USI and SUPSI
- About 30 people
    - Directors, Seniors, PostDocs, PhDs, . . .

Myself

Myself

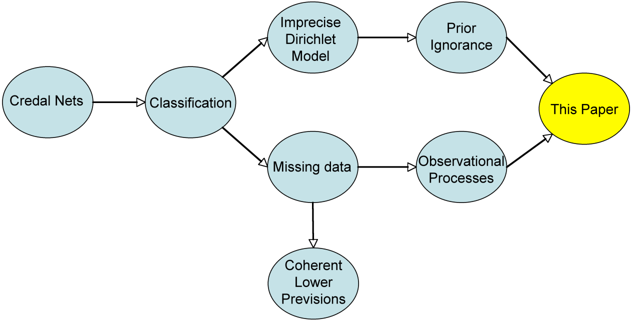
- MSc in Computer Science ('91), PhD in Applied Maths ('97)
- At IDSIA ever since

Myself

- MSc in Computer Science ('91), PhD in Applied Maths ('97)
- At IDSIA ever since
- My interests:

Myself

- MSc in Computer Science ('91), PhD in Applied Maths ('97)
- At IDSIA ever since
- My interests:



The imprecise probability group at IDSIA

-  Alessandro Antonucci, PhD
  - MSc in physics ('02), PhD in computer science ('08)
  - Alg. and theory for credal nets, environmental/military applications
-  Giorgio Corani, PhD
  - MSc in environmental eng. ('99), PhD in information eng. ('05)
  - Data mining, credal classification, dementia application
-  Alberto Piatti, PhD
  - MSc in maths ('01), PhD in finance ('06)
  - Statistics, credal net modeling, military applications
- And 3 new people: two postdocs (Alessio Benavoli, Cassio Polpo de Campos), one PhD student (Yi Sun)
  - Credal nets, data mining, bioinformatics

Background and motivation for this paper

Motivation

Motivation

- Focus on:

Motivation

- Focus on:
  - Personal (or subjective) probability
  - A variable X with values in a finite set $\mathcal{X} = \{x_1, \ldots, x_k\}$
  - An IID process and a predictive setup

Motivation

- Focus on:
  - Personal (or subjective) probability
  - A variable X with values in a finite set $\mathcal{X} = \{x_1, \ldots, x_k\}$
  - An IID process and a predictive setup
- Question:

Motivation

- Focus on:
  - Personal (or subjective) probability
  - A variable X with values in a finite set $\mathcal{X} = \{x_1, \ldots, x_k\}$
  - An IID process and a predictive setup
- Question:

  Can we learn about X from data under prior (*near*-)ignorance?

A bit of background

A bit of background

- Modeling prior ignorance: a long-standing issue in statistics

A bit of background

- Modeling prior ignorance: a long-standing issue in statistics
- The Bayesian viewpoint

A bit of background

- Modeling prior ignorance: a long-standing issue in statistics
- The Bayesian viewpoint
  - Choose a single prior distribution satisfying some principles

A bit of background

- Modeling prior ignorance: a long-standing issue in statistics
- The Bayesian viewpoint
  - Choose a single prior distribution satisfying some principles
  - E.g., symmetry principle+precision $\Rightarrow$ uniform prior

A bit of background

- Modeling prior ignorance: a long-standing issue in statistics
- The Bayesian viewpoint
  - Choose a single prior distribution satisfying some principles
  - E.g., symmetry principle+precision $\Rightarrow$ uniform prior
- Criticism: this appears to model indifference rather than ignorance

A bit of background

- Modeling prior ignorance: a long-standing issue in statistics
- The Bayesian viewpoint
  - Choose a single prior distribution satisfying some principles
  - E.g., symmetry principle+precision $\Rightarrow$ uniform prior
- Criticism: this appears to model indifference rather than ignorance
- More generally speaking:
  - Walley (1991) questions the idea that ignorance can be modeled by a single distribution

A bit of background

A bit of background

- What does Walley (1991, 1996) say then?

A bit of background

- What does Walley (1991, 1996) say then?
  - At least the symmetry and the embedding principles

A bit of background

- What does Walley (1991, 1996) say then?
  - At least the symmetry and the embedding principles
  - Symmetry+embedding ⇒ vacuous probability model for X

A bit of background

- What does Walley (1991, 1996) say then?
  - At least the symmetry and the embedding principles
  - Symmetry+embedding $\Rightarrow$ vacuous probability model for X
    - $\underline{P}(A) = 0, \overline{P}(A) = 1$ for every non-trivial $A \subseteq \mathcal{X}$

A bit of background

- What does Walley (1991, 1996) say then?
  - At least the symmetry and the embedding principles
  - Symmetry+embedding $\Rightarrow$ vacuous probability model for X
    - $\underline{P}(A) = 0, \overline{P}(A) = 1$ for every non-trivial $A \subseteq \mathcal{X}$
  - Note we need imprecise probabilities

A bit of background

- What does Walley (1991, 1996) say then?
  - At least the symmetry and the embedding principles
  - Symmetry+embedding $\Rightarrow$ vacuous probability model for X
    - $\underline{P}(A) = 0, \overline{P}(A) = 1$ for every non-trivial $A \subseteq \mathcal{X}$
  - Note we need imprecise probabilities
- Problem

A bit of background

- What does Walley (1991, 1996) say then?
  - At least the symmetry and the embedding principles
  - Symmetry+embedding $\Rightarrow$ vacuous probability model for X
    - $\underline{P}(A) = 0, \overline{P}(A) = 1$ for every non-trivial $A \subseteq \mathcal{X}$
  - Note we need imprecise probabilities
- Problem
  - Vacuous probability model $\Rightarrow$ vacuous prior beliefs

A bit of background

- What does Walley (1991, 1996) say then?
  - At least the symmetry and the embedding principles
  - Symmetry+embedding $\Rightarrow$ vacuous probability model for X
    - $\underline{P}(A) = 0, \overline{P}(A) = 1$ for every non-trivial $A \subseteq \mathcal{X}$
  - Note we need imprecise probabilities
- Problem
  - Vacuous probability model $\Rightarrow$ vacuous prior beliefs
  - Vacuous prior beliefs $\Rightarrow$ vacuous posterior beliefs (Walley, 1991)

A bit of background

- What does Walley (1991, 1996) say then?
  - At least the symmetry and the embedding principles
  - Symmetry+embedding ⇒ vacuous probability model for X
    - $\underline{P}(A) = 0, \overline{P}(A) = 1$ for every non-trivial $A \subseteq \mathcal{X}$
  - Note we need imprecise probabilities
- Problem
  - Vacuous probability model ⇒ vacuous prior beliefs
  - Vacuous prior beliefs ⇒ vacuous posterior beliefs (Walley, 1991)

⇒ Learning is not possible under prior ignorance!

Near-ignorance

Near-ignorance

- A brilliant idea

Near-ignorance

- A brilliant idea
- Consider the (unknown) *chances* of X

Near-ignorance

- A brilliant idea
- Consider the (unknown) *chances* of X
  - θ_i = physical probability that $X = x_i$

Near-ignorance

- A brilliant idea
- Consider the (unknown) *chances* of X
  - $\theta_i =$ physical probability that $X = x_i$
  - I.e., proportion of individuals in the population with feature x_i

Near-ignorance

- A brilliant idea
- Consider the (unknown) *chances* of X
  - $\theta_i =$ physical probability that $X = x_i$
  - I.e., proportion of individuals in the population with feature x_i
  - $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$

Near-ignorance

- A brilliant idea
- Consider the (unknown) *chances* of X
  - $\theta_i =$ physical probability that $X = x_i$
  - I.e., proportion of individuals in the population with feature x_i
  - $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$
- Choose a non-vacuous prior for θ such that:
  - $\underline{E}(\theta_A) = 0, \overline{E}(\theta_A) = 1$ for every non-trivial $A \subseteq \mathcal{X}$

Near-ignorance

- A brilliant idea
- Consider the (unknown) *chances* of X
  - $\theta_i =$ physical probability that $X = x_i$
  - I.e., proportion of individuals in the population with feature x_i
  - $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$
- Choose a non-vacuous prior for θ such that:
  - $\underline{E}(\theta_A) = 0, \overline{E}(\theta_A) = 1$ for every non-trivial $A \subseteq \mathcal{X}$
  - But $\underline{E}(\theta_A) = \underline{P}(A)$ and $\overline{E}(\theta_A) = \overline{P}(A)$

Near-ignorance

- A brilliant idea
- Consider the (unknown) *chances* of X
  - $\theta_i =$ physical probability that $X = x_i$
  - I.e., proportion of individuals in the population with feature x_i
  - $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$
- Choose a non-vacuous prior for θ such that:
  - $\underline{E}(\theta_A) = 0, \overline{E}(\theta_A) = 1$ for every non-trivial $A \subseteq \mathcal{X}$
  - But $\underline{E}(\theta_A) = \underline{P}(A)$ and $\overline{E}(\theta_A) = \overline{P}(A) \Rightarrow$ Vacuous prior probabilities for X without a vacuous prior for θ

Near-ignorance

- A brilliant idea
- Consider the (unknown) *chances* of X
  - θ_i = physical probability that $X = x_i$
  - I.e., proportion of individuals in the population with feature x_i
  - $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$
- Choose a non-vacuous prior for θ such that:
  - $\underline{E}(\theta_A) = 0, \overline{E}(\theta_A) = 1$ for every non-trivial $A \subseteq \mathcal{X}$
  - But $\underline{E}(\theta_A) = \underline{P}(A)$ and $\overline{E}(\theta_A) = \overline{P}(A) \Rightarrow$ Vacuous prior probabilities for X without a vacuous prior for θ

$\Rightarrow$ This enables us to have symmetry+embedding AND to learn!

Near-ignorance

- A brilliant idea
- Consider the (unknown) *chances* of X
  - $\theta_i =$ physical probability that $X = x_i$
  - I.e., proportion of individuals in the population with feature x_i
  - $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$
- Choose a non-vacuous prior for θ such that:
  - $\underline{E}(\theta_A) = 0, \overline{E}(\theta_A) = 1$ for every non-trivial $A \subseteq \mathcal{X}$
  - But $\underline{E}(\theta_A) = \underline{P}(A)$ and $\overline{E}(\theta_A) = \overline{P}(A) \Rightarrow$ Vacuous prior probabilities for X without a vacuous prior for θ

$\Rightarrow$ This enables us to have symmetry+embedding AND to learn!

$\Rightarrow$ Near-ignorance is a key approach to the problem
(Or at least this is the reason why it is so to some people)

An example: the *imprecise Dirichlet model*

The imprecise Dirichlet model (IDM)

The imprecise Dirichlet model (IDM)

- Walley (1996) has proposed an important model to learn under prior near-ignorance called *imprecise Dirichlet model*
  - Imprecise *Beta* model proposed also by Bernard (1996)

The imprecise Dirichlet model (IDM)

- Walley (1996) has proposed an important model to learn under prior near-ignorance called *imprecise Dirichlet model*
  - Imprecise *Beta* model proposed also by Bernard (1996)
- The IDM generalizes Bayesian learning from multinomial data

IDM: theoretical setting

IDM: theoretical setting

- Infinite population of individuals in k categories: $\mathcal{X} = \{x_1, \ldots, x_k\}$

IDM: theoretical setting

- Infinite population of individuals in k categories: $\mathcal{X} = \{x_1, \ldots, x_k\}$
- Chance θ_i: proportion of units of type x_i in the population

IDM: theoretical setting

- Infinite population of individuals in k categories: $\mathcal{X} = \{x_1, \ldots, x_k\}$
- Chance θ_i: proportion of units of type x_i in the population
- Vector of chances $\theta = (\theta_1, \ldots, \theta_k) \in \Theta$ unknown

IDM: theoretical setting

- Infinite population of individuals in k categories: $\mathcal{X} = \{x_1, \ldots, x_k\}$
- Chance θ_i: proportion of units of type x_i in the population
- Vector of chances $\theta = (\theta_1, \ldots, \theta_k) \in \Theta$ unknown
- Random sample $\mathbf{X} = \mathbf{x}$ from the population

IDM: theoretical setting

- Infinite population of individuals in k categories: $\mathcal{X} = \{x_1, \ldots, x_k\}$
- Chance θ_i: proportion of units of type x_i in the population
- Vector of chances $\theta = (\theta_1, \ldots, \theta_k) \in \Theta$ unknown
- Random sample $\mathbf{X} = \mathbf{x}$ from the population
- Posterior prediction: $P(\mathrm{X} = x_i \,|\, \mathbf{X} = \mathbf{x})$?

In a binomial case

In a binomial case

- People who can be healthy or diseased

In a binomial case

- People who can be healthy or diseased
- $\mathcal{X} = \{h, d\}$

In a binomial case

- People who can be healthy or diseased
- $\mathcal{X} = \{h, d\}$
- θ_1: proportion of *healthy* people in the population

In a binomial case

- People who can be healthy or diseased
- $\mathcal{X} = \{h, d\}$
- θ_1: proportion of *healthy* people in the population
- θ_2: proportion of *diseased* people in the population

In a binomial case

- People who can be healthy or diseased
- $\mathcal{X} = \{h, d\}$
- θ_1: proportion of *healthy* people in the population
- θ_2: proportion of *diseased* people in the population
- No prior knowledge about the proportions

In a binomial case

- People who can be healthy or diseased
- $\mathcal{X} = \{h, d\}$
- θ_1: proportion of *healthy* people in the population
- θ_2: proportion of *diseased* people in the population
- No prior knowledge about the proportions
- Random sample $\mathbf{X} = \mathbf{x}$ of people

In a binomial case

- People who can be healthy or diseased
- $\mathcal{X} = \{h, d\}$
- θ_1: proportion of *healthy* people in the population
- θ_2: proportion of *diseased* people in the population
- No prior knowledge about the proportions
- Random sample $\mathbf{X} = \mathbf{x}$ of people
- Posterior prediction: $P(\mathrm{X} = d \,|\, \mathbf{X} = \mathbf{x})$?

IDM: definition of the model

IDM: definition of the model

- Prior beliefs about θ represented by a set of densities (a credal set):

IDM: definition of the model

- Prior beliefs about θ represented by a set of densities (a credal set):

$$\mathcal{M}_0 = \{dir(s, \mathbf{t}) \mid \mathbf{t} \in \mathcal{T}\},$$

IDM: definition of the model

- Prior beliefs about θ represented by a set of densities (a credal set):

$$\mathcal{M}_0 = \{dir(s, \mathbf{t}) \mid \mathbf{t} \in \mathcal{T}\},$$

  where

IDM: definition of the model

- Prior beliefs about θ represented by a set of densities (a credal set):

$$\mathcal{M}_0 = \{dir(s, \mathbf{t}) \mid \mathbf{t} \in \mathcal{T}\},$$

where

$$dir(s, \mathbf{t}) = \frac{\Gamma(s)}{\Gamma(st_1) \cdots \Gamma(st_k)} \prod_{i=1}^{k} \theta_i^{st_i - 1}$$

is the *Dirichlet density*, with parameters s and $\mathbf{t} = (t_1, \ldots, t_k)$, and

IDM: definition of the model

- Prior beliefs about θ represented by a set of densities (a credal set):

$$\mathcal{M}_0 = \{dir(s, \mathbf{t}) \,|\, \mathbf{t} \in \mathcal{T}\},$$

where

$$dir(s, \mathbf{t}) = \frac{\Gamma(s)}{\Gamma(st_1) \cdots \Gamma(st_k)} \prod_{i=1}^{k} \theta_i^{st_i - 1}$$

is the *Dirichlet density*, with parameters s and $\mathbf{t} = (t_1, \ldots, t_k)$, and

$$\mathcal{T} = \{(t_1, \ldots, t_k) : 0 < t_i < 1, i = 1, \ldots, k, \sum_{i=1}^{k} t_i = 1\}$$

IDM: definition of the model

- Prior beliefs about θ represented by a set of densities (a credal set):

$$\mathcal{M}_0 = \{dir(s, \mathbf{t}) \,|\, \mathbf{t} \in \mathcal{T}\},$$

  where

$$dir(s, \mathbf{t}) = \frac{\Gamma(s)}{\Gamma(st_1) \cdots \Gamma(st_k)} \prod_{i=1}^{k} \theta_i^{st_i - 1}$$

  is the *Dirichlet density*, with parameters s and $\mathbf{t} = (t_1, \ldots, t_k)$, and

$$\mathcal{T} = \{(t_1, \ldots, t_k) : 0 < t_i < 1, i = 1, \ldots, k, \sum_{i=1}^{k} t_i = 1\}$$

- Is $\mathcal{M}_0$ a model of prior near-ignorance?

IDM: prior near-ignorance

IDM: prior near-ignorance

- Remember that $P(\mathrm{X} = x_i) = E(\theta_i)$

IDM: prior near-ignorance

- Remember that $P(\mathrm{X} = x_i) = E(\theta_i)$
- For Dirichlet densities it holds that $E(\theta_i) = t_i$

IDM: prior near-ignorance

- Remember that $P(\mathrm{X} = x_i) = E(\theta_i)$
- For Dirichlet densities it holds that $E(\theta_i) = t_i$
- $\underline{P}(\mathrm{X} = x_i)$

IDM: prior near-ignorance

- Remember that $P(\mathrm{X} = x_i) = E(\theta_i)$
- For Dirichlet densities it holds that $E(\theta_i) = t_i$
- $\underline{P}(\mathrm{X} = x_i) = \inf_{\mathcal{M}_0} E(\theta_i)$

IDM: prior near-ignorance

- Remember that $P(\mathrm{X} = x_i) = E(\theta_i)$
- For Dirichlet densities it holds that $E(\theta_i) = t_i$
- $\underline{P}(\mathrm{X} = x_i) = \inf_{\mathcal{M}_0} E(\theta_i) = \inf_{\mathcal{T}} t_i$

IDM: prior near-ignorance

- Remember that $P(\mathrm{X} = x_i) = E(\theta_i)$
- For Dirichlet densities it holds that $E(\theta_i) = t_i$
- $\underline{P}(\mathrm{X} = x_i) = \inf_{\mathcal{M}_0} E(\theta_i) = \inf_{\mathcal{T}} t_i = 0,$

IDM: prior near-ignorance

- Remember that $P(\mathrm{X} = x_i) = E(\theta_i)$
- For Dirichlet densities it holds that $E(\theta_i) = t_i$
- $\underline{P}(\mathrm{X} = x_i) = \inf_{\mathcal{M}_0} E(\theta_i) = \inf_{\mathcal{T}} t_i = 0,$
- $\overline{P}(\mathrm{X} = x_i)$

IDM: prior near-ignorance

- Remember that $P(\mathrm{X} = x_i) = E(\theta_i)$
- For Dirichlet densities it holds that $E(\theta_i) = t_i$
- $\underline{P}(\mathrm{X} = x_i) = \inf_{\mathcal{M}_0} E(\theta_i) = \inf_{\mathcal{T}} t_i = 0,$
- $\overline{P}(\mathrm{X} = x_i) = \sup_{\mathcal{M}_0} E(\theta_i)$

IDM: prior near-ignorance

- Remember that $P(\mathrm{X} = x_i) = E(\theta_i)$
- For Dirichlet densities it holds that $E(\theta_i) = t_i$
- $\underline{P}(\mathrm{X} = x_i) = \inf_{\mathcal{M}_0} E(\theta_i) = \inf_{\mathcal{T}} t_i = 0,$
- $\overline{P}(\mathrm{X} = x_i) = \sup_{\mathcal{M}_0} E(\theta_i) = \sup_{\mathcal{T}} t_i$

IDM: prior near-ignorance

- Remember that $P(\mathrm{X} = x_i) = E(\theta_i)$
- For Dirichlet densities it holds that $E(\theta_i) = t_i$
- $\underline{P}(\mathrm{X} = x_i) = \inf_{\mathcal{M}_0} E(\theta_i) = \inf_{\mathcal{T}} t_i = 0,$
- $\overline{P}(\mathrm{X} = x_i) = \sup_{\mathcal{M}_0} E(\theta_i) = \sup_{\mathcal{T}} t_i = 1,$

IDM: prior near-ignorance

- Remember that $P(\mathrm{X} = x_i) = E(\theta_i)$
- For Dirichlet densities it holds that $E(\theta_i) = t_i$
- $\underline{P}(\mathrm{X} = x_i) = \inf_{\mathcal{M}_0} E(\theta_i) = \inf_{\mathcal{T}} t_i = 0,$
- $\overline{P}(\mathrm{X} = x_i) = \sup_{\mathcal{M}_0} E(\theta_i) = \sup_{\mathcal{T}} t_i = 1,$

IDM: prior near-ignorance

- Remember that $P(\mathrm{X} = x_i) = E(\theta_i)$
- For Dirichlet densities it holds that $E(\theta_i) = t_i$
- $\underline{P}(\mathrm{X} = x_i) = \inf_{\mathcal{M}_0} E(\theta_i) = \inf_{\mathcal{T}} t_i = 0,$
- $\overline{P}(\mathrm{X} = x_i) = \sup_{\mathcal{M}_0} E(\theta_i) = \sup_{\mathcal{T}} t_i = 1,$

$\Rightarrow$ Vacuous predictive probabilities a priori

IDM: a posteriori

IDM: a posteriori

- Let $\mathbf{X} = \mathbf{x}$ be a sample with counts $\mathbf{a} = (a_1^{\mathbf{x}}, \ldots, a_k^{\mathbf{x}})$

IDM: a posteriori

- Let $\mathbf{X} = \mathbf{x}$ be a sample with counts $\mathbf{a} = (a_1^{\mathbf{x}}, \ldots, a_k^{\mathbf{x}})$
- Posterior inference: each prior updated by Bayes rule

IDM: a posteriori

- Let $\mathbf{X} = \mathbf{x}$ be a sample with counts $\mathbf{a} = (a_1^{\mathbf{x}}, \ldots, a_k^{\mathbf{x}})$
- Posterior inference: each prior updated by Bayes rule
- Posterior credal set:

IDM: a posteriori

- Let $\mathbf{X} = \mathbf{x}$ be a sample with counts $\mathbf{a} = (a_1^{\mathbf{x}}, \ldots, a_k^{\mathbf{x}})$
- Posterior inference: each prior updated by Bayes rule
- Posterior credal set:

$$\mathcal{M}_N = \left\{ dir(N+s, \mathbf{t}^{\mathbf{x}}) \,\middle|\, t_i^{\mathbf{x}} = \frac{a_i^{\mathbf{x}} + st_i}{N+s}, \, \mathbf{t} \in \mathcal{T} \right\}$$

IDM: a posteriori

- Let $\mathbf{X} = \mathbf{x}$ be a sample with counts $\mathbf{a} = (a_1^{\mathbf{x}}, \ldots, a_k^{\mathbf{x}})$
- Posterior inference: each prior updated by Bayes rule
- Posterior credal set:

$$\mathcal{M}_N = \left\{ dir(N+s, \mathbf{t}^{\mathbf{x}}) \,\middle|\, t_i^{\mathbf{x}} = \frac{a_i^{\mathbf{x}} + st_i}{N+s}, \, \mathbf{t} \in \mathcal{T} \right\}$$

- $\underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x})$

IDM: a posteriori

- Let $\mathbf{X} = \mathbf{x}$ be a sample with counts $\mathbf{a} = (a_1^{\mathbf{x}}, \ldots, a_k^{\mathbf{x}})$
- Posterior inference: each prior updated by Bayes rule
- Posterior credal set:

$$\mathcal{M}_N = \left\{ dir(N+s, \mathbf{t}^{\mathbf{x}}) \;\middle|\; t_i^{\mathbf{x}} = \frac{a_i^{\mathbf{x}} + st_i}{N+s}, \, \mathbf{t} \in \mathcal{T} \right\}$$

- $\underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \inf_{\mathcal{M}_N} E(\theta_i \,|\, \mathbf{x})$

IDM: a posteriori

- Let $\mathbf{X} = \mathbf{x}$ be a sample with counts $\mathbf{a} = (a_1^{\mathbf{x}}, \ldots, a_k^{\mathbf{x}})$
- Posterior inference: each prior updated by Bayes rule
- Posterior credal set:

$$\mathcal{M}_N = \left\{ dir(N+s, \mathbf{t}^{\mathbf{x}}) \,\middle|\, t_i^{\mathbf{x}} = \frac{a_i^{\mathbf{x}} + st_i}{N+s}, \mathbf{t} \in \mathcal{T} \right\}$$

- $\underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \inf_{\mathcal{M}_N} E(\theta_i \,|\, \mathbf{x}) = \inf_{\mathcal{T}} \frac{a_i^{\mathbf{x}} + st_i}{N+s}$

IDM: a posteriori

- Let $\mathbf{X} = \mathbf{x}$ be a sample with counts $\mathbf{a} = (a_1^{\mathbf{x}}, \ldots, a_k^{\mathbf{x}})$
- Posterior inference: each prior updated by Bayes rule
- Posterior credal set:

$$\mathcal{M}_N = \left\{ dir(N+s, \mathbf{t}^{\mathbf{x}}) \;\middle|\; t_i^{\mathbf{x}} = \frac{a_i^{\mathbf{x}} + st_i}{N+s}, \, \mathbf{t} \in \mathcal{T} \right\}$$

- $\underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \inf_{\mathcal{M}_N} E(\theta_i \,|\, \mathbf{x}) = \inf_{\mathcal{T}} \frac{a_i^{\mathbf{x}} + st_i}{N+s} = \frac{a_i^{\mathbf{x}}}{N+s}$

IDM: a posteriori

- Let $\mathbf{X} = \mathbf{x}$ be a sample with counts $\mathbf{a} = (a_1^{\mathbf{x}}, \ldots, a_k^{\mathbf{x}})$
- Posterior inference: each prior updated by Bayes rule
- Posterior credal set:

$$\mathcal{M}_N = \left\{ dir(N+s, \mathbf{t}^{\mathbf{x}}) \;\middle|\; t_i^{\mathbf{x}} = \frac{a_i^{\mathbf{x}} + st_i}{N+s}, \, \mathbf{t} \in \mathcal{T} \right\}$$

- $\underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \inf_{\mathcal{M}_N} E(\theta_i \,|\, \mathbf{x}) = \inf_{\mathcal{T}} \frac{a_i^{\mathbf{x}} + st_i}{N+s} = \frac{a_i^{\mathbf{x}}}{N+s}$
- $\overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x})$

IDM: a posteriori

- Let $\mathbf{X} = \mathbf{x}$ be a sample with counts $\mathbf{a} = (a_1^{\mathbf{x}}, \ldots, a_k^{\mathbf{x}})$
- Posterior inference: each prior updated by Bayes rule
- Posterior credal set:

$$\mathcal{M}_N = \left\{ dir(N+s, \mathbf{t}^{\mathbf{x}}) \;\middle|\; t_i^{\mathbf{x}} = \frac{a_i^{\mathbf{x}} + st_i}{N+s}, \mathbf{t} \in \mathcal{T} \right\}$$

- $\underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \inf_{\mathcal{M}_N} E(\theta_i \,|\, \mathbf{x}) = \inf_{\mathcal{T}} \frac{a_i^{\mathbf{x}} + st_i}{N+s} = \frac{a_i^{\mathbf{x}}}{N+s}$
- $\overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \sup_{\mathcal{M}_N} E(\theta_i \,|\, \mathbf{x})$

IDM: a posteriori

- Let $\mathbf{X} = \mathbf{x}$ be a sample with counts $\mathbf{a} = (a_1^{\mathbf{x}}, \ldots, a_k^{\mathbf{x}})$
- Posterior inference: each prior updated by Bayes rule
- Posterior credal set:

$$\mathcal{M}_N = \left\{ dir(N+s, \mathbf{t}^{\mathbf{x}}) \,\middle|\, t_i^{\mathbf{x}} = \frac{a_i^{\mathbf{x}} + st_i}{N+s}, \, \mathbf{t} \in \mathcal{T} \right\}$$

- $\underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \inf_{\mathcal{M}_N} E(\theta_i \,|\, \mathbf{x}) = \inf_{\mathcal{T}} \frac{a_i^{\mathbf{x}} + st_i}{N+s} = \frac{a_i^{\mathbf{x}}}{N+s}$
- $\overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \sup_{\mathcal{M}_N} E(\theta_i \,|\, \mathbf{x}) = \sup_{\mathcal{T}} \frac{a_i^{\mathbf{x}} + st_i}{N+s}$

IDM: a posteriori

- Let $\mathbf{X} = \mathbf{x}$ be a sample with counts $\mathbf{a} = (a_1^{\mathbf{x}}, \ldots, a_k^{\mathbf{x}})$
- Posterior inference: each prior updated by Bayes rule
- Posterior credal set:

$$\mathcal{M}_N = \left\{ dir(N+s, \mathbf{t}^{\mathbf{x}}) \,\middle|\, t_i^{\mathbf{x}} = \frac{a_i^{\mathbf{x}} + st_i}{N+s}, \, \mathbf{t} \in \mathcal{T} \right\}$$

- $\underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \inf_{\mathcal{M}_N} E(\theta_i \,|\, \mathbf{x}) = \inf_{\mathcal{T}} \frac{a_i^{\mathbf{x}} + st_i}{N+s} = \frac{a_i^{\mathbf{x}}}{N+s}$
- $\overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \sup_{\mathcal{M}_N} E(\theta_i \,|\, \mathbf{x}) = \sup_{\mathcal{T}} \frac{a_i^{\mathbf{x}} + st_i}{N+s} = \frac{a_i^{\mathbf{x}} + s}{N+s}$

IDM: a posteriori

- Let $\mathbf{X} = \mathbf{x}$ be a sample with counts $\mathbf{a} = (a_1^{\mathbf{x}}, \dots, a_k^{\mathbf{x}})$
- Posterior inference: each prior updated by Bayes rule
- Posterior credal set:

$$\mathcal{M}_N = \left\{ dir(N+s, \mathbf{t}^{\mathbf{x}}) \,\middle|\, t_i^{\mathbf{x}} = \frac{a_i^{\mathbf{x}} + st_i}{N+s}, \mathbf{t} \in \mathcal{T} \right\}$$

- $\underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \inf_{\mathcal{M}_N} E(\theta_i \,|\, \mathbf{x}) = \inf_{\mathcal{T}} \frac{a_i^{\mathbf{x}} + st_i}{N+s} = \frac{a_i^{\mathbf{x}}}{N+s}$
- $\overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \sup_{\mathcal{M}_N} E(\theta_i \,|\, \mathbf{x}) = \sup_{\mathcal{T}} \frac{a_i^{\mathbf{x}} + st_i}{N+s} = \frac{a_i^{\mathbf{x}} + s}{N+s}$
- How much ignorance left a posteriori?

IDM: degree of imprecision

IDM: degree of imprecision

- Degree of imprecision:

IDM: degree of imprecision

- Degree of imprecision:

$$\Delta(\mathrm{X} = x_i \,|\, \mathbf{x}) = \overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) - \underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x})$$

IDM: degree of imprecision

- Degree of imprecision:

$$\Delta(\mathrm{X} = x_i \,|\, \mathbf{x}) = \overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) - \underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \frac{s}{N + s}$$

IDM: degree of imprecision

- Degree of imprecision:

$$\Delta(\mathrm{X} = x_i \,|\, \mathbf{x}) = \overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) - \underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \frac{s}{N+s}$$

- Implications:

IDM: degree of imprecision

- Degree of imprecision:

$$\Delta(\mathrm{X} = x_i \,|\, \mathbf{x}) = \overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) - \underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \frac{s}{N+s}$$

- Implications:
  - $\Delta(\mathrm{X} = x_i \,|\, \mathbf{x})$ decreases with increasing N:

IDM: degree of imprecision

- Degree of imprecision:

$$\Delta(\mathrm{X} = x_i \,|\, \mathbf{x}) = \overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) - \underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \frac{s}{N+s}$$

- Implications:
  - $\Delta(\mathrm{X} = x_i \,|\, \mathbf{x})$ decreases with increasing N:

    Learning under prior near-ignorance is possible!

IDM: degree of imprecision

- Degree of imprecision:

$$\Delta(\mathrm{X} = x_i \,|\, \mathbf{x}) = \overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) - \underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \frac{s}{N + s}$$

- Implications:
  - $\Delta(\mathrm{X} = x_i \,|\, \mathbf{x})$ decreases with increasing N:

    Learning under prior near-ignorance is possible!
  - Large values of $s \Rightarrow$ decreases slowly

IDM: degree of imprecision

- Degree of imprecision:

$$\Delta(\mathrm{X} = x_i \,|\, \mathbf{x}) = \overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) - \underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \frac{s}{N+s}$$

- Implications:
  - $\Delta(\mathrm{X} = x_i \,|\, \mathbf{x})$ decreases with increasing N:

    Learning under prior near-ignorance is possible!
  - Large values of $s \Rightarrow$ decreases slowly
  - Small values of $s \Rightarrow$ decreases fast

IDM: degree of imprecision

- Degree of imprecision:

$$\Delta(\mathrm{X} = x_i \,|\, \mathbf{x}) = \overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) - \underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \frac{s}{N + s}$$

- Implications:
  - $\Delta(\mathrm{X} = x_i \,|\, \mathbf{x})$ decreases with increasing N:

    Learning under prior near-ignorance is possible!

  - Large values of $s \Rightarrow$ decreases slowly
  - Small values of $s \Rightarrow$ decreases fast
  - s: caution of inference

IDM: degree of imprecision

- Degree of imprecision:

$$\Delta(\mathrm{X} = x_i \,|\, \mathbf{x}) = \overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) - \underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \frac{s}{N+s}$$

- Implications:
  - $\Delta(\mathrm{X} = x_i \,|\, \mathbf{x})$ decreases with increasing N:

    Learning under prior near-ignorance is possible!
  - Large values of $s \Rightarrow$ decreases slowly
  - Small values of $s \Rightarrow$ decreases fast
  - s: caution of inference
    - Inference cautious enough to encompass several precise models

IDM: degree of imprecision

- Degree of imprecision:

$$\Delta(\mathrm{X} = x_i \,|\, \mathbf{x}) = \overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) - \underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \frac{s}{N+s}$$

- Implications:
  - $\Delta(\mathrm{X} = x_i \,|\, \mathbf{x})$ decreases with increasing N:

    Learning under prior near-ignorance is possible!
  - Large values of $s \Rightarrow$ decreases slowly
  - Small values of $s \Rightarrow$ decreases fast
  - s: caution of inference
    - Inference cautious enough to encompass several precise models
    - Support for $1 \leq s \leq 2$

IDM: degree of imprecision

- Degree of imprecision:

$$\Delta(\mathrm{X} = x_i \,|\, \mathbf{x}) = \overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) - \underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \frac{s}{N + s}$$

- Implications:
  - $\Delta(\mathrm{X} = x_i \,|\, \mathbf{x})$ decreases with increasing N:

    Learning under prior near-ignorance is possible!
  - Large values of $s \Rightarrow$ decreases slowly
  - Small values of $s \Rightarrow$ decreases fast
  - s: caution of inference
    - Inference cautious enough to encompass several precise models
    - Support for $1 \leq s \leq 2$
    - For a discussion see Bernard (2005) or Walley (1996)

IDM: degree of imprecision

- Degree of imprecision:

$$\Delta(\mathrm{X} = x_i \,|\, \mathbf{x}) = \overline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) - \underline{P}(\mathrm{X} = x_i \,|\, \mathbf{x}) = \frac{s}{N + s}$$

- Implications:
  - $\Delta(\mathrm{X} = x_i \,|\, \mathbf{x})$ decreases with increasing N:

    Learning under prior near-ignorance is possible!
  - Large values of $s \Rightarrow$ decreases slowly
  - Small values of $s \Rightarrow$ decreases fast
  - s: caution of inference
    - Inference cautious enough to encompass several precise models
    - Support for $1 \leq s \leq 2$
    - For a discussion see Bernard (2005) or Walley (1996)
  - $N \to \infty$ then $\Delta(\mathrm{X} = x_i \,|\, \mathbf{x}) \to 0$, $P(\mathrm{X} = x_i \,|\, \mathbf{x}) \to \frac{a_i^{\mathbf{x}}}{N}$

IDM: example

IDM: example

- Sample $\mathbf{x}$ contains 33 *healthy* and 4 *diseased* people

IDM: example

- Sample $\mathbf{x}$ contains 33 *healthy* and 4 *diseased* people
- With $s = 2$:

IDM: example

- Sample $\mathbf{x}$ contains 33 *healthy* and 4 *diseased* people
- With $s = 2$:

$$\underline{P}(\mathrm{X} = h \,|\, \mathbf{x}) = \frac{33}{37 + 2} \cong 0.85$$

IDM: example

- Sample $\mathbf{x}$ contains 33 *healthy* and 4 *diseased* people
- With $s = 2$:

$$\underline{P}(\mathrm{X} = h \,|\, \mathbf{x}) = \frac{33}{37 + 2} \cong 0.85$$

$$\overline{P}(\mathrm{X} = h \,|\, \mathbf{x}) = \frac{33 + 2}{37 + 2} \cong 0.90$$

IDM: example

- Sample $\mathbf{x}$ contains 33 *healthy* and 4 *diseased* people
- With $s = 2$:

$$\underline{P}(\mathrm{X} = h \,|\, \mathbf{x}) = \frac{33}{37+2} \cong 0.85$$

$$\overline{P}(\mathrm{X} = h \,|\, \mathbf{x}) = \frac{33+2}{37+2} \cong 0.90$$

- Analogously,

IDM: example

- Sample $\mathbf{x}$ contains 33 *healthy* and 4 *diseased* people
- With $s = 2$:

$$\underline{P}(\mathrm{X} = h \,|\, \mathbf{x}) = \frac{33}{37+2} \cong 0.85$$

$$\overline{P}(\mathrm{X} = h \,|\, \mathbf{x}) = \frac{33+2}{37+2} \cong 0.90$$

- Analogously,

$$\underline{P}(\mathrm{X} = d \,|\, \mathbf{x}) = \frac{4}{37+2} \cong 0.10$$

IDM: example

- Sample $\mathbf{x}$ contains 33 *healthy* and 4 *diseased* people
- With $s = 2$:

$$\underline{P}(\mathrm{X} = h \,|\, \mathbf{x}) = \frac{33}{37+2} \cong 0.85$$

$$\overline{P}(\mathrm{X} = h \,|\, \mathbf{x}) = \frac{33+2}{37+2} \cong 0.90$$

- Analogously,

$$\underline{P}(\mathrm{X} = d \,|\, \mathbf{x}) = \frac{4}{37+2} \cong 0.10$$

$$\overline{P}(\mathrm{X} = d \,|\, \mathbf{x}) = \frac{4+2}{37+2} \cong 0.15$$

Back to the general discussion; main result

Latent and manifest variables

Latent and manifest variables

- As before, X denotes whether a person is healthy or diseased, i.e., $\mathcal{X} = \{h, d\}$

Latent and manifest variables

- As before, X denotes whether a person is healthy or diseased, i.e., $\mathcal{X} = \{h, d\}$
- Something that we might have overlooked before:
  Will we actually observe X?

Latent and manifest variables

- As before, X denotes whether a person is healthy or diseased, i.e., $\mathcal{X} = \{h, d\}$
- Something that we might have overlooked before:
  Will we actually observe X?
  Perhaps not very realistic

Latent and manifest variables

- As before, X denotes whether a person is healthy or diseased, i.e., $\mathcal{X} = \{h, d\}$
- Something that we might have overlooked before:
  Will we actually observe X?
  Perhaps not very realistic
- What will we be likely to observe instead?

Latent and manifest variables

- As before, X denotes whether a person is healthy or diseased, i.e., $\mathcal{X} = \{h, d\}$
- Something that we might have overlooked before:
  Will we actually observe X?
  Perhaps not very realistic
- What will we be likely to observe instead?
  - E.g., the outcome of a medical test related to X

Latent and manifest variables

- As before, X denotes whether a person is healthy or diseased, i.e., $\mathcal{X} = \{h, d\}$
- Something that we might have overlooked before:
  Will we actually observe X?
  Perhaps not very realistic
- What will we be likely to observe instead?
  - E.g., the outcome of a medical test related to X
  - Call it S (for signal), another random variable

Latent and manifest variables

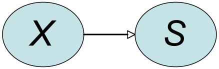
- As before, X denotes whether a person is healthy or diseased, i.e., $\mathcal{X} = \{h, d\}$
- Something that we might have overlooked before:
  Will we actually observe X?
  Perhaps not very realistic
- What will we be likely to observe instead?
  - E.g., the outcome of a medical test related to X
  - Call it S (for signal), another random variable
- X is said to be a latent variable

Latent and manifest variables

- As before, X denotes whether a person is healthy or diseased, i.e., $\mathcal{X} = \{h, d\}$
- Something that we might have overlooked before:
  Will we actually observe X?
  Perhaps not very realistic
- What will we be likely to observe instead?
  - E.g., the outcome of a medical test related to X
  - Call it S (for signal), another random variable
- X is said to be a latent variable
- S is said to be manifest

Latent and manifest variables

- As before, X denotes whether a person is healthy or diseased, i.e., $\mathcal{X} = \{h, d\}$
- Something that we might have overlooked before:
  Will we actually observe X?
  Perhaps not very realistic
- What will we be likely to observe instead?
  - E.g., the outcome of a medical test related to X
  - Call it S (for signal), another random variable
- X is said to be a latent variable
- S is said to be manifest
- The process that takes X in input and outputs S is said to be the observational process (or measurement process)



Ubiquity of latent variables

Ubiquity of latent variables

- Now let X denote the outcome of tossing a coin, i.e., $\mathcal{X} = \{h, t\}$

Ubiquity of latent variables

- Now let X denote the outcome of tossing a coin, i.e., $\mathcal{X} = \{h, t\}$
- Will we actually (be certain to) observe X?

Ubiquity of latent variables

- Now let X denote the outcome of tossing a coin, i.e., $\mathcal{X} = \{h, t\}$
- Will we actually (be certain to) observe X?
  Perhaps not very realistic either ...

Ubiquity of latent variables

- Now let X denote the outcome of tossing a coin, i.e., $\mathcal{X} = \{h, t\}$
- Will we actually (be certain to) observe X?
  Perhaps not very realistic either ...
  - To observe something we must employ a "mechanism"
    - Even only our eyes

Ubiquity of latent variables

- Now let X denote the outcome of tossing a coin, i.e., $\mathcal{X} = \{h, t\}$
- Will we actually (be certain to) observe X?
  Perhaps not very realistic either ...
  - To observe something we must employ a "mechanism"
    - Even only our eyes
  - Every mechanism appears to be imperfect in practice

Ubiquity of latent variables

- Now let X denote the outcome of tossing a coin, i.e., $\mathcal{X} = \{h, t\}$
- Will we actually (be certain to) observe X?
  Perhaps not very realistic either ...
  - To observe something we must employ a "mechanism"
    - Even only our eyes
  - Every mechanism appears to be imperfect in practice
    - I.e., once in a while we might confuse h with t

Ubiquity of latent variables

- Now let X denote the outcome of tossing a coin, i.e., $\mathcal{X} = \{h, t\}$
- Will we actually (be certain to) observe X?
  Perhaps not very realistic either ...
  - To observe something we must employ a “mechanism”
    - Even only our eyes
  - Every mechanism appears to be imperfect in practice
    - I.e., once in a while we might confuse h with t
    - Imperfection may well be tiny

Ubiquity of latent variables

- Now let X denote the outcome of tossing a coin, i.e., $\mathcal{X} = \{h, t\}$
- Will we actually (be certain to) observe X?
  Perhaps not very realistic either . . .
  - To observe something we must employ a "mechanism"
    - Even only our eyes
  - Every mechanism appears to be imperfect in practice
    - I.e., once in a while we might confuse h with t
    - Imperfection may well be tiny
- Strictly speaking, the observation of X is another variable, say S
  - S takes values in $\mathcal{S} = \mathcal{X} = \{h, t\}$

Ubiquity of latent variables

- Now let X denote the outcome of tossing a coin, i.e., $\mathcal{X} = \{h, t\}$
- Will we actually (be certain to) observe X?
  Perhaps not very realistic either ...
  - To observe something we must employ a "mechanism"
    - Even only our eyes
  - Every mechanism appears to be imperfect in practice
    - I.e., once in a while we might confuse h with t
    - Imperfection may well be tiny
- Strictly speaking, the observation of X is another variable, say S
  - S takes values in $\mathcal{S} = \mathcal{X} = \{h, t\}$
  - X latent, S manifest

Ubiquity of latent variables

- Now let X denote the outcome of tossing a coin, i.e., $\mathcal{X} = \{h, t\}$
- Will we actually (be certain to) observe X?
  Perhaps not very realistic either ...
  - To observe something we must employ a "mechanism"
    - Even only our eyes
  - Every mechanism appears to be imperfect in practice
    - I.e., once in a while we might confuse h with t
    - Imperfection may well be tiny
- Strictly speaking, the observation of X is another variable, say S
  - S takes values in $\mathcal{S} = \mathcal{X} = \{h, t\}$
  - X latent, S manifest

$\Rightarrow$ Latent variables appear to arise as soon as we make observations

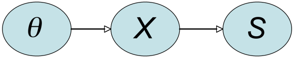
Data generation

Data generation

- The overall process of data generation:

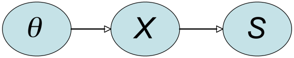
Data generation

- The overall process of data generation:



Data generation

- The overall process of data generation:



  - Note we are assuming that θ is not relevant to S once we know X

Main result

Main result

- Say that

Main result

- Say that
    - Your probability for $X = x_i$ is vacuous a priori

Main result

- Say that
  - Your probability for $X = x_i$ is vacuous a priori
    - I.e., you use an arbitrary set of priors $\mathcal{M}_0$ to model near-ignorance

Main result

- Say that
  - Your probability for $X = x_i$ is vacuous a priori
    - I.e., you use an arbitrary set of priors $\mathcal{M}_0$ to model near-ignorance
  - Your data are collected in a vector $\mathbf{s} = (s_1, \ldots, s_N)$ of signals

Main result

- Say that
  - Your probability for $X = x_i$ is vacuous a priori
    - I.e., you use an arbitrary set of priors $\mathcal{M}_0$ to model near-ignorance
  - Your data are collected in a vector $\mathbf{s} = (s_1, \ldots, s_N)$ of signals
    - Either discrete or continuous

Main result

- Say that
  - Your probability for $X = x_i$ is vacuous a priori
    - I.e., you use an arbitrary set of priors $\mathcal{M}_0$ to model near-ignorance
  - Your data are collected in a vector $\mathbf{s} = (s_1, \ldots, s_N)$ of signals
    - Either discrete or continuous
    - Related in whatever (but known) way to N realizations of X

Main result

- Say that
  - Your probability for $X = x_i$ is vacuous a priori
    - I.e., you use an arbitrary set of priors $\mathcal{M}_0$ to model near-ignorance
  - Your data are collected in a vector $\mathbf{s} = (s_1, \ldots, s_N)$ of signals
    - Either discrete or continuous
    - Related in whatever (but known) way to N realizations of X
- If
  - The likelihood $P(\mathbf{s}|\theta)$ is continuous and positive

Main result

- Say that
  - Your probability for $X = x_i$ is vacuous a priori
    - I.e., you use an arbitrary set of priors $\mathcal{M}_0$ to model near-ignorance
  - Your data are collected in a vector $\mathbf{s} = (s_1, \ldots, s_N)$ of signals
    - Either discrete or continuous
    - Related in whatever (but known) way to N realizations of X
- If
  - The likelihood $P(\mathbf{s}|\theta)$ is continuous and positive
- Then
  - Your probability for $X = x_i$ is vacuous a posteriori!

Main result

- Say that
  - Your probability for $X = x_i$ is vacuous a priori
    - I.e., you use an arbitrary set of priors $\mathcal{M}_0$ to model near-ignorance
  - Your data are collected in a vector $\mathbf{s} = (s_1, \ldots, s_N)$ of signals
    - Either discrete or continuous
    - Related in whatever (but known) way to N realizations of X
- If
  - The likelihood $P(\mathbf{s}|\theta)$ is continuous and positive
- Then
  - Your probability for $X = x_i$ is vacuous a posteriori!

$\Rightarrow$ Learning is not possible under prior near-ignorance!

Main result

- Say that
  - Your probability for $X = x_i$ is vacuous a priori
    - I.e., you use an arbitrary set of priors $\mathcal{M}_0$ to model near-ignorance
  - Your data are collected in a vector $\mathbf{s} = (s_1, \ldots, s_N)$ of signals
    - Either discrete or continuous
    - Related in whatever (but known) way to N realizations of X
- If
  - The likelihood $P(\mathbf{s}|\theta)$ is continuous and positive
- Then
  - Your probability for $X = x_i$ is vacuous a posteriori!

  $\Rightarrow$ Learning is not possible under prior near-ignorance!

- Analogous result if we wish to predict the next N' units

Some insight: core result about deterministic chances

Some insight: core result about deterministic chances

- Denote with $\mathbf{e_i}$ the i-th vector of deterministic chances:

$$\mathbf{e_i} = (0, \ldots, 0, \underbrace{1}_{i\text{-th place}}, 0, \ldots, 0).$$

Some insight: core result about deterministic chances

- Denote with $\mathbf{e_i}$ the i-th vector of deterministic chances:

$$\mathbf{e_i} = (0, \ldots, 0, \underbrace{1}_{i\text{-th place}}, 0, \ldots, 0).$$

- $P(\mathbf{S} = \mathbf{s} \mid \theta)$ continuous in θ

Some insight: core result about deterministic chances

- Denote with $\mathbf{e_i}$ the i-th vector of deterministic chances:

$$\mathbf{e_i} = (0, \ldots, 0, \underbrace{1}_{i\text{-th place}}, 0, \ldots, 0).$$

- $P(\mathbf{S} = \mathbf{s} \,|\, \theta)$ continuous in θ
- If $P(\mathbf{S} = \mathbf{s} \,|\, \mathbf{e_i}) > 0$, then

Some insight: core result about deterministic chances

- Denote with $\mathbf{e_i}$ the i-th vector of deterministic chances:

$$\mathbf{e_i} = (0, \ldots, 0, \underbrace{1}_{i\text{-th place}}, 0, \ldots, 0).$$

- $P(\mathbf{S} = \mathbf{s} \,|\, \theta)$ continuous in θ
- If $P(\mathbf{S} = \mathbf{s} \,|\, \mathbf{e_i}) > 0$, then

$$\overline{P}(\mathrm{X} = x_i \,|\, \mathbf{S} = \mathbf{s}) = 1$$

and consequently,

Some insight: core result about deterministic chances

- Denote with $\mathbf{e_i}$ the i-th vector of deterministic chances:

$$\mathbf{e_i} = (0, \ldots, 0, \underbrace{1}_{i\text{-th place}}, 0, \ldots, 0).$$

- $P(\mathbf{S} = \mathbf{s} \,|\, \theta)$ continuous in θ
- If $P(\mathbf{S} = \mathbf{s} \,|\, \mathbf{e_i}) > 0$, then

$$\overline{P}(\mathrm{X} = x_i \,|\, \mathbf{S} = \mathbf{s}) = 1$$

  and consequently,

$$\underline{P}(\mathrm{X} = x_j \,|\, \mathbf{S} = \mathbf{s}) = 0$$

  for all $j \neq i$

Are our results pointing to “pathological” situations?

Are our results pointing to “pathological” situations?

- Much probably not
- The condition on the likelihood seems to be very frequently satisfied

Are our results pointing to “pathological” situations?

- Much probably not
- The condition on the likelihood seems to be very frequently satisfied
- Medical example again

Are our results pointing to "pathological" situations?

- Much probably not
- The condition on the likelihood seems to be very frequently satisfied
- Medical example again
  - X is the health status, in $\mathcal{X} = \{h, d\}$
  - S is the outcome of a medical test, in $\mathcal{S} = \{-, +\}$

Are our results pointing to "pathological" situations?

- Much probably not
- The condition on the likelihood seems to be very frequently satisfied
- Medical example again
  - X is the health status, in $\mathcal{X} = \{h, d\}$
  - S is the outcome of a medical test, in $\mathcal{S} = \{-, +\}$
  - Say that the test may be wrong once in a while:
    - $P(-|d) = \epsilon'$, $P(+|h) = \epsilon''$: $\epsilon', \epsilon'' > 0$

Are our results pointing to "pathological" situations?

- Much probably not
- The condition on the likelihood seems to be very frequently satisfied
- Medical example again
  - X is the health status, in $\mathcal{X} = \{h, d\}$
  - S is the outcome of a medical test, in $\mathcal{S} = \{-, +\}$
  - Say that the test may be wrong once in a while:
    - $P(-|d) = \epsilon'$, $P(+|h) = \epsilon''$: $\epsilon', \epsilon'' > 0$
    - The opposite is not very realistic

Are our results pointing to “pathological” situations?

- Much probably not
- The condition on the likelihood seems to be very frequently satisfied
- Medical example again
  - X is the health status, in $\mathcal{X} = \{h, d\}$
  - S is the outcome of a medical test, in $\mathcal{S} = \{-, +\}$
  - Say that the test may be wrong once in a while:
    - $P(-|d) = \epsilon'$, $P(+|h) = \epsilon''$: $\epsilon', \epsilon'' > 0$
    - The opposite is not very realistic
- Then the likelihood is continuous and everywhere positive
  - Learning is not possible

Are our results pointing to “pathological” situations?

- Much probably not
- The condition on the likelihood seems to be very frequently satisfied
- Medical example again
  - X is the health status, in $\mathcal{X} = \{h, d\}$
  - S is the outcome of a medical test, in $\mathcal{S} = \{-, +\}$
  - Say that the test may be wrong once in a while:
    - $P(-|d) = \epsilon'$, $P(+|h) = \epsilon''$: $\epsilon', \epsilon'' > 0$
    - The opposite is not very realistic
- Then the likelihood is continuous and everywhere positive
  - Learning is not possible
- Note that ϵ', ϵ'' can be arbitrarily small (yet positive)

Are our results pointing to "pathological" situations?

- Much probably not
- The condition on the likelihood seems to be very frequently satisfied
- Medical example again
  - X is the health status, in $\mathcal{X} = \{h, d\}$
  - S is the outcome of a medical test, in $\mathcal{S} = \{-, +\}$
  - Say that the test may be wrong once in a while:
    - $P(-|d) = \epsilon'$, $P(+|h) = \epsilon''$: $\epsilon', \epsilon'' > 0$
    - The opposite is not very realistic
- Then the likelihood is continuous and everywhere positive
  - Learning is not possible
- Note that ϵ', ϵ'' can be arbitrarily small (yet positive)

$\Rightarrow$ We cannot neglect the observational process,
however tiny the imperfection!

Some intuition based on the special case of the IDM

A simple setup based on the IDM

A simple setup based on the IDM

- X and S taking values from $\mathcal{X}$

A simple setup based on the IDM

- X and S taking values from $\mathcal{X}$
- Random sample $\mathbf{s}$

A simple setup based on the IDM

- X and S taking values from $\mathcal{X}$
- Random sample s
- Imperfect observational mechanism modeled by emission matrix Λ:
    - $\lambda_{ij} = P(S = x_i \,|\, \mathrm{X} = x_j)$ known for all i, j

A simple setup based on the IDM

- X and S taking values from $\mathcal{X}$
- Random sample s
- Imperfect observational mechanism modeled by emission matrix Λ:
  - $\lambda_{ij} = P(S = x_i \,|\, \mathrm{X} = x_j)$ known for all i, j
- If $\lambda_{ij} > 0$ for all i, j then we cannot learn from s

Intuitive explanation

Intuitive explanation

- Consider an investigator with precise beliefs $t_i \to 1$
(Similar arguments can be used when $t_i \to 0$)

Intuitive explanation

- Consider an investigator with precise beliefs $t_i \to 1$
  (Similar arguments can be used when $t_i \to 0$)
- Belief a priori: population completely of individuals of type x_i

Intuitive explanation

- Consider an investigator with precise beliefs $t_i \to 1$
  (Similar arguments can be used when $t_i \to 0$)
- Belief a priori: population completely of individuals of type x_i
- The investigator observes $S = x_j$ and knows that $\lambda_{ji} > 0$

Intuitive explanation

- Consider an investigator with precise beliefs $t_i \to 1$
  (Similar arguments can be used when $t_i \to 0$)
- Belief a priori: population completely of individuals of type x_i
- The investigator observes $S = x_j$ and knows that $\lambda_{ji} > 0$
- Then he thinks that an error has occurred

Intuitive explanation

- Consider an investigator with precise beliefs $t_i \to 1$
  (Similar arguments can be used when $t_i \to 0$)
- Belief a priori: population completely of individuals of type x_i
- The investigator observes $S = x_j$ and knows that $\lambda_{ji} > 0$
- Then he thinks that an error has occurred
- This degenerate behavior is impossible only if $\lambda_{ji} = 0$

Intuitive explanation

- Consider an investigator with precise beliefs $t_i \to 1$
  (Similar arguments can be used when $t_i \to 0$)
- Belief a priori: population completely of individuals of type x_i
- The investigator observes $S = x_j$ and knows that $\lambda_{ji} > 0$
- Then he thinks that an error has occurred
- This degenerate behavior is impossible only if $\lambda_{ji} = 0$

$\Rightarrow$ to learn at least partial perfection is needed

Intuitive explanation

- Consider an investigator with precise beliefs $t_i \to 1$
  (Similar arguments can be used when $t_i \to 0$)
- Belief a priori: population completely of individuals of type x_i
- The investigator observes $S = x_j$ and knows that $\lambda_{ji} > 0$
- Then he thinks that an error has occurred
- This degenerate behavior is impossible only if $\lambda_{ji} = 0$

  $\Rightarrow$ to learn at least partial perfection is needed
- The IDM encompasses the case $t_i \to 1$, from which the problem

Degeneracy on the border

Degeneracy on the border

- Consider the following example:

Degeneracy on the border

- Consider the following example:
  - IDM in the binomial case (with caution parameter s=2)

Degeneracy on the border

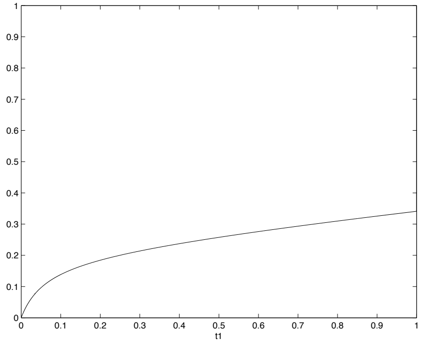
- Consider the following example:
  - IDM in the binomial case (with caution parameter s=2)
  - Counts for heads and tails (variable S): $(12, 23)$

Degeneracy on the border

- Consider the following example:
  - IDM in the binomial case (with caution parameter s=2)
  - Counts for heads and tails (variable S): $(12, 23)$
  - Probability of confusing heads with tails and vice versa equal to 0.2

Degeneracy on the border

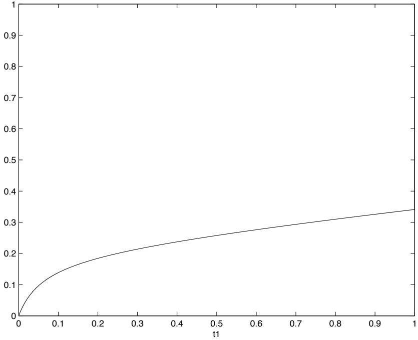
- Consider the following example:
  - IDM in the binomial case (with caution parameter s=2)
  - Counts for heads and tails (variable S): $(12, 23)$
  - Probability of confusing heads with tails and vice versa equal to 0.2



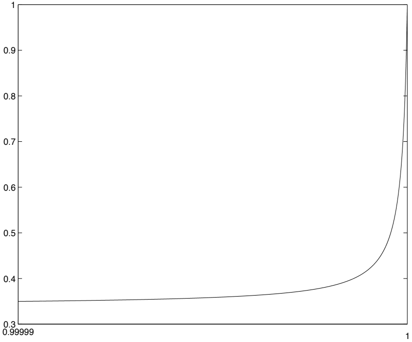
$P(heads|\mathbf{s})$ as function of t_1

Degeneracy on the border

- Consider the following example:
  - IDM in the binomial case (with caution parameter s=2)
  - Counts for heads and tails (variable S): $(12, 23)$
  - Probability of confusing heads with tails and vice versa equal to 0.2



$P(heads|\mathbf{s})$ as function of t_1



Again, but $0.99999 < t_1 < 1$

Extreme priors: how to "solve" the problem (and yet losing near-ignorance)

Extreme priors: how to “solve” the problem (and yet losing near-ignorance)

- The previous discussion suggests a possible solution of the problem

Extreme priors: how to "solve" the problem (and yet losing near-ignorance)

- The previous discussion suggests a possible solution of the problem
- Vacuous probabilities arise because of the presence, in $\mathcal{M}_0$, of extreme priors, arbitrarily close to the degenerate ones

Extreme priors: how to "solve" the problem (and yet losing near-ignorance)

- The previous discussion suggests a possible solution of the problem
- Vacuous probabilities arise because of the presence, in $\mathcal{M}_0$, of extreme priors, arbitrarily close to the degenerate ones
- Given a small probability of errors, a slight restriction of $\mathcal{M}_0$ avoids the problem

Extreme priors: how to "solve" the problem (and yet losing near-ignorance)

- The previous discussion suggests a possible solution of the problem
- Vacuous probabilities arise because of the presence, in $\mathcal{M}_0$, of extreme priors, arbitrarily close to the degenerate ones
- Given a small probability of errors, a slight restriction of $\mathcal{M}_0$ avoids the problem
    - E.g., in the IDM set $\epsilon < t_i < 1 - \epsilon$ for all i and small positive ϵ

Extreme priors: how to "solve" the problem (and yet losing near-ignorance)

- The previous discussion suggests a possible solution of the problem
- Vacuous probabilities arise because of the presence, in $\mathcal{M}_0$, of extreme priors, arbitrarily close to the degenerate ones
- Given a small probability of errors, a slight restriction of $\mathcal{M}_0$ avoids the problem
    - E.g., in the IDM set $\epsilon < t_i < 1 - \epsilon$ for all i and small positive ϵ
    - Yet, we lose near-ignorance

Extreme priors: how to "solve" the problem (and yet losing near-ignorance)

- The previous discussion suggests a possible solution of the problem
- Vacuous probabilities arise because of the presence, in $\mathcal{M}_0$, of extreme priors, arbitrarily close to the degenerate ones
- Given a small probability of errors, a slight restriction of $\mathcal{M}_0$ avoids the problem
  - E.g., in the IDM set $\epsilon < t_i < 1 - \epsilon$ for all i and small positive ϵ
  - Yet, we lose near-ignorance
  - Moreover: how can we justify the choice of a certain value ϵ?

A note on manifest variables

A note on manifest variables

- So far we have focused on predicting latent values

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches
- A "clean" approach:

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches
- A "clean" approach:
  - We admit that there is a latent and a manifest level

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches
- A "clean" approach:
  - We admit that there is a latent and a manifest level
  - Yet, we want both to learn and make predictions on the manifest level only

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches
- A "clean" approach:
  - We admit that there is a latent and a manifest level
  - Yet, we want both to learn and make predictions on the manifest level only
  - Then

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches
- A "clean" approach:
  - We admit that there is a latent and a manifest level
  - Yet, we want both to learn and make predictions on the manifest level only
  - Then
    - We are obliged to apply a modified IDM

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches
- A "clean" approach:
  - We admit that there is a latent and a manifest level
  - Yet, we want both to learn and make predictions on the manifest level only
  - Then
    - We are obliged to apply a modified IDM
    - The learning problem pops up again, this time in predicting S

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches
- A "clean" approach:
  - We admit that there is a latent and a manifest level
  - Yet, we want both to learn and make predictions on the manifest level only
  - Then
    - We are obliged to apply a modified IDM
    - The learning problem pops up again, this time in predicting S
- A "dirty" approach:

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches
- A "clean" approach:
  - We admit that there is a latent and a manifest level
  - Yet, we want both to learn and make predictions on the manifest level only
  - Then
    - We are obliged to apply a modified IDM
    - The learning problem pops up again, this time in predicting S
- A "dirty" approach:
  - We *pretend* that there is no latent level

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches
- A "clean" approach:
  - We admit that there is a latent and a manifest level
  - Yet, we want both to learn and make predictions on the manifest level only
  - Then
    - We are obliged to apply a modified IDM
    - The learning problem pops up again, this time in predicting S
- A "dirty" approach:
  - We *pretend* that there is no latent level
  - I.e., that observations are perfect even if we know they are not (This is what all the people who use the IDM in practice do)

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches
- A "clean" approach:
  - We admit that there is a latent and a manifest level
  - Yet, we want both to learn and make predictions on the manifest level only
  - Then
    - We are obliged to apply a modified IDM
    - The learning problem pops up again, this time in predicting S
- A "dirty" approach:
  - We *pretend* that there is no latent level
  - I.e., that observations are perfect even if we know they are not (This is what all the people who use the IDM in practice do)
  - Then

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches
- A "clean" approach:
  - We admit that there is a latent and a manifest level
  - Yet, we want both to learn and make predictions on the manifest level only
  - Then
    - We are obliged to apply a modified IDM
    - The learning problem pops up again, this time in predicting S
- A "dirty" approach:
  - We *pretend* that there is no latent level
  - I.e., that observations are perfect even if we know they are not (This is what all the people who use the IDM in practice do)
  - Then
    - We can apply the usual IDM

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches
- A "clean" approach:
  - We admit that there is a latent and a manifest level
  - Yet, we want both to learn and make predictions on the manifest level only
  - Then
    - We are obliged to apply a modified IDM
    - The learning problem pops up again, this time in predicting S
- A "dirty" approach:
  - We *pretend* that there is no latent level
  - I.e., that observations are perfect even if we know they are not (This is what all the people who use the IDM in practice do)
  - Then
    - We can apply the usual IDM
    - Learning is possible

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches
- A “clean” approach:
  - We admit that there is a latent and a manifest level
  - Yet, we want both to learn and make predictions on the manifest level only
  - Then
    - We are obliged to apply a modified IDM
    - The learning problem pops up again, this time in predicting S
- A “dirty” approach:
  - We *pretend* that there is no latent level
  - I.e., that observations are perfect even if we know they are not (This is what all the people who use the IDM in practice do)
  - Then
    - We can apply the usual IDM
    - Learning is possible
    - Empirically works very well, yet theoretically is unexplored

A note on manifest variables

- So far we have focused on predicting latent values
  - I.e., X rather than S
- One might say that focusing on manifest values is more reasonable
  - Looks reasonable especially when the imperfection is tiny
- Two approaches
- A "clean" approach:
  - We admit that there is a latent and a manifest level
  - Yet, we want both to learn and make predictions on the manifest level only
  - Then
    - We are obliged to apply a modified IDM
    - The learning problem pops up again, this time in predicting S
- A "dirty" approach:
  - We *pretend* that there is no latent level
  - I.e., that observations are perfect even if we know they are not (This is what all the people who use the IDM in practice do)
  - Then
    - We can apply the usual IDM
    - Learning is possible
    - Empirically works very well, yet theoretically is unexplored
- This is a strange paradox still to be solved

Discussion

Discussion

- Initial question

  Can we learn about X from data under prior near-ignorance?

Discussion

- Initial question

  Can we learn about X from data under prior near-ignorance?

- Our answer (in the current setup):

Discussion

- Initial question

  Can we learn about X from data under prior near-ignorance?

- Our answer (in the current setup): not very likely when X is latent

Discussion

- Initial question

  Can we learn about X from data under prior near-ignorance?

- Our answer (in the current setup): not very likely when X is latent
  - ... and X appears to be latent most of the times

Discussion

- Initial question

  Can we learn about X from data under prior near-ignorance?
- Our answer (in the current setup): not very likely when X is latent
  - ... and X appears to be latent most of the times
- What to do?

Discussion

- Initial question

  Can we learn about X from data under prior near-ignorance?

- Our answer (in the current setup): not very likely when X is latent
  - ... and X appears to be latent most of the times
- What to do?

  Optimist Search for ways out that may preserve learning+near-ignorance

Discussion

- Initial question

  Can we learn about X from data under prior near-ignorance?

- Our answer (in the current setup): not very likely when X is latent
  - ... and X appears to be latent most of the times
- What to do?

  Optimist Search for ways out that may preserve learning+near-ignorance

  Skeptic
  - Perhaps forget about the possibility of adopting near-ignorance
  - Wonder whether symmetry+embedding can be really met in practice
  - Study forms of weak beliefs that are stronger than near-ignorance

Discussion

- Initial question

  Can we learn about X from data under prior near-ignorance?

- Our answer (in the current setup): not very likely when X is latent
  - ... and X appears to be latent most of the times
- What to do?

  Optimist: Search for ways out that may preserve learning+near-ignorance

  Skeptic:
  - Perhaps forget about the possibility of adopting near-ignorance
  - Wonder whether symmetry+embedding can be really met in practice
  - Study forms of weak beliefs that are stronger than near-ignorance
- Or (perhaps easier) show that our results are wrong!