Coarse categorical data under epistemic and ontologic uncertainty

Comparison and extensions of some approaches

Julia Plaß

17th of June 2013

Outline

1. Introduction to the problem of coarse data
2. Approaches for dealing with epistemic uncertainty
3. Distribution on the power set as an approach for dealing with ontologic uncertainty
4. A multinomial logit model based approach under the second kind of uncertainty
   - Accounting for epistemic uncertainty
   - Accounting for ontologic uncertainty
5. Conclusion

Introduction to the problem of coarsened data

Two kinds of uncertainty

- First kind of uncertainty: Sampling variability
- Second kind of uncertainty: Lack of information

$\Downarrow$ **Epistemic uncertainy**

$\Downarrow$ **Ontologic uncertainty**

Epistemic uncertainty

- Imprecise observation of something precise
- True underlying coarsening mechanism available
- Different types of coarsening
  - Rounding
  - Grouping
  - Heaping
  - Censoring
  - ...

Epistemic uncertainty

- Imprecise observation of something precise
- True underlying coarsening mechanism available
- Different types of coarsening
  - Rounding e.g. needed time to go to university
  - Grouping
  - Heaping
  - Censoring
  - ...

Epistemic uncertainty

- Imprecise observation of something precise
- True underlying coarsening mechanism available
- Different types of coarsening
  - Rounding
  - Grouping
  - Heaping
  - Censoring
  - ...

classes of wages per month for working students:
$[0, 200), [200, 400), [400, 600), \ldots$

Epistemic uncertainty

- Imprecise observation of something precise
- True underlying coarsening mechanism available
- Different types of coarsening
  - Rounding — e.g. Age heaping
  - Grouping — G=0: true age truncated to the next lowest month,
  - Heaping — G=1 to the next lowest half year,
  - Censoring — G=2 to the next lowest year
  - ...

Epistemic uncertainty

- Imprecise observation of something precise
- True underlying coarsening mechanism available
- Different types of coarsening
  - Rounding
  - Grouping
  - Heaping
  - Censoring e.g. failure time data
  - ...

Ontologic uncertainty

- Precise observation of something imprecise
- Coarse observations are true observations
  $\Rightarrow$ No coarsening mechanism available
- Imprecision because of Indecision
  e.g. some respondents are indecisive between electing party A and party B
  $\Rightarrow$ category "A or B" (AB) represents the truth

Questions to be answered

- Are there some general approaches for dealing with...
  - ... epistemic uncertainty?
  - ... ontologic uncertainty?
- How can those types of uncertainty be involved within a regression model?
- How can those approaches be compared?

Here:

- Categorical data only
- Coarse dependent variable only

Approaches for dealing with epistemic uncertainty

Initial problem:

$$P(\mathcal{Y} = \mathfrak{y}) = \sum_{Y} \underbrace{P(\mathcal{Y} = \mathfrak{y}) | Y = y)}_{q} \cdot P(Y = y)$$

Possible solutions:

- Assuming ignorability: Coarsening at random
- Set valued results by procedures that avoid making unjustified assumptions
  - Partial identification
  - Sensitivity analysis

Coarsening at random (CAR)

Likelihood according to Heitjan and Rubin (1991)

$$L(\theta, \gamma, \mathfrak{y}) = \int_{\mathfrak{y}} q(\mathfrak{y}|y, \gamma)\ f(y, \theta) dy$$

Types of coarsening:

nonstochastic:

$$q(\mathfrak{y}|y, \gamma) = r(\mathfrak{y}|y, \theta) = \begin{cases} 1, & \text{if } \mathfrak{y} = \mathcal{Y}(y) \\ 0, & \text{if } \mathfrak{y} \neq \mathcal{Y}(y) \end{cases}$$

stochastic:

$$q(\mathfrak{y}|y, \gamma) = \int_{\Gamma} r(\mathfrak{y}|y, g)\ h(g|y, \gamma) dg.$$

Coarsening at random (CAR)

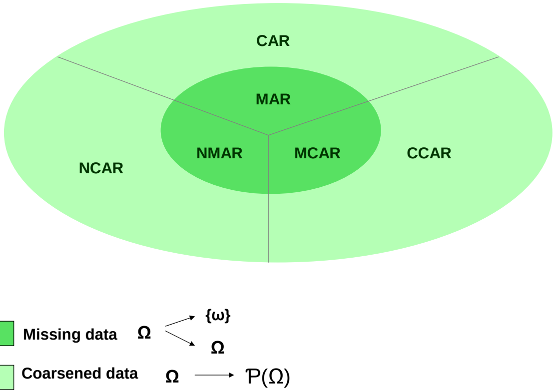
"Under which circumstances can stochastic nature of the coarsening be ignored?"

$\Rightarrow$ CAR + distinct parameters

"The data are CAR if,
for the fixed observed value of $\mathfrak{y}$ and for each value of γ,
$q(\mathfrak{y}|y,\gamma)$ takes the same value for all $y \in \mathfrak{y}$,
i.e., for all values of y that are consistent with $\mathfrak{y}$."
– Heitjan and Rubin, 1991, p.2248 –

Relation to the missing data problem

Missing as a special case of coarsening



Partial identification

- Identification region in context of missing data according to Manksi (2003):

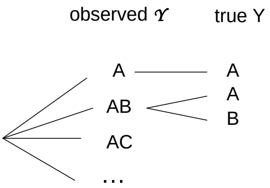
$$H[P(Y=y)] \equiv [P(Y=y|g=1)\cdot P(g=1) + \underbrace{P(Y=y|g=0)}_{\gamma}\cdot P(g=0), \gamma \in \Gamma_Y].$$

- Identification region in context of coarsened data:

$$\begin{aligned} H[P(Y=A)] \equiv \quad & [\underbrace{P(Y=A|\mathcal{Y}=A)}_{1} P(\mathcal{Y}=A) + \\ & \underbrace{P(Y=A|\mathcal{Y}=AB)}_{\gamma_1}\cdot P(\mathcal{Y}=AB) + \\ & \underbrace{P(Y=A|\mathcal{Y}=AC)}_{\gamma_2} P(\mathcal{Y}=AC), \\ & \forall \text{ possible } P(Y=A|\mathcal{Y}=\mathfrak{y}) = \gamma_i,\ i=1,2] \end{aligned}$$

Partial identification: Different points of view

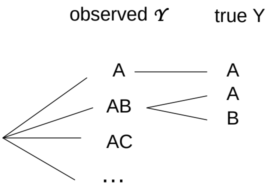
Conditioning on the observed variable $P(Y=y|\Upsilon=y)$ ....



$$
\begin{aligned}
P(Y=A) = {} & P(\Upsilon=A)\,\overbrace{P(Y=A|\Upsilon=A)}^{1} + \\
& P(\Upsilon=AB)\,\overbrace{P(Y=A|\Upsilon=AB)}^{\gamma_1} + \\
& P(\Upsilon=AC)\,\overbrace{P(Y=A|\Upsilon=AC)}^{\gamma_2}
\end{aligned}
$$

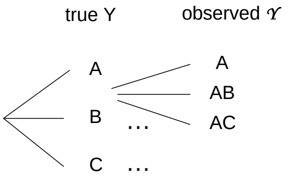
Partial identification: Different points of view

Conditioning on the observed variable $P(Y=y|\Upsilon=y)$ ....



$$
\begin{aligned}
P(Y=A) = {} & P(\Upsilon=A)\,\overbrace{P(Y=A|\Upsilon=A)}^{1} + \\
& P(\Upsilon=AB)\,\overbrace{P(Y=A|\Upsilon=AB)}^{\gamma_1} + \\
& P(\Upsilon=AC)\,\overbrace{P(Y=A|\Upsilon=AC)}^{\gamma_2}
\end{aligned}
$$

Conditioning on the true variable $P(\Upsilon=y|Y=y)$ ....



$$
\begin{aligned}
P(\Upsilon=AB) = {} & P(Y=A)\,\overbrace{P(\Upsilon=AB|Y=A)}^{q_1} + \\
& P(Y=B)\,\overbrace{P(\Upsilon=AB|Y=B)}^{q_2} + \\
& P(Y=C)\,\overbrace{P(\Upsilon=AC|Y=C)}^{0}
\end{aligned}
$$

Partial identification: Different points of view

	Approach 1	Approach 2
Starting point	P(Y=y)= $= \sum_{\mathcal{Y}} P(\mathcal{Y} = \mathfrak{y} \mid Y = y) P(Y = y)$	$P(\mathcal{Y} = \mathfrak{y}) =$ $= \sum_{\mathcal{Y}} P(\mathcal{Y} = \mathfrak{y} \mid Y = y) P(\mathcal{Y} = \mathfrak{y})$
Assumptions on...	$P(Y = y \mid \mathcal{Y} = \mathfrak{y}) = \gamma$ (conditioning on observed variable	$P(\mathcal{Y} = \mathfrak{y} \mid Y = y) = q$ (conditioning on true variable)
Empirical evidence	- $\gamma \in [0, 1]$	$\overline{q}_1 \leq \frac{P(\mathcal{Y}=AB)}{P(\mathcal{Y}=A)+P(\mathcal{Y}=AB)}$ $\overline{q}_2 \leq \frac{P(\mathcal{Y}=AB)}{P(\mathcal{Y}=B)+P(\mathcal{Y}=AB)}$ - No lower bound $\underline{q}_1$ and $\underline{q}_2$
Further assumptions	- Make plausible set-valued assumptions about γ - Evaluate by contentual aspects if $\gamma_1 > \gamma_2$ or vice versa	- CAR - Assumption about $R = \frac{q_2}{q_1}$

Sensitivity analysis

Foundations:

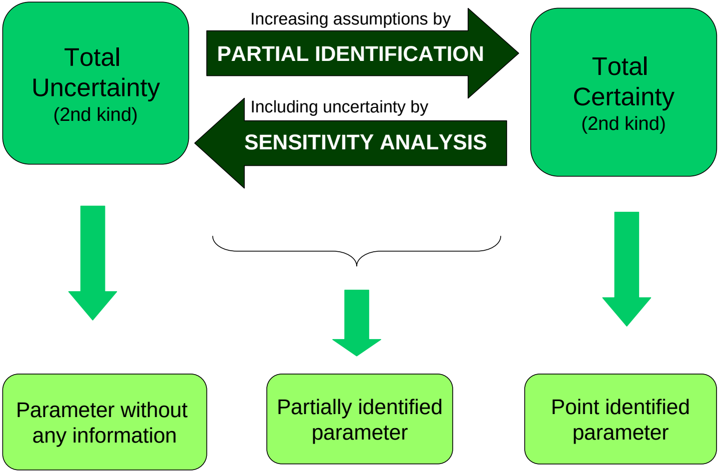
- Ignorance region $ir(\theta, \Delta)$: whole collection of θ-values that result from different δ
- Sensitivity parameter δ: Parameter of interest is identified given the value of δ
- Selection model in context of missing data

$$\pi_{g_1 g_2, ij} = p_{ij} \cdot q_{g_1 g_2 | ij}.$$

Application in the framework of coarse data

- Selection model: $\pi_{AB,A} = p_A \cdot q_{AB|A}$
- Additional restriction: $q_{A|A} + q_{AB|A} + q_{AC|A} = 1$
  $\Rightarrow q_{AC|A} \leq 1 - q_{A|A} - q_{AB|A}$

Comparison of partial identification and sensitivity analysis



Dealing with ontologic uncertainty

- Introduction of the distribution on the power set in order to deal with ontologic uncertainty (*-notation)
- Foundations adapted from
  - Random set theory: Finite random sets (basic idea)
  - Dempster-Shafer theory (interpretation of notions, prediction)

Finite random sets

Important notions:

- Definition of finite random sets:

  A finite random set with values in $\mathcal{P}(\Omega)$ is a map $X : \Omega \to \mathcal{P}(\Omega)$ such that $X^{-1}(\{A\}) = \{\omega \in \Omega : X(\omega) = A\} \in \mathcal{A}$, for any $A \subseteq \Omega$.

- Definition of distribution on the power set and capacity functional (DST: belief, plausibility)

General analysis of coarse data under ontologic uncertainty

Coarse observations can be regarded as own outcomes $\Rightarrow$ analysis on the power set $\mathcal{P}(\Omega) \setminus \emptyset = \Omega^\star$

$$\begin{aligned} P^\star : \mathcal{P}(\Omega^\star) = \mathcal{P}(\mathcal{P}(\Omega) \setminus \emptyset) &\to \mathbb{R} \\ A^\star &\to \mathbb{R}, \end{aligned}$$

Dempster-Shafer theory - Introduction

Example: Who has filled the role of Santa Claus this year? A, B or C?
Query set $Q = [B, C]$

person no.	1	2	3	4	5	6
guess of person	$[A, B]$	$[A, C]$	A	C	[A, B, C]	[B, C]

- Measure of belief: include all guesses g_i that are fully contained within the query set (i.e. $g_i \subseteq Q$)
  $\Rightarrow \; Bel(Q) = \frac{2}{6} = \frac{1}{3}$
- Measure of plausibility: involve all guesses g_i that intersect the query set Q (i.e. $g_i \cap Q \neq \emptyset$)
  $\Rightarrow \; Pl(Q) = \frac{5}{6}$

Dempster-Shafer theory - Important notions

- *Basic probability assignement* $m : \mathcal{P}(\Omega) \to [0,1]$
    - $m(\emptyset) = 0$ and $\sum_{A \subseteq \Omega} m(A) = 1$
    - confidence that can be exactly commited to A
- *Belief function* $Bel : \mathcal{P}(\Omega) \to [0,1]$
    - $Bel(\Omega) = 1,\ Bel(\emptyset) = 0$
    - ∞-monotone, i.e. $Bel(\cup_{i=1}^{k} A_j) \geq \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,k\}} (-1)^{|I|+1} Bel(\cap_{i \in I} A_i)$
    - calculation: $Bel(Q) = \sum_{A \subseteq Q} m(A)$
- *Plausibility function* $Pl : \mathcal{P}(\Omega) \to [0,1]$
    - $Pl(\Omega) = 1,\ Pl(\emptyset) = 0$
    - alternating of infinite order, i.e. $Pl(\cap_{j=1}^{k}) \leq \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} Pl(\cup_{i \in I} K_i)$
    - calculation: $Pl(Q) = \sum_{A \cap Q \neq \emptyset} m(A)$

Belief and plausibility function as instruments for prediction

- Analysis based on distributions on the power set as generalization of classical probability theory
- Interpretation of *Pl* and *Bel* as lower and upper bound respectively

Prediction under the presence of ontologic uncertainty:

$\Pi^\star$: family of distributions on the power set, $F^\star : \Omega^\star \to [0,1]$

$$\begin{aligned} \underline{F}^\star(Q)^\star &= \sum_{A^\star \subseteq Q^\star} m^\star(A^\star) = \inf\{F^\star(A^\star) | F^\star \in \Pi^\star\} \\ \overline{F^\star}(Q^\star) &= \sum_{A^\star \cap Q^\star \neq \emptyset} m^\star(A^\star) = \sup\{F^\star(A) | F^\star \in \Pi^\star\} \\ \Rightarrow F^\star(Q^\star) &= [\underline{F}^\star(Q^\star), \overline{F^\star}(Q^\star)] \end{aligned}$$

where the length of the interval indicates the extent of ontologic uncertainty

Modelling approach with coarse data

- Goal: Involve 1.) epistemic and 2.) ontologic uncertainty within the dependent variable

- Precise multinomial logit model as a starting point:
  - $Y_i \in \{1,...,c\}$ is categorical and of nominal scale
  - The probability of occurrence for category r is determined by

$$P(Y_i = r|\mathbf{x}_i) = \pi_{ir} = \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta}_r)}{1 + \sum_{s=1}^{c-1} \exp(\mathbf{x}_i^T \boldsymbol{\beta}_s)}$$

  - ... and for the reference category by

$$P(Y_i = c|\mathbf{x}_i) = \pi_{ic} = 1 - \pi_{i1} - ... - \pi_{ic-1} = \frac{1}{1 + \sum_{s=1}^{c-1} \exp(\mathbf{x}_i^T \boldsymbol{\beta}_s)}$$

  - Solving for the linear predictor:

$$\log \frac{\pi_{ir}}{\pi_{ic}} = \mathbf{x}_i^T \boldsymbol{\beta}_r, \;\; r = 1, ..., c$$

Epistemic uncertainty - Data generating process

True categories: "A", "B", Observed categories: "A", "B", "A or B"

Data for *iid*-model

- *iid* assumption $\Rightarrow \pi_{iA} = \pi_A$ and $\pi_{iB} = \pi_B$
- Different combinations of $q_1 = P(Y_{coarse} = \text{A or B}|Y = \text{A})$ and $q_2 = P(Y_{coarse} = \text{A or B}|Y = \text{B}) \Rightarrow Ycoarse1, ..., Ycoarse81$
- 100 datasets of that kind:

Y	Ycoarse1	Ycoarse2	$\cdots$	Ycoarse81
B	A or B	B	$\cdots$	B
A	A	A	$\cdots$	A or B
B	B	B	$\cdots$	A or B
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A	A	A	$\cdots$	A or B

Epistemic uncertainty - Data generating process

True categories: "A", "B", Observed categories: "A", "B", "A or B"

Data for model with covariates

- sampling probabilities π_{iA} and π_{iB} are dependent on underlying values of covariates $X_{i1} \sim Po(3)$ and $X_{i2} \sim \mathcal{N}(0, 4)$
- same way of coarsening as in *iid*-model
- 100 datasets of that kind:

Y	X1	X2	Ycoarse1	Ycoarse2	...	Ycoarse81
A	7	0.2456983	A	A or B	···	A
A	1	1.7636975	A	A	···	A
A	5	0.8042766	A	A	···	A or B
B	2	0.5196141	B	B	···	B
⋮	⋮	⋮	⋮	⋮	⋮	⋮
B	3	-5.134471	B	A or B	···	A or B
A	1	-0.7402479	A	A	···	A
A	2	2.448102	A	A	···	A or B

Epistemic uncertainty - Models of interest

Model 1: *iid* model

$$\begin{aligned} L(q, \pi_{iA}) &= \prod_{\mathcal{Y}_i} P(\mathcal{Y} = \mathfrak{y}) \\ &= \prod_{i:\mathcal{Y}_i = A} \underbrace{P(\mathcal{Y} = A \mid Y = A)}_{(1-q_1)} \pi_A \prod_{i:\mathcal{Y}_i = B} \underbrace{P(\mathcal{Y} = B \mid Y = B}_{(1-q_2)} (1 - \pi_A) \\ &\quad \prod_{i:\mathcal{Y}_i = AB} \underbrace{P(\mathcal{Y} = A \text{ or } B \mid Y = A)}_{q_1} \pi_A + \underbrace{P(\mathcal{Y} = A \text{ or } B \mid Y = B)}_{q_2} (1 - \pi_A). \end{aligned}$$

$\Rightarrow$ estimators: $\hat{\pi}_A$, $\hat{q}_1$, $\hat{q}_2$

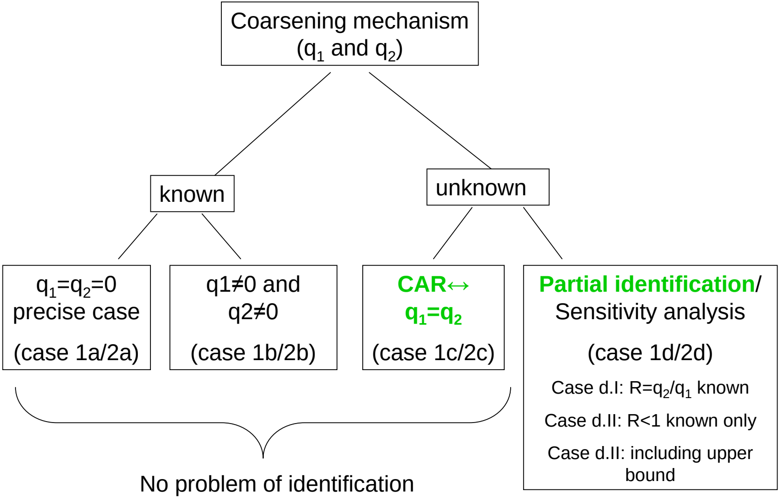
Model 2: Including covariates

$$\begin{aligned} L(q_1, q_2, \beta_A) &= \prod_{i=1}^{N_1} (1 - q_1) \frac{\exp(\beta_{A0} + x_{i1}\beta_{A1} + x_{i2}\beta_{A2})}{1 + \exp(\beta_{A0} + x_{i1}\beta_{A1} + x_{i2}\beta_{A2})} \prod_{i=N_1+1}^{N_2} (1 - q_2) \frac{1}{1 + \exp(\beta_{A0} + x_{i1}\beta_{A1} + x_{i2}\beta_{A2})} \\ &\quad \prod_{i=N_2+1}^{N} q_1 \frac{\exp(\beta_{A0} + x_{i1}\beta_{A1} + x_{i2}\beta_{A2})}{1 + \exp(\beta_{A0} + x_{i1}\beta_{A1} + x_{i2}\beta_{A2})} + \frac{q_2}{1 + \exp(\beta_{A0} + x_{i1}\beta_{A1} + x_{i2}\beta_{A2})} \end{aligned}$$

$\Rightarrow$ estimators: $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{q}_1$, $\hat{q}_2$

In both models: Problem of identification

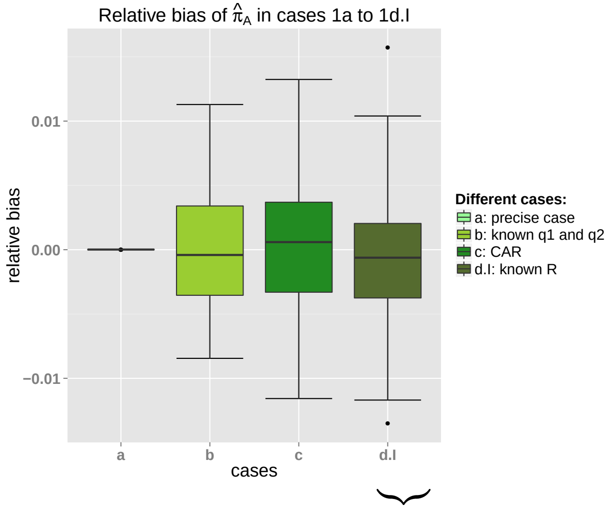
Epistemic uncertainty - Models of interest



No problem of identification

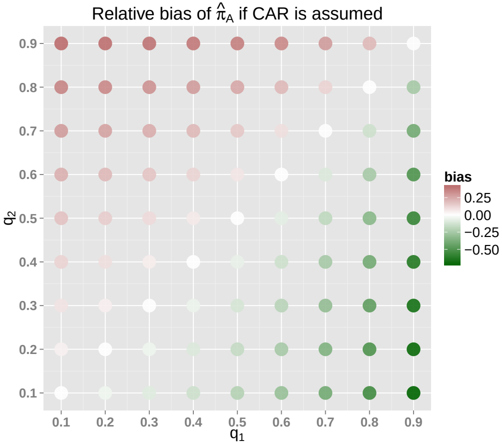
→ point identified solutions

Model 1 under epistemic uncertainty - Cases 1a to d.I



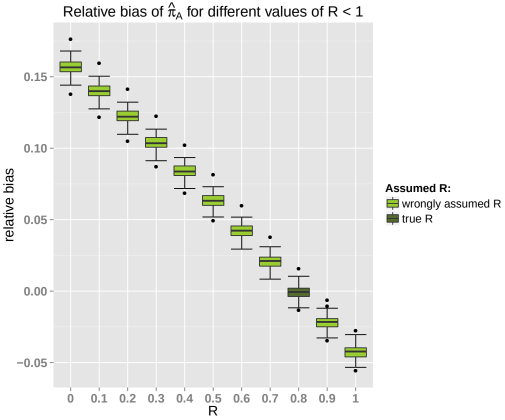
Generalization of CAR

Model 1c: Consequences if CAR is not valid



- Median rel. bias of $\hat{\pi}_A$ for combinations of q_1 and q_2
- Max. rel. bias of 0.72 if $q_1 = 0.9$ and $q_2 = 0.1$
- No symmetric problem

Model 1d.II: Assumption of $R < 1$ only



- Median relative bias between -0.05 and 0.18
- $\hat{\pi}_A \in [0.64, 0.78]$ if assumption is involved (true $\pi_A = 0.67$)
- $\hat{\pi}_A \in \left[\frac{n_A}{n}, \frac{n_A+n_{AB}}{n}\right] = [0.40, 0.77]$ if no assumption implied

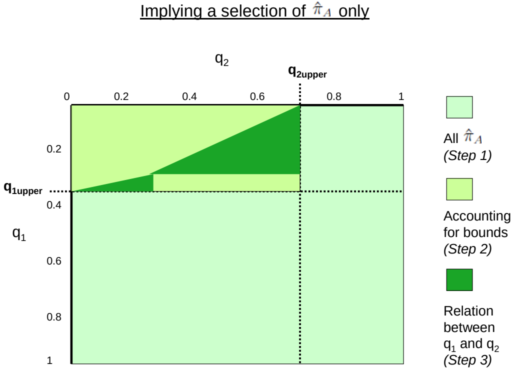
Model 1d.III: Involving upper bounds $\bar{q}_1$ and $\bar{q}_2$

How can the set of possible $\hat{\pi}_A$ be restricted by using the empirical evidence only?

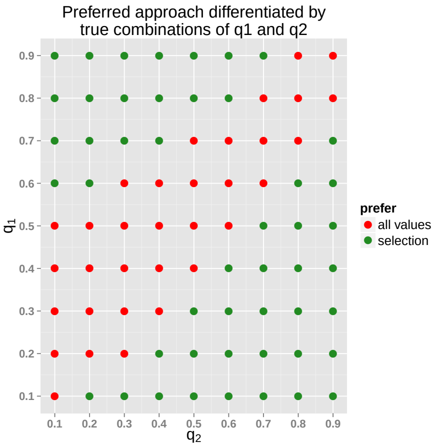
- First implying upper bounds only
- Involve relation between q_1 and q_2 additionally: $P(\mathcal{Y} = AB) = \pi_A q_1 + (1-\pi) q_2$

$$q_1 = \frac{n_{AB} - n q_2 (1-\pi_A)}{n \pi_A}$$

$$q_2 = \frac{n_{AB} - n q_1 \pi_A}{n(1-\pi_A)}$$



Model 1d.III: Involving upper bounds $\bar{q}_1$ and $\bar{q}_2$



- Median relative bias for the first dataset are depicted
- Method seems to be reasonable in cases that strongly differ from CAR only:

$$q_1 = q_2 = q$$

$$\Leftrightarrow \quad \frac{n_{AB} - qn(1-\pi_A)}{n\pi_A} = \frac{n_{AB} - qn\pi_A}{n(1-\pi_A)}$$

$\Rightarrow$ only valid if π_A=0.5

Multinomial logit model under ontologic uncertainty

Idea and particularity of the model

- Coarse values represent the truth
  $\Rightarrow$ Multinomial logit model with coarse observations (e.g. "A or B") as own categories
- No further changes compared to precise multinomial logit model

$\Rightarrow$ Predictions by means of Dempster-Shafer theory
$\Rightarrow$ How far does it make sense to imply additional assumptions?
$\Rightarrow$ Comparison of results under epistemic and ontologic uncertainty
Evaluation of estimators if wrong type of uncertainty is assumed

Conclusion

- Important to distinguish between epistemic and ontologic uncertainty
- For dealing with epistemic uncertainty some methods of the framework of missing data can be applied
- For dealing with ontologic uncertainty *-notation could be introduced
  - General dealing of coarse data and prediction
  - Formal background can be provided by random set theory and DST
- Multinomial logit model can be extended by accounting for...
  - ... epistemic uncertainty: Extending likelihood $\Rightarrow$ Identification problem $\Rightarrow$ Identifying restrictions as CAR or partial identification
  - ... ontologic uncertainty: Extending model by implying coarse categories as own categories

Literature

Heitjan, D. and Rubin, D.
Ignorability and Coarse Data.
Annals of Statistics, Volume 19, Issue 4, 1991.

Manski, C.
Partial Identification of Probability Distributions.
Springer, 2003.

Nguyen, H.
An Introduction to Random Sets.
CRC press, 2006.

Zadeh, L.
A Simple View of the Dempster-Shafer Theory of Evidence and its Implication for the Rule of Combination.
AI Magazine, Volume 7, Issue 2, 1986.