Statistical Modelling under Epistemic Data Imprecision

Some Results on Estimating Multinomial Distributions and Logistic Regression for Coarse Categorical Data

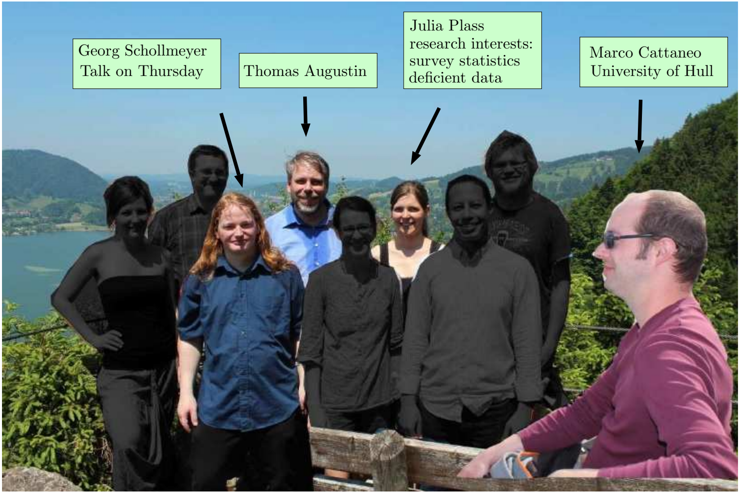
Julia Plass*, Thomas Augustin*,
Marco Cattaneo**, Georg Schollmeyer*

*Department of Statistics, Ludwigs-Maximilians University and
**Department of Mathematics, University of Hull

&

21st of July 2015

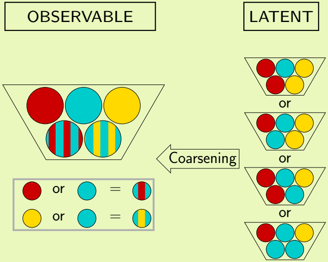
Our working group



Epistemic vs. ontic interpretation (Couso, Dubois, Sánchez, 2014)

Epistemic imprecision:

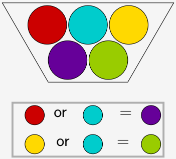
"Imprecise observation of something precise"



⇒ Truth is hidden due to the underlying coarsening mechanism

Ontic imprecision:

"Precise observation of something imprecise"

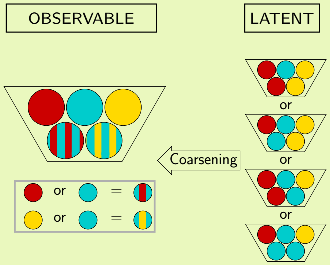


⇒ Truth is represented by coarse observation

Examples of data under epistemic imprecision

Epistemic imprecision:

"Imprecise observation of something precise"



$\Rightarrow$ Truth is hidden due to the underlying coarsening mechanism

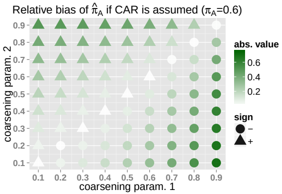
Examples:

- Matched data sets with partially overlapping variables
- Coarsening as anonymization technique
- Missing data as special case

Here: PASS-data
$\Omega_{\mathcal{Y}} = \{<, \geq, \text{na}\}$
"<1000", "≥ 1000" and
"$<1000€$ or $\geq 1000€$" (na)

Already existing approaches

- Still common to **enforce precise results**
  $\Rightarrow$ Biased results:



- Variety of **set-valued approaches**
  - via random sets (e.g. Nguyen, 2006)
  - via likelihood-based belief function (Denœux, 2014)
  - using Bayesian approaches (de Cooman, Zaffalon, 2004)
  - via profile likelihood (Cattaneo, Wiencierz, 2012)

Here: Likelihood-based approach
influenced by methodology of partial identification (Manski, 2003)
coarse categorical data only

Basic idea for the i.i.d. case (regression cf. poster)

OBSERVABLE

$\mathcal{Y}$ coarse data

$p_{\mathscr{Y}_i} = P(\mathcal{Y}_i = \mathscr{Y}_i),\ i = 1, \dots, n$

Observation model $\mathcal{Q}$

error-freeness

coarsening mechanism

$q_{\mathscr{Y}|y} = P(\mathcal{Y} = \mathscr{Y} | Y = y)$

LATENT

Y latent variable

Main goal:

Estimation of $\pi_{ij} = P(Y_i = j)$

$\pi_{i1} = \pi_1, \dots, \pi_{iK} = \pi_K$

$\boldsymbol{\gamma} = (\boldsymbol{q}_{\mathscr{Y}|\boldsymbol{y}}^T,\ \boldsymbol{\pi}_{\boldsymbol{y}}^T)^T$

Use random-set perspective and determine maximum-likelihood estimator $\hat{p}_{\mathscr{Y}}$

Likelihood for parameters $\boldsymbol{p} = (p_1, \dots, p_{|\Omega_{\mathcal{Y}}|-1})^T$

$L(\boldsymbol{p}) \propto \prod_{\mathscr{Y} \in \Omega_{\mathcal{Y}}} p_{\mathscr{Y}}^{n_{\mathscr{Y}}}$ is *uniquely* maximized by

$\boxed{\hat{p}_{\mathscr{Y}} = \frac{n_{\mathscr{Y}}}{n},}\quad \mathscr{Y} \in \{1, \dots, |\Omega_{\mathcal{Y}}| - 1\}$

and thus $\hat{p}_{|\Omega_{\mathcal{Y}}|} = 1 - \sum_{m=1}^{|\Omega_{\mathcal{Y}}|-1} \hat{p}_m$.

Use the **connection between $\boldsymbol{p}$ and $\boldsymbol{\gamma}$**

$$\Phi(\boldsymbol{\gamma}) = \mathbf{p}$$

and the **invariance of the likelihood** under parameter transformations, i.e.:

$$\boxed{\hat{\Gamma} = \{\boldsymbol{\gamma} \mid \Phi(\boldsymbol{\gamma}) = \hat{\boldsymbol{p}}\}}$$

$\hat{\pi}_y \in \left[\frac{n_{\{y\}}}{n}, \frac{\sum_{\mathscr{Y} \ni y} n_{\mathscr{Y}}}{n}\right]$

$\hat{q}_{\mathscr{Y}|y} \in \left[0, \frac{n_{\mathscr{Y}}}{n_{\{y\}} + n_{\mathscr{Y}}}\right]$

Basic idea for the i.i.d. case (regression cf. poster)

OBSERVABLE

$\mathcal{Y}$ coarse data

$p_{\mathscr{Y}_i} = P(\mathcal{Y}_i = \mathscr{Y}_i),\ i = 1, \dots, n$

Observation model $\mathcal{Q}$

error-freeness

coarsening mechanism

$q_{\mathscr{Y}|y} = P(\mathcal{Y} = \mathscr{Y} | Y = y)$

LATENT

Y latent variable

Main goal:

Estimation of $\pi_{ij} = P(Y_i = j)$

$\pi_{i1} = \pi_1, \dots, \pi_{iK} = \pi_K$

$\boldsymbol{\gamma} = (\boldsymbol{q}^T_{\mathscr{Y}|\boldsymbol{y}},\ \boldsymbol{\pi}^T_{\boldsymbol{y}})^T$

Use the **connection between $\boldsymbol{p}$ and $\boldsymbol{\gamma}$**

$$\Phi(\boldsymbol{\gamma}) = \mathbf{p}$$

Use random-set perspective and determine maximum-likelihood estimator $\hat{p}_{\mathscr{Y}}$

Likelihood for parameters $\boldsymbol{p} = (p_1, \dots, p_{|\Omega_{\mathcal{Y}}|-1})^T$

$L(\boldsymbol{p}) \propto \prod_{\mathscr{Y} \in \Omega_{\mathcal{Y}}} p_{\mathscr{Y}}^{n_{\mathscr{Y}}}$ is *uniquely* maximized by

$\boxed{\hat{p}_{\mathscr{Y}} = \frac{n_{\mathscr{Y}}}{n},}\qquad \mathscr{Y} \in \{1, \dots, |\Omega_{\mathcal{Y}}| - 1\}$

and thus $\hat{p}_{|\Omega_{\mathcal{Y}}|} = 1 - \sum_{m=1}^{|\Omega_{\mathcal{Y}}|-1} \hat{p}_m$.

and the **invariance of the likelihood** under parameter transformations, i.e.:

$$\hat{\Gamma} = \{\boldsymbol{\gamma} \mid \Phi(\boldsymbol{\gamma}) = \hat{\boldsymbol{p}}\}$$

$\hat{\pi}_y \in \left[\frac{n_{\{y\}}}{n},\ \frac{\sum_{\mathscr{Y} \ni y} n_{\mathscr{Y}}}{n}\right]$

$\hat{q}_{\mathscr{Y}|y} \in \left[0,\ \frac{n_{\mathscr{Y}}}{n_{\{y\}} + n_{\mathscr{Y}}}\right]$

Basic idea for the i.i.d. case (regression cf. poster)

OBSERVABLE

$\mathcal{Y}$ coarse data

$p_{\mathscr{Y}_i} = P(\mathcal{Y}_i = \mathscr{Y}_i),\ i = 1, \dots, n$

Observation model $\mathcal{Q}$

error-freeness

coarsening mechanism

$q_{\mathscr{Y}|y} = P(\mathcal{Y} = \mathscr{Y} | Y = y)$

LATENT

Y latent variable

Main goal:

Estimation of $\pi_{ij} = P(Y_i = j)$

$\pi_{i1} = \pi_1, \dots, \pi_{iK} = \pi_K$

$\boldsymbol{\gamma} = (\boldsymbol{q}_{\mathscr{Y}|\boldsymbol{y}}^T,\ \boldsymbol{\pi}_{\boldsymbol{y}}^T)^T$

Use the **connection between $\boldsymbol{p}$ and $\boldsymbol{\gamma}$**

$$\Phi(\boldsymbol{\gamma}) = \mathbf{p}$$

Use random-set perspective and determine maximum-likelihood estimator $\hat{p}_{\mathscr{Y}}$

Likelihood for parameters $\boldsymbol{p} = (p_1, \dots, p_{|\Omega_{\mathcal{Y}}|-1})^T$

$L(\boldsymbol{p}) \propto \prod_{\mathscr{Y} \in \Omega_{\mathcal{Y}}} p_{\mathscr{Y}}^{n_{\mathscr{Y}}}$ is *uniquely* maximized by

$\boxed{\hat{p}_{\mathscr{Y}} = \frac{n_{\mathscr{Y}}}{n},}$ $\quad \mathscr{Y} \in \{1, \dots, |\Omega_{\mathcal{Y}}| - 1\}$

and thus $\hat{p}_{|\Omega_{\mathcal{Y}}|} = 1 - \sum_{m=1}^{|\Omega_{\mathcal{Y}}|-1} \hat{p}_m$.

and the **invariance of the likelihood** under parameter transformations, i.e.:

$$\boxed{\hat{\Gamma} = \{\boldsymbol{\gamma} \mid \Phi(\boldsymbol{\gamma}) = \hat{\boldsymbol{p}}\}}$$

$\hat{\pi}_y \in \left[\frac{n_{\{y\}}}{n},\ \frac{\sum_{\mathscr{Y} \ni y} n_{\mathscr{Y}}}{n}\right]$

$\hat{q}_{\mathscr{Y}|y} \in \left[0,\ \frac{n_{\mathscr{Y}}}{n_{\{y\}} + n_{\mathscr{Y}}}\right]$

Basic idea for the i.i.d. case (regression cf. poster)

OBSERVABLE

$\mathcal{Y}$ coarse data

$p_{\mathscr{Y}_i} = P(\mathcal{Y}_i = \mathscr{Y}_i),\ i = 1, \dots, n$

Observation model $\mathcal{Q}$

error-freeness

coarsening mechanism
$q_{\mathscr{Y}|y} = P(\mathcal{Y} = \mathscr{Y} | Y = y)$

LATENT

Y latent variable

Main goal:
Estimation of $\pi_{ij} = P(Y_i = j)$
$\pi_{i1} = \pi_1, \dots, \pi_{iK} = \pi_K$

$\boldsymbol{\gamma} = (\boldsymbol{q}_{\mathscr{Y}|\boldsymbol{y}}^T,\ \boldsymbol{\pi}_{\boldsymbol{y}}^T)^T$

Use random-set perspective and determine maximum-likelihood estimator $\hat{p}_{\mathscr{Y}}$

Likelihood for parameters $\boldsymbol{p} = (p_1, \dots, p_{|\Omega_{\mathcal{Y}}|-1})^T$
$L(\boldsymbol{p}) \propto \prod_{\mathscr{Y} \in \Omega_{\mathcal{Y}}} p_{\mathscr{Y}}^{n_{\mathscr{Y}}}$ is *uniquely* maximized by
$\boxed{\hat{p}_{\mathscr{Y}} = \frac{n_{\mathscr{Y}}}{n},}$ $\quad \mathscr{Y} \in \{1, \dots, |\Omega_{\mathcal{Y}}| - 1\}$
and thus $\hat{p}_{|\Omega_{\mathcal{Y}}|} = 1 - \sum_{m=1}^{|\Omega_{\mathcal{Y}}|-1} \hat{p}_m$.

Use the **connection between $\boldsymbol{p}$ and $\boldsymbol{\gamma}$**

$$\Phi(\boldsymbol{\gamma}) = \mathbf{p}$$

and the **invariance of the likelihood** under parameter transformations, i.e.:

$$\boxed{\hat{\Gamma} = \{\boldsymbol{\gamma} \mid \Phi(\boldsymbol{\gamma}) = \hat{\boldsymbol{p}}\}}$$

$\hat{\pi}_y \in \left[\frac{n_{\{y\}}}{n}, \frac{\sum_{\mathscr{Y} \ni y} n_{\mathscr{Y}}}{n}\right]$
$\hat{q}_{\mathscr{Y}|y} \in \left[0, \frac{n_{\mathscr{Y}}}{n_{\{y\}} + n_{\mathscr{Y}}}\right]$

Basic idea for the i.i.d. case (regression cf. poster)

OBSERVABLE

$\mathcal{Y}$ coarse data

$p_{\mathscr{Y}_i} = P(\mathcal{Y}_i = \mathscr{y}_i),\ i = 1, \dots, n$

Observation model $\mathcal{Q}$

error-freeness

coarsening mechanism

$q_{\mathscr{y}|y} = P(\mathcal{Y} = \mathscr{y} | Y = y)$

LATENT

Y latent variable

Main goal:

Estimation of $\pi_{ij} = P(Y_i = j)$

$\pi_{i1} = \pi_1, \dots, \pi_{iK} = \pi_K$

$\boldsymbol{\gamma} = (\boldsymbol{q}_{\mathscr{y}|\boldsymbol{y}}^T,\ \boldsymbol{\pi}_{\boldsymbol{y}}^T)^T$

Use random-set perspective and determine maximum-likelihood estimator $\hat{p}_{\mathscr{y}}$

Likelihood for parameters $\boldsymbol{p} = (p_1, \dots, p_{|\Omega_{\mathcal{Y}}|-1})^T$

$L(\boldsymbol{p}) \propto \prod_{\mathscr{y} \in \Omega_{\mathcal{Y}}} p_{\mathscr{y}}^{n_{\mathscr{y}}}$ is *uniquely* maximized by

$\boxed{\hat{p}_{\mathscr{y}} = \frac{n_{\mathscr{y}}}{n},}$ $\quad \mathscr{y} \in \{1, \dots, |\Omega_{\mathcal{Y}}| - 1\}$

and thus $\hat{p}_{|\Omega_{\mathcal{Y}}|} = 1 - \sum_{m=1}^{|\Omega_{\mathcal{Y}}|-1} \hat{p}_m.$

Use the **connection between $\boldsymbol{p}$ and $\boldsymbol{\gamma}$**

$$\Phi(\boldsymbol{\gamma}) = \mathbf{p}$$

and the **invariance of the likelihood** under parameter transformations, i.e.:

$$\boxed{\hat{\Gamma} = \{\boldsymbol{\gamma} \mid \Phi(\boldsymbol{\gamma}) = \hat{\boldsymbol{p}}\}}$$

$\hat{\pi}_y \in \left[\frac{n_{\{y\}}}{n},\ \frac{\sum_{\mathscr{y} \ni y} n_{\mathscr{y}}}{n}\right]$

$\hat{q}_{\mathscr{y}|y} \in \left[0,\ \frac{n_{\mathscr{y}}}{n_{\{y\}} + n_{\mathscr{y}}}\right]$

Basic idea for the i.i.d. case (regression cf. poster)

OBSERVABLE

$\mathcal{Y}$ coarse data

$p_{\mathscr{Y}_i} = P(\mathcal{Y}_i = \mathscr{y}_i),\ i = 1, \dots, n$

Observation model $\mathcal{Q}$

error-freeness

coarsening mechanism
$q_{\mathscr{y}|y} = P(\mathcal{Y} = \mathscr{y} | Y = y)$

LATENT

Y latent variable

Main goal:
Estimation of $\pi_{ij} = P(Y_i = j)$
$\pi_{i1} = \pi_1, \dots, \pi_{iK} = \pi_K$

$\boldsymbol{\gamma} = (\boldsymbol{q}^T_{\mathscr{y}|\boldsymbol{y}},\ \boldsymbol{\pi}^T_{\boldsymbol{y}})^T$

Use the **connection between $\boldsymbol{p}$ and $\boldsymbol{\gamma}$**

$$\Phi(\boldsymbol{\gamma}) = \mathbf{p}$$

Use random-set perspective and determine maximum-likelihood estimator $\hat{p}_{\mathscr{y}}$

Likelihood for parameters $\boldsymbol{p} = (p_1, \dots, p_{|\Omega_\mathcal{Y}|-1})^T$
$L(\boldsymbol{p}) \propto \prod_{\mathscr{y} \in \Omega_\mathcal{Y}} p_{\mathscr{y}}^{n_{\mathscr{y}}}$ is *uniquely* maximized by

$\boxed{\hat{p}_{\mathscr{y}} = \frac{n_{\mathscr{y}}}{n},}$ $\quad \mathscr{y} \in \{1, \dots, |\Omega_\mathcal{Y}| - 1\}$

and thus $\hat{p}_{|\Omega_\mathcal{Y}|} = 1 - \sum_{m=1}^{|\Omega_\mathcal{Y}|-1} \hat{p}_m$.

and the **invariance of the likelihood** under parameter transformations, i.e.:

$$\hat{\Gamma} = \{\boldsymbol{\gamma} \mid \Phi(\boldsymbol{\gamma}) = \hat{\boldsymbol{p}}\}$$

$\hat{\pi}_y \in \left[\frac{n_{\{y\}}}{n},\ \frac{\sum_{\mathscr{y} \ni y} n_{\mathscr{y}}}{n}\right]$

$\hat{q}_{\mathscr{y}|y} \in \left[0,\ \frac{n_{\mathscr{y}}}{n_{\{y\}} + n_{\mathscr{y}}}\right]$

Basic idea for the i.i.d. case (regression cf. poster)

OBSERVABLE

$\mathcal{Y}$ coarse data

$p_{\mathscr{y}_i} = P(\mathcal{Y}_i = \mathscr{y}_i),\ i = 1, \dots, n$

Observation model $\mathcal{Q}$ (arrow pointing left)

error-freeness

coarsening mechanism

$q_{\mathscr{y}|y} = P(\mathcal{Y} = \mathscr{y} | Y = y)$

LATENT

Y latent variable

Main goal:

Estimation of $\pi_{ij} = P(Y_i = j)$

$\pi_{i1} = \pi_1, \dots, \pi_{iK} = \pi_K$

$\boldsymbol{\gamma} = (\boldsymbol{q}_{\mathscr{y}|\boldsymbol{y}}^T,\ \boldsymbol{\pi}_{\boldsymbol{y}}^T)^T$

Use random-set perspective and determine maximum-likelihood estimator $\hat{p}_{\mathscr{y}}$

Likelihood for parameters $\boldsymbol{p} = (p_1, \dots, p_{|\Omega_{\mathcal{Y}}|-1})^T$

$L(\boldsymbol{p}) \propto \prod_{\mathscr{y} \in \Omega_{\mathcal{Y}}} p_{\mathscr{y}}^{n_{\mathscr{y}}}$ is *uniquely* maximized by

$\boxed{\hat{p}_{\mathscr{y}} = \frac{n_{\mathscr{y}}}{n},}$ $\quad \mathscr{y} \in \{1, \dots, |\Omega_{\mathcal{Y}}| - 1\}$

and thus $\hat{p}_{|\Omega_{\mathcal{Y}}|} = 1 - \sum_{m=1}^{|\Omega_{\mathcal{Y}}|-1} \hat{p}_m$.

Use the **connection between $\boldsymbol{p}$ and $\boldsymbol{\gamma}$**

$$\Phi(\boldsymbol{\gamma}) = \mathbf{p}$$

and the **invariance of the likelihood** under parameter transformations, i.e.:

$$\boxed{\hat{\Gamma} = \{\boldsymbol{\gamma} \mid \Phi(\boldsymbol{\gamma}) = \hat{\boldsymbol{p}}\}}$$

$\hat{\pi}_y \in \left[\frac{n_{\{y\}}}{n},\ \frac{\sum_{\mathscr{y} \ni y} n_{\mathscr{y}}}{n}\right]$

$\hat{q}_{\mathscr{y}|y} \in \left[0,\ \frac{n_{\mathscr{y}}}{n_{\{y\}} + n_{\mathscr{y}}}\right]$

Illustration (PASS data)

$n_< = 238,\ n_\geq = 835,\ n_{\text{na}} = 338$

$\hat{\pi}_< \in \left[\frac{238}{1411},\ \frac{238+338}{1411}\right]$

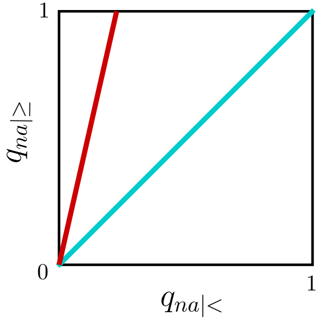
Starting from point-identifying assumptions, we use sensitivity parameters to allow inclusion of partial knowledge.

Assumption about exact value of $R = \frac{q_{na|\geq}}{q_{na|<}}$ (Nordheim, 1984):

e.g. $\mathcal{Q}$ specified by R=1, R=4

where R=1 corresponds to CAR

(Heitjan, Rubin, 1991).



Starting from point-identifying assumptions, we use sensitivity parameters to allow inclusion of partial knowledge.

Assumption about exact value of $R = \frac{q_{na|\geq}}{q_{na|<}}$ (Nordheim, 1984):

e.g. $\mathcal{Q}$ specified by R=1, R=4

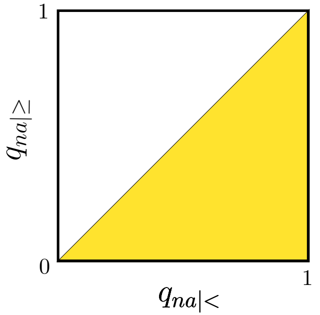
where R=1 corresponds to CAR

(Heitjan, Rubin, 1991).

Rough evaluation of R:

e.g. $\mathcal{Q}$ specified by R $\leq$ 1:

low income group has a higher tendency to report "na"



Summary and outlook

- Via the observation model $\mathcal{Q}$ maximum-likelihood estimators referring to the latent variable may be obtained for both cases
  - ... the homogeneous case
  - ... the case with categorical covariates (cf. poster)
- Proper inclusion of auxiliary information via further restrictions on $\mathcal{Q}$

Next steps:

- Inclusion of auxiliary information via sets of priors
- Likelihood-based hypothesis tests and uncertainty regions for coarse categorical data
- Consideration of other "deficiency" processes

References

- Couso, I., Dubois, D., Sánchez, L.
  Random Sets and Random Fuzzy Sets as Ill-Perceived Random Variables, Springer, 2014.
- Heitjan, D., Rubin, D.
  Ignorability and Coarse Data, Annals of Statistics, 1991.
- Manski, C.
  Partial Identification of Probability Distributions, Springer, 2003.
- E. Nordheim.
  Inference from nonrandomly missing categorical data:An example from a genetic study on Turner's syndrome, J. Am. Stat. Assoc., 1984.
- Vansteelandt, S., Goetghebeur, E., Kenward, M., Molenberghs, G.
  Ignorance and uncertainty regions as inferential tools in a sensitivity analysis, Stat. Sin., 2006.