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
Our working group

Thomas Augustin
Marco Cattaneo
University of Hull
Julia Plass
research interests:
survey statistics
deficient data
Georg Schollmeyer
Talk on Thursday
Marco Cattaneo
University of Hull
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”

OBSERVABLE  |  LATENT

⇒ 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
Ω\(\gamma\) = \{<, ≥, na\}
“< 1000”, “≥ 1000” and “< 1000€ or ≥ 1000€” (na)
Already existing approaches

- Still common to enforce precise results
  ⇒ 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)

Likelihood for parameters \( p = (p_1, \ldots, p_{|\Omega_Y|-1})^T \) is uniquely maximized by
\[
\hat{p}_Y = \frac{n_Y}{n}, \quad Y \in \{1, \ldots, |\Omega_Y| - 1\}
\]
and thus \( \hat{p}_{|\Omega_Y|} = 1 - \sum_{m=1}^{|\Omega_Y|-1} \hat{p}_m \).

Observation model \( Q \)

Error-freeness

Coarsening mechanism
\[
q_{\mathcal{Y}|y} = P(Y = \mathcal{Y}|Y = y)
\]

Main goal:

Estimation of \( \pi_{ij} = P(Y_i = j) \)
\[
\pi_{i1} = \pi_1, \ldots, \pi_{iK} = \pi_K
\]

Use the connection between \( p \) and \( \gamma \)

\[
\Phi(\gamma) = p
\]

Use random-set perspective and determine maximum-likelihood estimator \( \hat{p}_\mathcal{Y} \)

Likelihood for parameters \( p = (p_1, \ldots, p_{|\Omega_Y|-1})^T \)
\[
L(p) \propto \prod_{\mathcal{Y} \in \Omega_Y} p_{\mathcal{Y}}^{n_{\mathcal{Y}}}
\]

is uniquely maximized by
\[
\hat{p}_\mathcal{Y} = \frac{n_{\mathcal{Y}}}{n}, \quad \mathcal{Y} \in \{1, \ldots, |\Omega_Y| - 1\}
\]

and thus \( \hat{p}_{|\Omega_Y|} = 1 - \sum_{m=1}^{|\Omega_Y|-1} \hat{p}_m \).

\( Y \) latent variable

\( \gamma = (q_{\mathcal{Y}|y}^T, \pi_y^T)^T \)

and the invariance of the likelihood under parameter transformations, i.e.:
\[
\hat{\Gamma} = \{\gamma \mid \Phi(\gamma) = \hat{p}\}
\]

\( \hat{\pi}_y \in \left[ \frac{n_{(y)}}{n}, \frac{\sum_{\mathcal{Y} \in \Omega_Y} n_{\mathcal{Y}}}{n} \right] \)
\( \hat{q}_{\mathcal{Y}|y} \in [0, \frac{n_{\mathcal{Y}}}{n_{(y)} + n_{\mathcal{Y}}}] \)
Basic idea for the i.i.d. case (regression cf. poster)

Likelihood for parameters $p = (p_1, \ldots, p_{|Ω_Y|−1})^T$ is uniquely maximized by $\hat{p}_Y = \frac{n_Y}{n}$, $Y \in \{1, \ldots, |Ω_Y|−1\}$

Observation model $Q_Y$

Error-freeness

$q_{Y|y} = P(Y = y|Y = y)$

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

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

Likelihood for parameters $p = (p_1, \ldots, p_{|Ω_Y|−1})^T$

$L(p) \propto \prod_{\mathcal{Y} \in Ω_Y} p^{n_{\mathcal{Y}}} \frac{p_{\mathcal{Y}}}{n}$ is uniquely maximized by

$\hat{p}_{\mathcal{Y}} = \frac{n_{\mathcal{Y}}}{n}$, $\mathcal{Y} \in \{1, \ldots, |Ω_Y|−1\}$

and thus $\hat{p}_{|Ω_Y|} = 1 - \sum_{m=1}^{|Ω_Y|−1} \hat{p}_m$. 

Use the connection between $p$ and $γ$

$Φ(γ) = p$

and the invariance of the likelihood under parameter transformations, i.e.: $Γ = \{γ | Φ(γ) = \hat{p}\}$

Main goal:

Estimation of $π_{ij} = P(Y_i = j)$

$π_{i1} = π_1, \ldots, π_{iK} = π_K$

γ = $(q_{\mathcal{Y}|y}^T, π_Y^T)^T$

LATENT

Y latent variable

OBSERVABLE

Y coarse data

Use the connection between $p$ and $γ$
Basic idea for the i.i.d. case (regression cf. poster)

Likelihood for parameters \( p = (p_1, \ldots, p_{|\Omega_Y| - 1})^T \) is uniquely maximized by
\[
\hat{p}_Y = \frac{n_Y}{n}, \quad Y \in \{1, \ldots, |\Omega_Y| - 1\}
\]
and thus \( \hat{p}_i = 1 - \sum_{m=1}^{\Omega_Y - 1} \hat{p}_m. \)

Likelihood for parameters \( \gamma \)

\[
L(p) \propto \prod_{\gamma \in \Omega_Y} p_{\gamma}^{n_{\gamma}}
\]
is uniquely maximized by
\[
\hat{p}_i = \frac{n_{\gamma}}{n}, \quad \gamma \in \{1, \ldots, |\Omega_Y| - 1\}
\]
and thus \( \hat{p}_\gamma = 1 - \sum_{m=1}^{\Omega_Y - 1} \hat{p}_m. \)

Use random-set perspective and determine maximum-likelihood estimator \( \hat{p}_\gamma \)

\[
\hat{\Gamma} = \{\gamma \mid \Phi(\gamma) = \hat{p}\}
\]

\( \Phi(\gamma) = p \)

Main goal:

\[
\gamma = (q_T \gamma|_Y, \pi_T y)^T
\]

Observation model \( Q \)

Error-freeness

\[
q_Y|_y = P(Y = \gamma|Y = y)
\]

Coarsening mechanism

\[
p_Y i = P(Y_i = Y)
\]

Use the connection between \( p \) and \( \gamma \)

\( \Phi(\gamma) = p \)

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

\[
\hat{\gamma} = (q_T \gamma|_y, \pi_T y)^T
\]

LATENT

Y latent variable

OBSERVABLE

\( Y \) coarse data

Use random-set perspective and determine maximum-likelihood estimator \( \hat{p}_\gamma \)

Likelihood for parameters \( p = (p_1, \ldots, p_{|\Omega_Y| - 1})^T \)

Observation model \( Q \)

Error-freeness

\[
q_Y|_y = P(Y = \gamma|Y = y)
\]

Coarsening mechanism

\[
p_Y i = P(Y_i = Y)
\]

Use the connection between \( p \) and \( \gamma \)

\( \Phi(\gamma) = p \)

and the invariance of the likelihood under parameter transformations, i.e.:
Basic idea for the i.i.d. case (regression cf. poster)

Likelihood for parameters \( p = (p_1, \ldots, p_{|\Omega_Y|-1})^T \)
is uniquely maximized by
\[
\hat{p}_Y = \frac{n_Y}{n}, \quad Y \in \{1, \ldots, |\Omega_Y|-1\}
\]
and thus
\[
\hat{p}_{|\Omega_Y|} = 1 - \sum_{m=1}^{\Omega_Y} \hat{p}_m.
\]

Use random-set perspective and determine maximum-likelihood estimator \( \hat{p}_Y \)

Observation model \( Q \) error-freeness

Coarsening mechanism

Main goal:

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

Error-freeness

Main goal:

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

Use the connection between \( p \) and \( \gamma \)

\[
\Phi(\gamma) = p
\]

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

\[
\hat{\Gamma} = \{\gamma \mid \Phi(\gamma) = \hat{p}\}
\]

\[
\hat{\gamma} = (q^T_{\gamma|y}, \pi_y^T)^T
\]

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

\[
\hat{\pi}_y \in \left[\frac{n_{\pi_y}}{n}, \frac{\sum_{\gamma \neq y} n_{\gamma}}{n}\right]
\]

LATENT

Y latent variable

OBSERVABLE

\( \mathcal{Y} \) coarse data

Use the connection between \( p \) and \( \gamma \)

<table>
<thead>
<tr>
<th>( \gamma )</th>
<th>( \Phi(\gamma) = p )</th>
</tr>
</thead>
</table>

| \( \hat{\gamma} \) | \( (q^T_{\gamma|y}, \pi_y^T)^T \) |
|---|---|

Error-freeness

Coarsening mechanism

Main goal:

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

Use random-set perspective and determine maximum-likelihood estimator \( \hat{p}_Y \)

Likelihood for parameters \( p = (p_1, \ldots, p_{|\Omega_Y|-1})^T \)

\[
L(p) \propto \prod_{\mathcal{Y} \in \Omega_Y} p_{\mathcal{Y}}^{n_{\mathcal{Y}}}
\]
is uniquely maximized by

\[
\hat{p}_{\mathcal{Y}} = \frac{n_{\mathcal{Y}}}{n}, \quad \mathcal{Y} \in \{1, \ldots, |\Omega_Y|-1\}
\]

and thus

\[
\hat{p}_{|\Omega_Y|} = 1 - \sum_{m=1}^{\Omega_Y} \hat{p}_m.
\]
Basic idea for the i.i.d. case (regression cf. poster)

Likelihood for parameters $p = (p_1, \ldots, p_{|\Omega Y| - 1})^T$
is uniquely maximized by

$$\hat{p}_Y = \frac{n_Y}{n}, \quad Y \in \{1, \ldots, |\Omega Y| - 1\}$$

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

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

Likelihood for parameters $p = (p_1, \ldots, p_{|\Omega Y| - 1})^T$

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

$$\hat{p}_\mathcal{Y} = \frac{n_{\mathcal{Y}}}{n}, \quad \mathcal{Y} \in \{1, \ldots, |\Omega Y| - 1\}$$

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

Main goal:

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

$$\tilde{\pi}_1 = \pi_1, \ldots, \tilde{\pi}_K = \pi_K$$

Use the connection between $p$ and $\gamma$

$$\Phi(\gamma) = p$$

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

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

Observation model $Q$

error-freeness

coarsening mechanism

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

$$p_Y \propto \prod_{Y \in \Omega Y} p_n Y_Y$$

Observation model $Q$

coarse data

error-freeness

$q_Y | y = P(Y = y | Y = y)$

$$p_Y i = P(Y_i = Y)$$

$$i = 1, \ldots, n$$

LATENT

$Y$ latent variable

and thus

$$\hat{\pi}_y \in \left[\frac{n_y}{n}, \frac{\sum_{\mathcal{Y} \in \Omega Y} n_{\mathcal{Y}}}{n}\right]$$

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

$\mathcal{Y}$ coarse data

$\mathcal{Y}$ coarse data
Basic idea for the i.i.d. case (regression cf. poster)

Likelihood for parameters $p = (p_1, \ldots, p_{|\Omega_Y| - 1})^T$ is uniquely maximized by

$$\hat{p}_Y = \frac{n_Y}{n}, \quad Y \in \{1, \ldots, |\Omega_Y| - 1\}$$

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

Observation model $Q$

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

coarsening mechanism

$$p_{\mathcal{Y}} = P(Y = \mathcal{Y}), \quad i = 1, \ldots, n$$

Observation model $Q$

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

coarsening mechanism

$$p_{\mathcal{Y}} = P(Y = \mathcal{Y}), \quad i = 1, \ldots, n$$

Observation model $Q$

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

coarsening mechanism

$$p_{\mathcal{Y}} = P(Y = \mathcal{Y}), \quad i = 1, \ldots, n$$

Main goal:

$$\gamma = (q_{\mathcal{Y}|y}^T, \pi_{y}^T)^T$$

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

Likelihood for parameters $p = (p_1, \ldots, p_{|\Omega_Y| - 1})^T$

$$L(p) \propto \prod_{\mathcal{Y} \in \Omega_Y} p_{\mathcal{Y}}^{n_{\mathcal{Y}}}$$

is uniquely maximized by

$$\hat{p}_{\mathcal{Y}} = \frac{n_{\mathcal{Y}}}{n}, \quad \mathcal{Y} \in \{1, \ldots, |\Omega_Y| - 1\}$$

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

Use the connection between $p$ and $\gamma$

$$\Phi(\gamma) = p$$

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

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

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

Use random-set perspective and determine maximum-likelihood estimator $\hat{p}_Y$
**Basic idea for the i.i.d. case (regression cf. poster)**

**OBSERVABLE**

\( Y \) coarse data

\[ p_{\mathcal{Y}_i} = P(Y_i = \mathcal{Y}_i), \ i = 1, \ldots, n \]

Use random-set perspective and determine maximum-likelihood estimator \( \hat{p}_{\mathcal{Y}} \)

Likelihood for parameters \( p = (p_1, \ldots, p_{|\Omega_Y| - 1})^T \)

\[ L(p) \propto \prod_{\mathcal{Y} \in \Omega_Y} p_{\mathcal{Y}}^{n_{\mathcal{Y}}} \]

is uniquely maximized by

\[ \hat{p}_{\mathcal{Y}} = \frac{n_{\mathcal{Y}}}{n}, \ \mathcal{Y} \in \{1, \ldots, |\Omega_Y| - 1\} \]

and thus

\[ \hat{p}_{|\Omega_Y|} = 1 - \sum_{m=1}^{|\Omega_Y| - 1} \hat{p}_m. \]

**LATENT**

**Observation model**

\( Q \) error-freeness

**Y latent variable**

Main goal:

\[ \hat{\gamma} = (q_{\mathcal{Y}|y}^T, \pi_y^T)^T \]

Use the connection between \( p \) and \( \gamma \)

\[ \Phi(\gamma) = p \]

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

\[ \hat{\Gamma} = \{\gamma \mid \Phi(\gamma) = \hat{p}\} \]

**Illustration (PASS data)**

\( n_\prec = 238, n_\succ = 835, n_{\text{na}} = 338 \)

\[ \hat{n}_\prec \in \left[ \frac{238}{1411}, \frac{238 + 338}{1411} \right] \]
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. $Q$ 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. $Q$ specified by $R=1$, $R=4$

where $R=1$ corresponds to CAR (Heitjan, Rubin, 1991).

**Rough evaluation of** $R$:

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

low income group has a higher tendency to report “na”
Summary and outlook

- Via the observation model $Q$ 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 $Q$

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
Couso, I., Dubois, D., Sánchez, L.  

Heitjan, D., Rubin, D.  

Manski, C.  

E. Nordheim.  

Vansteelandt, S., Goetghebeur, E., Kenward, M., Molenberghs, G.  