Testing of coarsening mechanisms: Coarsening at random versus subgroup independence

SMPS 2016, Rome

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

*Department of Statistics, Ludwig-Maximilians University and
**School of Mathematics and Physical Sciences, University of Hull

&

12th of September 2016

What's the problem?

Common dealing with incomplete data: assumptions

$\Rightarrow$ Missing at random (MAR) / coarsening at random (CAR)

$\Rightarrow$ Frequently: assumptions only for pragmatic reasons

What's the problem?

Common dealing with incomplete data: assumptions

$\Rightarrow$ Missing at random (MAR) / coarsening at random (CAR)

$\Rightarrow$ Frequently: assumptions only for pragmatic reasons

What are we doing?

Two types of uninformative coarsening:

The coarsening is independent of ...

- ... the true underlying value (CAR)
- ... the covariate's value (subgroup independence (SI))

What are we doing?

Two types of uninformative coarsening:

The coarsening is independent of ...

- ... the true underlying value (CAR)
- ... the covariate's value (subgroup independence (SI))

Here: categorical setting

1.) No assumptions about the coarsening
2.) CAR: intuitive insight why not testable
3.) SI: construction of Likelihood Ratio test

Coarse data (categorical setting)

Epistemic interpretation of coarse data (Couso, Dubois, 2014):

LATENT

random sample

$Y_1, \dots, Y_n$

$y_i \in \Omega_Y$

$i = 1, \dots, n$

coarsening

$y_i \in \mathfrak{y}_i$

OBSERVABLE

random sample

$\mathcal{Y}_1, \dots, \mathcal{Y}_n$

$\mathfrak{y}_i \in \mathcal{P}(\Omega_Y) \setminus \{\emptyset\}$

$i = 1, \dots, n$

Coarse data (categorical setting)

Epistemic interpretation of coarse data (Couso, Dubois, 2014):

LATENT		OBSERVABLE
random sample	coarsening $\longrightarrow$	random sample
$Y_1, \ldots, Y_n$		$\mathcal{Y}_1, \ldots, \mathcal{Y}_n$
$y_i \in \Omega_Y$	$y_i \in \mathcal{Y}_i$	$\mathcal{Y}_i \in \mathcal{P}(\Omega_Y) \setminus \{\emptyset\}$
$i = 1, \ldots, n$		$i = 1, \ldots, n$

		$\mathcal{Y}$							
		$\{a\}$	$\{b\}$	$\{c\}$	$\{a,b\}$	$\{a,c\}$	$\{b,c\}$	$\{a,b,c\}$	sum
X	0	$n_{0\{a\}}$	$n_{0\{b\}}$	$n_{0\{c\}}$	$n_{0\{a,b\}}$	$n_{0\{a,c\}}$	$n_{0\{b,c\}}$	$n_{0\{a,b,c\}}$	n_0
	1	$n_{1\{a\}}$	$n_{1\{b\}}$	$n_{1\{c\}}$	$n_{1\{a,b\}}$	$n_{1\{a,c\}}$	$n_{1\{b,c\}}$	$n_{1\{a,b,c\}}$	n_1

Table: Contingency table for coarse data

Coarse data (categorical setting)

Epistemic interpretation of coarse data (Couso, Dubois, 2014):

LATENT

random sample

$Y_1, \ldots, Y_n$

$y_i \in \Omega_Y$

$i = 1, \ldots, n$

coarsening

$y_i \in \pmb{y}_i$

OBSERVABLE

random sample

$\mathcal{Y}_1, \ldots, \mathcal{Y}_n$

$\pmb{y}_i \in \mathcal{P}(\Omega_Y) \setminus \{\emptyset\}$

$i = 1, \ldots, n$

		$\mathcal{Y}$			
		$\{a\}$	$\{b\}$	$\{a,b\}$	sum
X	0	$n_{0\{a\}}$	$n_{0\{b\}}$	$n_{0\{a,b\}}$	n_0
	1	$n_{1\{a\}}$	$n_{1\{b\}}$	$n_{1\{a,b\}}$	n_1

Table: Contingency table for missing data

Coarse data (categorical setting)

Epistemic interpretation of coarse data (Couso, Dubois, 2014):

LATENT		OBSERVABLE
random sample $Y_1, \ldots, Y_n$	coarsening	random sample $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$
$y_i \in \Omega_Y$, $i = 1, \ldots, n$	$y_i \in \mathfrak{y}_i$	$\mathfrak{y}_i \in \mathcal{P}(\Omega_Y) \setminus \{\emptyset\}$, $i = 1, \ldots, n$

- **UBII**: receipt of Unemployment Benefit II
- $\mathcal{Y}$: categorical income
  - $\{a\}$: < 1000€
  - $\{b\}$: ≥ 1000€
  - $\{a,b\}$: $<$ or $\geq$

		$\mathcal{Y}$			
		{a}	{b}	{a,b}	sum
UBII	0	38	385	95	518
	1	36	42	9	87

Table: PASS, w5 (Trappmann, 2010)

Coarse data (categorical setting)

Epistemic interpretation of coarse data (Couso, Dubois, 2014):

LATENT

random sample

$Y_1, \ldots, Y_n$

$y_i \in \Omega_Y$

$i = 1, \ldots, n$

coarsening

$y_i \in \mathscr{y}_i$

OBSERVABLE

random sample

$\mathcal{Y}_1, \ldots, \mathcal{Y}_n$

$\mathscr{y}_i \in \mathcal{P}(\Omega_Y) \setminus \{\emptyset\}$

$i = 1, \ldots, n$

		$\mathcal{Y}$		
		a	b	sum
UBII	0	38+**95**	385	518
	1	36+**9**	42	87

Table: potential true table

		$\mathcal{Y}$			
		{a}	{b}	{a,b}	sum
UBII	0	38	385	**95**	518
	1	36	42	**9**	87

Table: observed table

Coarse data (categorical setting)

Epistemic interpretation of coarse data (Couso, Dubois, 2014):

LATENT

random sample

$Y_1, \ldots, Y_n$

$y_i \in \Omega_Y$

$i = 1, \ldots, n$

coarsening

$y_i \in \mathcal{y}_i$

OBSERVABLE

random sample

$\mathcal{Y}_1, \ldots, \mathcal{Y}_n$

$\mathcal{y}_i \in \mathcal{P}(\Omega_Y) \setminus \{\emptyset\}$

$i = 1, \ldots, n$

		$\mathcal{Y}$		
		a	b	sum
UBII	0	38	385+**95**	518
	1	36	42+**9**	87

Table: potential true table

		$\mathcal{Y}$			
		{a}	{b}	{a,b}	sum
UBII	0	38	385	**95**	518
	1	36	42	**9**	87

Table: observed table

Coarse data (categorical setting)

Epistemic interpretation of coarse data (Couso, Dubois, 2014):

LATENT

random sample

$Y_1, \ldots, Y_n$

$y_i \in \Omega_Y$

$i = 1, \ldots, n$

coarsening

$y_i \in \mathcal{y}_i$

OBSERVABLE

random sample

$\mathcal{Y}_1, \ldots, \mathcal{Y}_n$

$\mathcal{y}_i \in \mathcal{P}(\Omega_Y) \setminus \{\emptyset\}$

$i = 1, \ldots, n$

		$\mathcal{Y}$		
		a	b	sum
UBII	0	38+**20**	385+**75**	518
	1	36+**5**	42+**4**	87

Table: potential true table

		$\mathcal{Y}$			
		{a}	{b}	{a,b}	sum
UBII	0	38	385	**95**	518
	1	36	42	**9**	87

Table: observed table

LATENT		OBSERVABLE
$\pi_{xy} :=$ $P(Y_i = y \mid X_i = x)$	(error-freeness) **Observation model** $q_{\mathfrak{y}\mid xy} :=$ $P(\mathcal{Y}_i = \mathfrak{y} \mid X_i = x, Y_i = y)$	$p_{x\mathfrak{y}} :=$ $P(\mathcal{Y}_i = \mathfrak{y} \mid X_i = x)$

General ML estimation (Plass, Augustin, Cattaneo, Schollmeyer, 2015)

LATENT

$\vartheta = (\pi_{xy}^{T},\ q_{\mathfrak{y}|xy}^{T})^{T}$

OBSERVABLE

$p_{x\mathfrak{y}} :=$
$P(\mathcal{Y}_i = \mathfrak{y} | X_i = x)$

LATENT	OBSERVABLE	
$\vartheta = (\pi_{xy}^T,\ q_{\mathfrak{y}\mid xy}^T)^T$	$p_{x\mathfrak{y}} := P(\mathcal{Y}_i = \mathfrak{y} \mid X_i = x)$	unique $\arg_{p_{x\mathfrak{y}}} L(p_{x\mathfrak{y}} \mid \text{data}) = \hat{p}_{x\mathfrak{y}} = \frac{n_{x\mathfrak{y}}}{n_x}$

1.) Determine MLE of observed variable distribution

General ML estimation (Plass, Augustin, Cattaneo, Schollmeyer, 2015)

LATENT	OBSERVABLE	
$\vartheta^{(1)}$, $\vartheta^{(2)}$, $\vartheta^{(3)}$, $\dots$	$p_{x\mathfrak{y}} := P(\mathcal{Y}_i = \mathfrak{y} \mid X_i = x)$	unique $\arg_{p_{x\mathfrak{y}}} L(p_{x\mathfrak{y}} \mid \text{data}) = \hat{p}_{x\mathfrak{y}} = \frac{n_{x\mathfrak{y}}}{n_x}$

$$\Phi : [0,1]^{\dim(\Theta_{lat})} \longrightarrow [0,1]^{\dim(\Theta_{obs})}$$

1.) Determine MLE of observed variable distribution

2.) Use connection between both worlds

$$p_{x\mathfrak{y}} = \sum_{y \in \mathfrak{y}} \left(\pi_{xy} \cdot q_{\mathfrak{y}|xy} \right).$$

LATENT

$\vartheta^{(1)}$

$\vartheta^{(2)}$

$\vartheta^{(3)}$

$\dots$

OBSERVABLE

$p_{x\mathfrak{y}} := P(\mathcal{Y}_i = \mathfrak{y} | X_i = x)$

$\arg_{\vartheta^{(1)}} L(\vartheta^{(1)} | \text{data})$

$\arg_{\vartheta^{(2)}} L(\vartheta^{(2)} | \text{data})$

$\arg_{\vartheta^{(3)}} L(\vartheta^{(3)} | \text{data})$

$\dots$

$$\Phi : [0,1]^{\dim(\Theta_{lat})} \longrightarrow [0,1]^{\dim(\Theta_{obs})}$$

1.) Determine MLE of observed variable distribution

2.) Use connection between both worlds

$$p_{x\mathfrak{y}} = \sum_{y \in \mathfrak{y}} \left(\pi_{xy} \cdot q_{\mathfrak{y}|xy} \right).$$

3.) Use invariance of the likelihood

$$\hat{\pi}_{xy} \in \left[\frac{n_{x\{y\}}}{n_x}, \frac{\sum_{\mathfrak{y} \ni y} n_{x\mathfrak{y}}}{n_x} \right], \quad \hat{q}_{\mathfrak{y}|xy} \in \left[0, \frac{n_{x\mathfrak{y}}}{n_{x\{y\}} + n_{x\mathfrak{y}}} \right].$$

General ML estimation (Plass, Augustin, Cattaneo, Schollmeyer, 2015)

LATENT		OBSERVABLE	
$\vartheta^{(\mathrm{CAR})}$	$\longrightarrow$	$p_{x\mathfrak{y}} := P(\mathcal{Y}_i = \mathfrak{y} \mid X_i = x)$	unique $\arg_{\vartheta^{(\mathrm{CAR})}} L(\vartheta^{(\mathrm{CAR})} \mid \mathrm{data})$

$$\Phi : [0,1]^{\dim(\Theta_{\mathrm{lat}})} \longrightarrow [0,1]^{\dim(\Theta_{\mathrm{obs}})}$$

1.) Determine MLE of observed variable distribution

2.) Use connection between both worlds

$$p_{x\mathfrak{y}} = \sum_{y \in \mathfrak{y}} \left(\pi_{xy} \cdot q_{\mathfrak{y}|xy} \right).$$

3.) Use invariance of the likelihood

$$\hat{\pi}_{xy} \in \left[\frac{n_{x\{y\}}}{n_x}, \frac{\sum_{\mathfrak{y} \ni y} n_{x\mathfrak{y}}}{n_x} \right], \quad \hat{q}_{\mathfrak{y}|xy} \in \left[0, \frac{n_{x\mathfrak{y}}}{n_{x\{y\}} + n_{x\mathfrak{y}}} \right].$$

Coarsening at random (CAR)

Definition of CAR (Heitjan, Rubin, 1991):
For each fixed $\mathcal{y}$ and x, $q_{\mathcal{y}|xy}$ takes the same value $\forall y \in \mathcal{y}$.

Coarsening at random (CAR)

Definition of CAR (Heitjan, Rubin, 1991):

For each fixed $\mathfrak{y}$ and x, $q_{\mathfrak{y}|xy}$ takes the same value $\forall y \in \mathfrak{y}$.

Illustrated by the example:

- The probability of $\{a,b\}$ is taken to be independent of the true income category in both subgroups split by UBII:

$$q_{\{a,b\}|0\mathbf{a}} = q_{\{a,b\}|0\mathbf{b}} \quad \text{and} \quad q_{\{a,b\}|1\mathbf{a}} = q_{\{a,b\}|1\mathbf{b}}$$

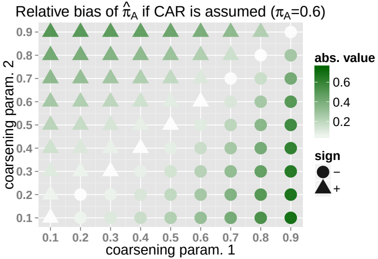
Definition of CAR (Heitjan, Rubin, 1991):

For each fixed $\mathcal{y}$ and x, $q_{\mathcal{y}|xy}$ takes the same value $\forall y \in \mathcal{y}$.

Illustrated by the example:

- The probability of $\{a,b\}$ is taken to be independent of the true income category in both subgroups split by UBII:

$$q_{\{a,b\}|0\mathbf{a}} = q_{\{a,b\}|0\mathbf{b}} \quad \text{and} \quad q_{\{a,b\}|1\mathbf{a}} = q_{\{a,b\}|1\mathbf{b}}$$



Coarsening at random (CAR)

Definition of CAR (Heitjan, Rubin, 1991):
For each fixed $\boldsymbol{y}$ and x, $q_{\boldsymbol{y}|xy}$ takes the same value $\forall y \in \boldsymbol{y}$.

Illustrated by the example:

- The probability of $\{a,b\}$ is taken to be independent of the true income category in both subgroups split by UBII:

$$q_{\{a,b\}|0\mathbf{a}} = q_{\{a,b\}|0\mathbf{b}} \quad \text{and} \quad q_{\{a,b\}|1\mathbf{a}} = q_{\{a,b\}|1\mathbf{b}}$$

- Resulting estimators:

$$\hat{\pi}_{xa}^{(CAR)} = \frac{n_{x\{a\}}}{n_{x\{a\}} + n_{x\{b\}}}, \quad \hat{q}_{\{a,b\}|xa}^{(CAR)} = \hat{q}_{\{a,b\}|xb}^{(CAR)} = \frac{n_{x\{a,b\}}}{n_x}$$

Estimators for subgroup 0:

$$\hat{\pi}_{0a} \in [0.07, 0.26], \quad \hat{q}_{\{a,b\}|0a} \in [0,\ 0.71], \quad \hat{q}_{\{a,b\}|0b} \in [0,\ 0.20]$$

$$\hat{\pi}_{0a}^{(CAR)} = 0.09, \qquad \hat{q}_{\{a,b\}|0a}^{(CAR)} = 0.18, \qquad \hat{q}_{\{a,b\}|0b}^{(CAR)} = 0.18$$

Testing CAR? ⇒ :-((e.g. Jaeger, 2006)

Estimators for subgroup 0:

$$\hat{\pi}_{0a} \in [0.07, 0.26], \quad \hat{q}_{\{a,b\}|0a} \in [0,\ 0.71], \quad \hat{q}_{\{a,b\}|0b} \in [0,\ 0.20]$$

$$\hat{\pi}^{(CAR)}_{0a} = 0.09, \qquad \hat{q}^{(CAR)}_{\{a,b\}|0a} = 0.18, \qquad \hat{q}^{(CAR)}_{\{a,b\}|0b} = 0.18$$

$\vartheta^{(CAR)}$

$$\hat{\pi}^{(CAR)}_{xa} = \frac{n_{x\{a\}}}{n_{x\{a\}} + n_{x\{b\}}}$$

$$\hat{q}^{(CAR)}_{\{a,b\}|xa} = \frac{n_{x\{a,b\}}}{n_x} = \hat{q}_{\{a,b\}|xb}$$

$\underline{\vartheta}$

$$\underline{\hat{\pi}}_{xa} = \frac{n_{x\{a\}}}{n_x}$$

$$\overline{\hat{q}}_{\{a,b\}|xb} = \frac{n_{x\{a,b\}}}{n_{x\{a,b\}} + n_{x\{b\}}}$$

$$\underline{\hat{q}}_{\{a,b\}|xa} = 0$$

$\Big\downarrow$ CAR

$\overline{\vartheta}$

$$\overline{\hat{\pi}}_{xa} = \frac{n_{x\{a\}} + n_{x\{a,b\}}}{n_x}$$

$$\overline{\hat{q}}_{\{a,b\}|xa} = \frac{n_{x\{a,b\}}}{n_{x\{a,b\}} + n_{x\{a\}}}$$

$$\underline{\hat{q}}_{\{a,b\}|xb} = 0$$

⇒ CAR is generally not testable, unless further assumptions about the coarsening are justified

Subgroup independence (SI)

Subgroup independence (Plass, Augustin, Cattaneo, Schollmeyer, 2015):
For each fixed $\mathfrak{y}$ and $y \in \mathfrak{y}$, $q_{\mathfrak{y}|xy}$ takes the same value $\forall x \in \Omega_X$.

Subgroup independence (SI)

Subgroup independence (Plass, Augustin, Cattaneo, Schollmeyer, 2015):
For each fixed $\mathfrak{y}$ and $y \in \mathfrak{y}$, $q_{\mathfrak{y}|xy}$ takes the same value $\forall x \in \Omega_X$.

Illustrated by the example:

- The probability of $\{a, b\}$ is taken to be independent of the receipt of the UBII given y:

  $q_{\{a,b\}|\mathbf{0}a} = q_{\{a,b\}|\mathbf{1}a}$ and $q_{\{a,b\}|\mathbf{0}b} = q_{\{a,b\}|\mathbf{1}b}$

Subgroup independence (SI)

Subgroup independence (Plass, Augustin, Cattaneo, Schollmeyer, 2015):
For each fixed $\mathfrak{y}$ and $y \in \mathfrak{y}$, $q_{\mathfrak{y}|xy}$ takes the same value $\forall x \in \Omega_X$.

Illustrated by the example:

- The probability of $\{a, b\}$ is taken to be independent of the receipt of the UBII given y:

  $q_{\{a,b\}|\mathbf{0}a} = q_{\{a,b\}|\mathbf{1}a}$ and $q_{\{a,b\}|\mathbf{0}b} = q_{\{a,b\}|\mathbf{1}b}$

- Resulting estimators (if well-defined and inside [0,1]):

$$\hat{\pi}^{(SI)}_{xa} = \frac{n_{x\{a\}}}{n_x}\frac{v}{w}, \qquad \hat{q}^{(SI)}_{\{a,b\}|xa} = 1 - \frac{w}{v}, \quad \hat{q}^{(SI)}_{\{a,b\}|xb} = 1 - \frac{w}{z}$$

with $v = n_0 n_{1\{b\}} - n_{0\{b\}} n_1$, $w = n_{0\{a\}} n_{1\{b\}} - n_{0\{b\}} n_{1\{a\}}$ and $z = n_{0A} n_1 - n_{1A} n_0$

Testing SI? ⇒ :-)

Estimators for subgroup 0:

$$\hat{\pi}_{0a} \in [0.07, 0.26], \quad \hat{q}_{\{a,b\}|0a} \in [0,\ 0.71], \quad \hat{q}_{\{a,b\}|0b} \in [0,\ 0.198]$$

$$\hat{\pi}_{0a}^{(SI)} = 0.42, \qquad \hat{q}_{\{a,b\}|0a}^{(SI)} = -0.04, \qquad \hat{q}_{\{a,b\}|0b}^{(SI)} = 0.20$$

Testing SI? ⇒ :-)

Estimators for subgroup 0:

$$\hat{\pi}_{0a} \in [0.07, 0.26], \quad \hat{q}_{\{a,b\}|0a} \in [0,\ 0.71], \quad \hat{q}_{\{a,b\}|0b} \in [0,\ 0.198]$$

$$\hat{\pi}_{0a}^{(SI)} = 0.42, \qquad \hat{q}_{\{a,b\}|0a}^{(SI)} = -0.04, \qquad \hat{q}_{\{a,b\}|0b}^{(SI)} = 0.20$$

⇒ There are data situations that might hint to (partial) incompatibility with SI

⇒ SI is testable in our setting

Testing SI: Hypothesis and test statistic

- Hypothesis:

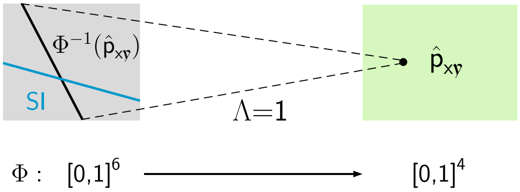
$$\begin{aligned} H_0 &: \quad q_{\mathfrak{y}|0y} = q_{\mathfrak{y}|1y} \text{ for all } y \in \Omega_Y = \{a,b\},\ \mathfrak{y} \in \mathcal{P}(\Omega_Y) \setminus \emptyset \\ H_1 &: \quad q_{\mathfrak{y}|0y} \neq q_{\mathfrak{y}|1y} \text{ for some } y \in \Omega_Y = \{a,b\},\ \mathfrak{y} \in \mathcal{P}(\Omega_Y) \setminus \emptyset. \end{aligned}$$

- Test based on test statistic:

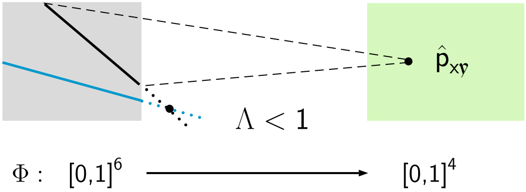
$$\Lambda(\mathfrak{y}_1, \dots, \mathfrak{y}_n, x_1, \dots, x_n) = \frac{\sup_{H_0} L(\vartheta|\mathfrak{y}_1, \dots, \mathfrak{y}_n, x_1, \dots, x_n)}{\sup_{H_0 \cup H_1} L(\vartheta|\mathfrak{y}_1, \dots, \mathfrak{y}_n, x_1, \dots, x_n)},$$

with $\vartheta = (\pi_{0a}, \pi_{1a}, q_{\{a,b\}|0a}, q_{\{a,b\}|1a}, q_{\{a,b\}|0b}, q_{\{a,b\}|1b})^T$

- No evidence to reject SI



- Some evidence to reject SI



- Result of the data example:
  $\Lambda \approx 0.93 \Rightarrow$ slight evidence against SI

Studying more general settings...

Now: number of subgroups m, $|\Omega_Y| = k$

CAR is generally known to be ...

- ... point-identifying
  (Heitjan, Rubin, 1991)
- ... not testable
  (e.g. Jaeger, 2006)

Studying more general settings...

Now: number of subgroups m, $|\Omega_{\mathrm{Y}}| = k$

CAR is generally known to be ...

- ... point-identifying (Heitjan, Rubin, 1991)
- ... not testable (e.g. Jaeger, 2006)

SI: A determination of the number of degrees of freedom df is crucial:

$$\mathrm{df} = \dim(\Theta_{\mathrm{obs}}) - \dim(\Theta_{\mathrm{lat,\,SI}}) = 2^{\mathrm{m}-1}(2\mathrm{k}-\mathrm{m})-(\mathrm{m}+1)(\mathrm{k}-1)-1$$

- Point-identification and testability are only valid if sufficient subgroups are available inducing $\mathrm{df} \geq 0$
- If $\mathrm{df} \geq 0$, a LR-Test on SI may be constructed with ...
  - ... test statistic $\mathrm{T} = -2 \cdot \log(\Lambda)$ (Wilks, 1938)
  - ... asymptotic distribution of T under H_0
    - $\frac{1}{2}\delta_0 + \frac{1}{2}\chi^2_1$ if df=0
    - χ^2_{df} if df>0,

  where δ_0 is the Dirac distribution at 0 (Chernoff, 1954)

Summary: CAR versus SI

	CAR	SI
Point-identifying?	always	in specific settings
Testability	generally impossible	possible in specific settings

Construction of a hypothesis test:

- H_0: SI, H_1: no SI
- Test statistic based on the Likelihood Ratio

Next step: Generalized version of LR-Test on SI

References

Couso, Dubois.
Statistical Reasoning with Set-Valued Information: Ontic vs. Epistemic Views, IJAR, 2014.

Chernoff.
On the distribution of the likelihood ratio, Ann. Stat. Math., 1954.

Heitjan, Rubin.
Ignorability and Coarse Data, Annals of Statistics, 1991.

Jaeger.
On testing the missing at random assumption, ECML, 2006.

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

Plass, Augustin, Cattaneo, Schollmeyer.
Statistical modelling under epistemic data imprecision, ISIPTA, 2015.

Trappmann, Gundert, Wenzig, Gebhardt.
PASS: a household panel survey for research on unemployment and poverty, Schmollers Jahrbuch, 2010.

Wilks.
The large-sample distribution of the likelihood ratio for testing composite hypotheses, Ann. Stat., 1938.