UNIVERSITY OF OSLO

integreat

Bridging Statistical and Logical Perspectives on Inconsistency

Lorenzo Zuccato [1] Anne-Marie George [2] Valeria Vitelli [3] Arnoldo Frigessi [3] Martin Giese [2] Ida Scheel [1]

Introduction

Inconsistency in real-world data is handled very differently across fields:

- **Logic & KR:** the principle of explosion makes classical reasoning collapse under inconsistency; **repair-based semantics** (AR, CAR, . . .) restore meaningful inference via maximally consistent subsets.
- **Statistics:** a *deterministic model* encoding the ideal logical rule is relaxed with an **error layer**, assigning non-zero probability to all observations and enabling likelihood-based inference.

Contribution: A unified framework formalizing statistical inconsistency in logical terms, enabling direct comparison of repair-based semantics and statistical models.

Unified Framework

- **Statistical Model:** $\mathcal{M} = (\mathcal{X}, \mathcal{Y}, \Theta, \{P_\theta\})$,
  $\mathcal{X}$: input space, $\mathcal{Y}$: output space, Θ: parameter space, $\{P_\theta\}$: parametric family.
- **Induced Logical Framework:** $\mathcal{L}_\mathcal{M} = (\Theta,\ \mathcal{X}\times\mathcal{Y},\ \vDash)$ defined by:

$$\theta \ \vDash\ (x, y) \quad \Longleftrightarrow \quad y \ \in\ \operatorname{supp}(P_\theta(Y \mid X = x)) .$$

- **Interpretations:** $\theta \in \Theta$. | **Assertions:** $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

A dataset $\mathcal{D}$ is **inconsistent** with $\mathcal{M}$ iff $L(\theta \mid \mathcal{D}) = 0,\ \forall\, \theta \in \Theta$.
Key equivalence:

$$\mathcal{D} \text{ is inconsistent with } \mathcal{M} \iff \mathcal{D} \text{ is unsatisfiable in } \mathcal{L}_\mathcal{M}.$$

This correspondence enables **repair-based semantics** for statistical datasets.
Repair: $R \subseteq \mathcal{D}$ is a maximal consistent subset (inclusion- or cardinality-maximal).

- **AR:** $\mathcal{D} \vDash_{\mathrm{AR}} f \iff \forall R \in \mathcal{R}(\mathcal{D}),\ R \vDash f.$
- **CAR:** $\mathcal{D} \vDash_{\mathrm{CAR}} f \iff \forall R \in \mathcal{R}(\mathrm{cl}(\mathcal{D})),\ R \vDash f.$

Statistical view: cardinality repairs ≡ minimal deletions ≡ error minimization.

Binary Classification

Goal: classify unseen x under data inconsistent with a monotonicity assumption, i.e. given x, try to deduce $(x, 0)$ or $(x, 1)$.

- **Data:** $\mathcal{D} = \{(x_i, y_i)\}$, with $\mathcal{X} = \mathbb{R}$ and $\mathcal{Y} = \{0, 1\}$.
- **Deterministic model:** threshold $\theta \in \mathbb{R}$ separates two classes:

$$P_\theta(Y = y \mid X = x) = \mathbf{1}\{(x > \theta, y = 1) \lor (x \le \theta, y = 0)\}.$$

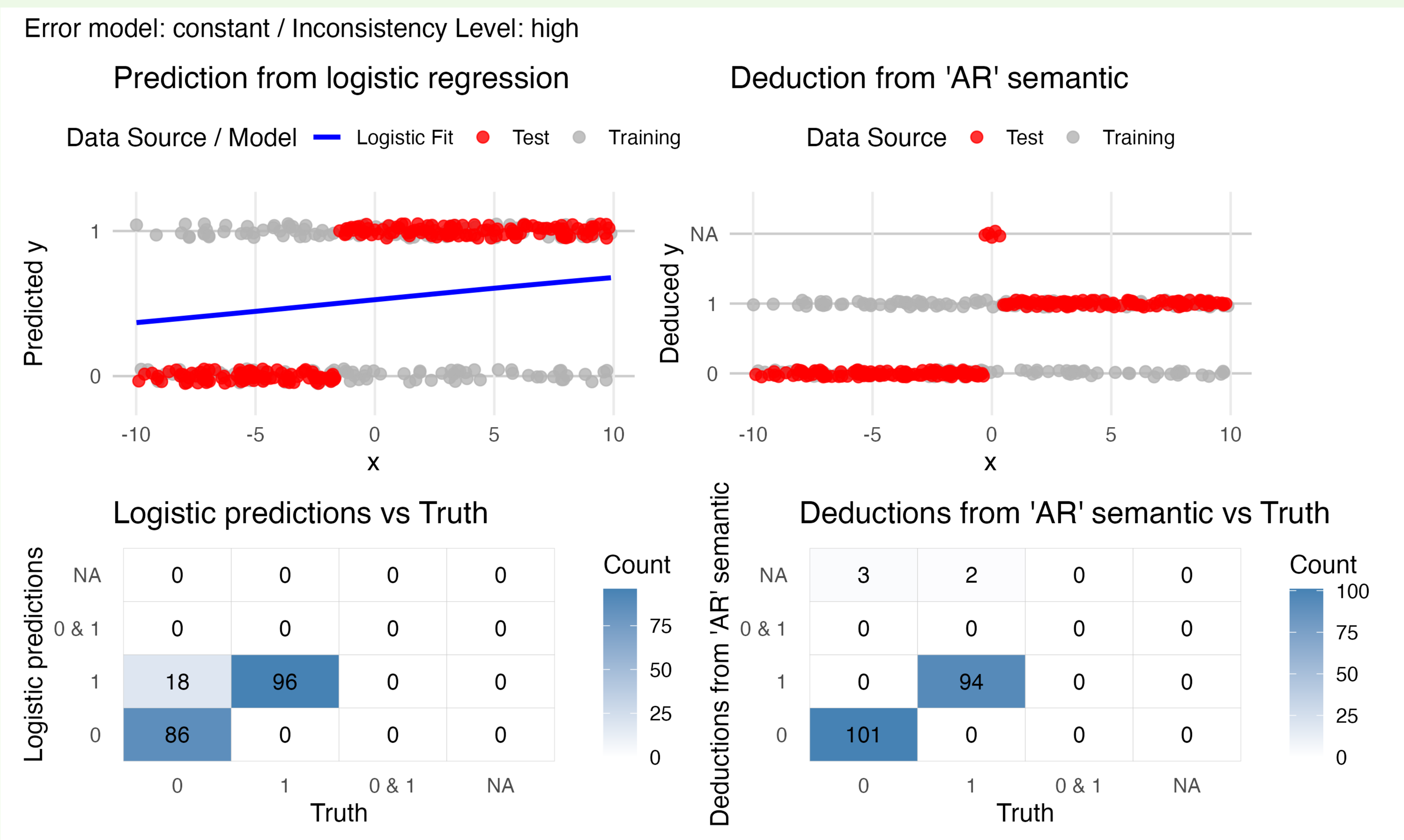
Inconsistency occurs when monotonicity is violated ($y = 1$ points left of $y = 0$ points).

- **Error model:** logistic regression $P_{\theta,s}(Y = 1 \mid X = x) = \sigma((x - \theta)/s)$, by MLE.
- **Repairs:** inclusion-maximal repairs are threshold-indexed consistent subsets

$$R_{x_j} = \{(x_i, y_i) \in \mathcal{D} \mid x_i \le x_j, y_i = 0 \lor x_i > x_j, y_i = 1\}.$$

Cardinality repairs minimize deletions; AR yields safe prediction intervals.

- **Experiments:** 201 labeled points in $[-10, 10]$; compare logistic regression and cardinality-AR predictions of unseen x under different inconsistency levels.

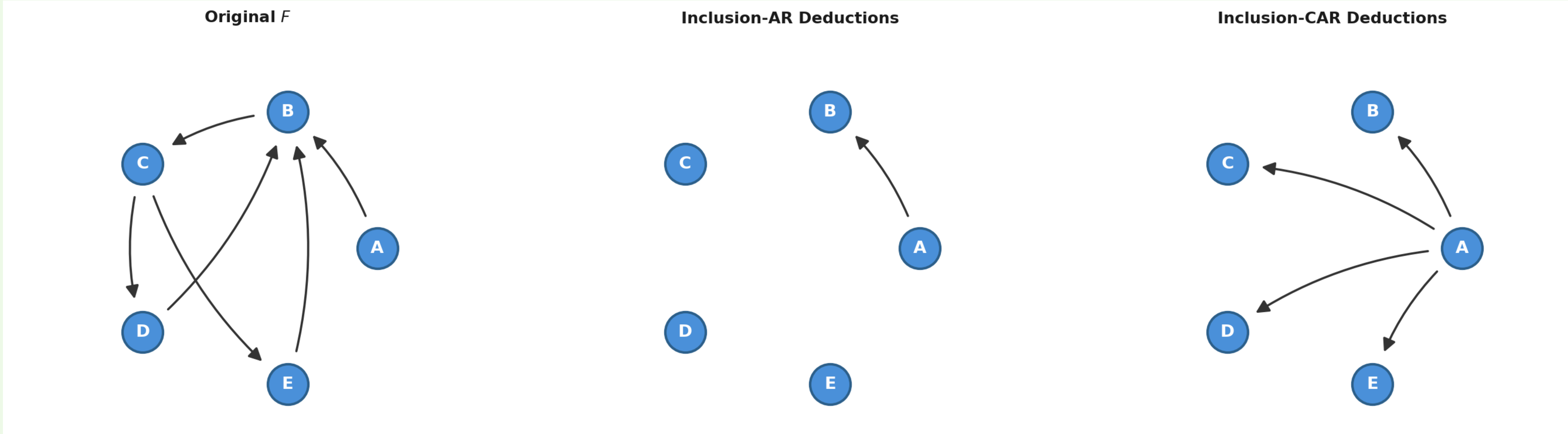


High constant noise: cardinality-AR **outperforms** logistic regression, being robust to distant mislabelled points. **Logistic noise:** comparable accuracy.

Pairwise Preferences

Goal: infer missing pairwise comparisons from N assessors expressing partial preferences over n items, under cycle-induced inconsistency, i.e., given an assesor k and a pair of items (i, j), try to deduce $i \succ j$ or $j \succ i$.

The preference graph for each assessor contains items as nodes and comparisons as directed edges; cycles represent violations of transitivity and inconsistency.



- **Data:** $\mathcal{D} = \{(x_i, y_i)\}_i$, where $\mathcal{X} = \{(i, j) : i < j \le n\}$ represents item pairs and $\mathcal{Y} = \{0, 1\}$ comparison outcomes: $y = 0$ means $i \succ j$, while $y = 1$ means $j \succ i$.
- **Deterministic model:** a ranking $\boldsymbol{\rho} \in S_n$ determines all pairwise preferences:

$$P_{\boldsymbol{\rho}}(Y = y \mid X = (i, j)) = \mathbf{1}\{(\rho_i < \rho_j, y = 1) \lor (\rho_j < \rho_i, y = 0)\}.$$

- **Error model: Mallows model with Bernoulli noisy pair comparisons.**

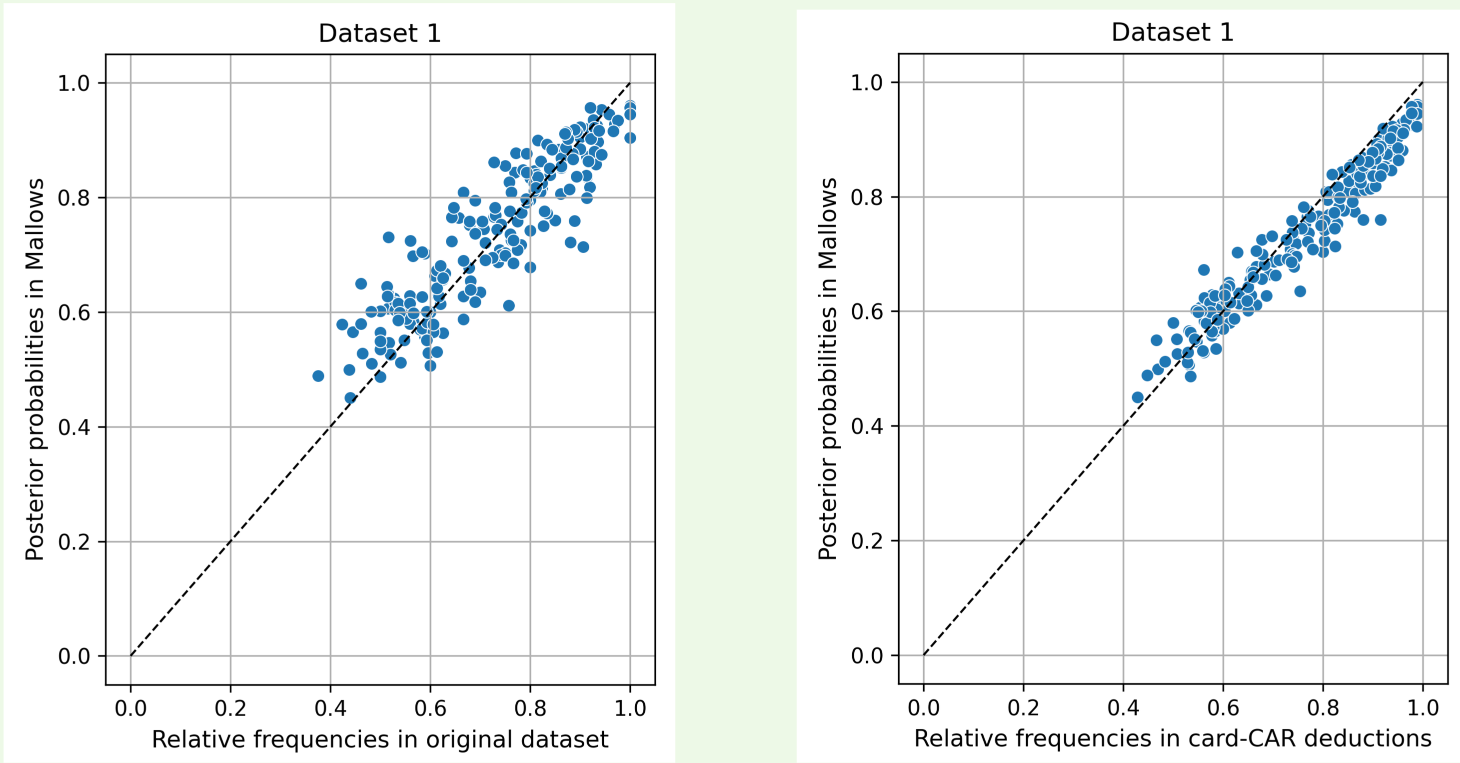
Each assessor k has a latent ranking $\boldsymbol{\pi}_k$ sampled around a consensus ranking $\boldsymbol{\rho}$:

$$P(\boldsymbol{\pi}_k \mid \boldsymbol{\rho}, \alpha) \propto \exp(-\alpha \cdot d(\boldsymbol{\pi}_k, \boldsymbol{\rho})) .$$

Pair comparisons are generated from $\boldsymbol{\pi}_k$ and flipped with error probability $p \le 1/2$. Bayesian inference estimates $\{\boldsymbol{\pi}_k\}_{k=1}^N$, $\boldsymbol{\rho}$, and p to predict missing comparisons.

- **Repairs:** maximal acyclic subgraphs of the preference graph.
  - **Inclusion-AR:** cycle hitting-set feasibility (coNP-hard).
  - **Cardinality-AR:** Minimum Feedback Arc Set (NP-hard).
  - **CAR:** polynomial post-processing of AR deductions.
- **Experiments:** $N = 200$ assessors, $n = 20$ items, with inconsistencies wrt full rankings.

We compare Mallows estimated probabilities for each pair comparison with: relative frequencies in the original dataset and in the cardinality-CAR deductions dataset.



CAR deductions are **more correlated** with Mallows posteriors than raw comparisons, indicating compatible repairs. Under **low inconsistency**, Mallows may over-pool assessors and favour non-optimal repairs. Under **high inconsistency**, this effect is reduced, and Mallows prioritizes optimal repairs over population pooling.

Conclusions and Future Work

- Statistical inconsistency and logical unsatisfiability are **formally equivalent** under $\mathcal{L}_\mathcal{M}$.
- **Cardinality-based repairs** are the logical counterpart of **likelihood-based error minimisation**: a bridge between the two fields.
- The two paradigms act at **different levels**: statistics is *inferential and completion-oriented* – it assumes latent structure with noise; logic is *conservative but structurally expressive* – it retains only what is deductively entailed.
- The approaches are **complementary**: logic for deterministic rules, statistics for population-level uncertainty.

Future work: richer statistical models (continuous variables); structured observations (disjunctions, implications, rank constraints) beyond atomic facts; population-level repair semantics; broader notions of statistical inconsistency.

Norwegian Centre of Excellence

[1]Department of Mathematics, UiO [2]Department of Informatics, UiO
[3]Oslo Centre for Biostatistics and Epidemiology, UiO