Bayesian Posterior Estimation of Gene Regulatory Networks from Zero-Inflated Single-Cell Data

Marco Scutari
scutari@bnlearn.com

Dalle Molle Institute for
Artificial Intelligence (IDSIA)

August 24, 2026

→ Informative Graphical Priors

Parameter Estimation with Zero-Inflated Counts

Informative Priors from Annotation Databases

We have much prior information in annotation databases (think KEGG). Why not use it?

1. Look up the pathways linking your proteins, and construct a (tentative, possibly partial) a priori network $\mathcal{G}_p$.
2. Define a loss function measuring structural differences from $\mathcal{G}_p$.
3. Place the loss in an energy-based prior a la Imoto *et al.* [2]:

$$\pi\left(\mathcal{G}\right) \propto \exp\left\{-\beta E\left(\mathcal{G}\right)\right\}, \beta \in \mathbb{R}^{+}.$$

Posterior inference then follows as usual, $\mathrm{P}\left(\mathcal{G} \mid \mathcal{D}\right) \propto \pi\left(\mathcal{G}\right) \mathrm{P}\left(\mathcal{D} \mid \mathcal{G}\right)$.

The question: Which energy function $E(\cdot)$?

LGE: a Causal Informative Prior

$$\pi(\mathcal{G}) \propto \exp\left\{-\beta\left[w\widetilde{\mathrm{SHD}}(\mathcal{G}, \mathcal{G}_p) + (1-w)\widetilde{\mathrm{SID}}(\mathcal{G}, \mathcal{G}_p)\right]\right\}$$

A local-global prior (LGE) centred on $\mathcal{G}_p$:

- Local: SHD penalises arc-by-arc [5], but not the effect of a misplaced arc on the causal structure.
- Global: SID penalises differences in interventional pathways [3], but misplaced arcs arcs may not affect them.
- Hyperparameters: $\beta \in \mathbb{R}^+$ is a decay parameter, $w \in [0, 1]$ weights local against global.

Both on a common scale, so $w \in [0, 1]$ makes sense:

$$\widetilde{\mathrm{SHD}} = \mathrm{SHD}\,/|A_p| \qquad \text{and} \qquad \widetilde{\mathrm{SID}} = \mathrm{SID}\,/[N(N-1)].$$

Properties

Normalising constant. Let $\mathcal{G}_p = (V, A_p)$ with N nodes, $P = N(N-1)/2$, and $q = \exp\{-\beta w/|A_p|\}$. Then, for all $\beta \geqslant 0$ and $w \in [0,1]$,

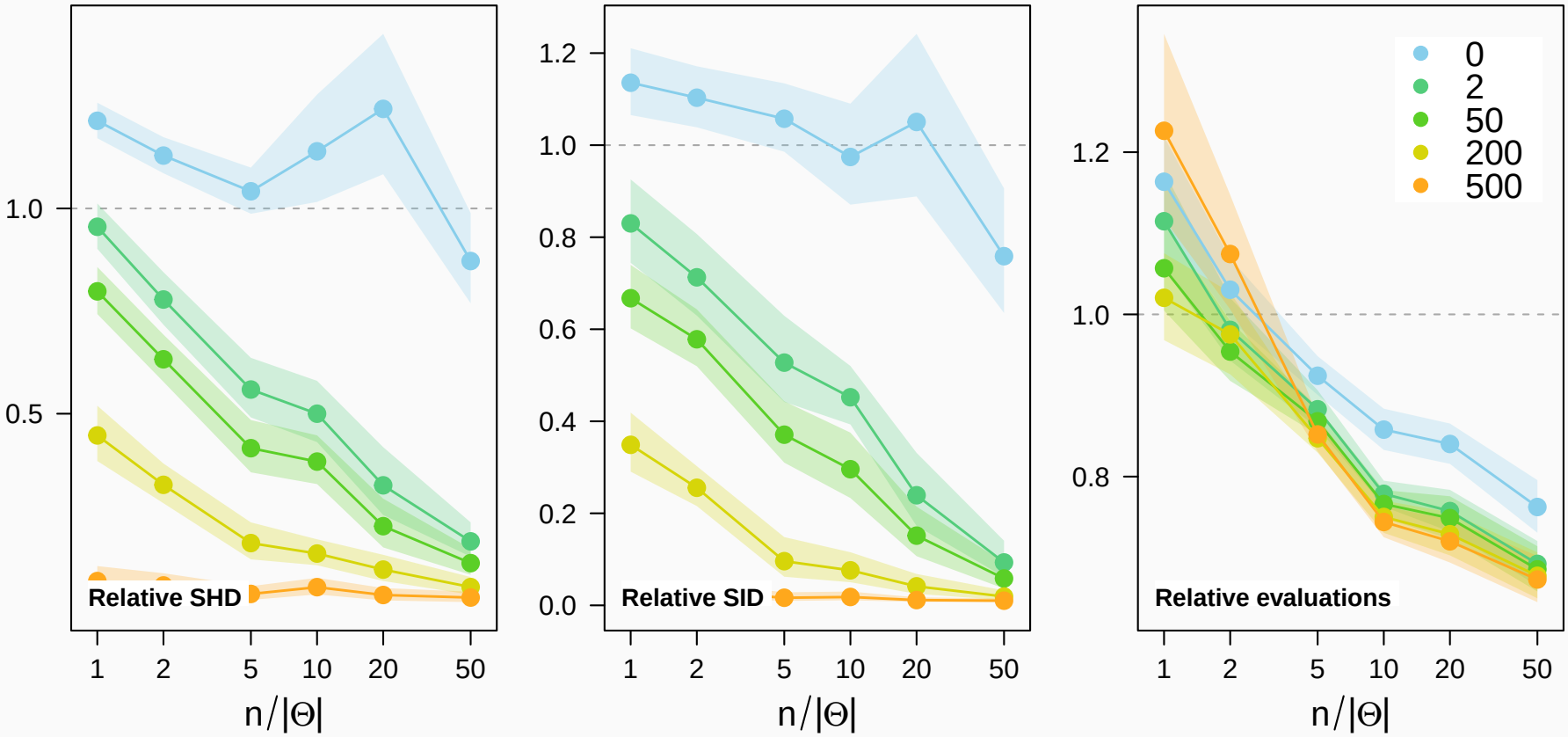
$$\max\left\{(1+q)^{P-|A_p|},\ e^{-\beta(1-w)}(1+q)^P\right\} \leqslant Z(\beta, w) \leqslant \min\left\{(1+2q)^P,\ N!\,(1+q)^P\right\}.$$

Calibrating β. Define the ball $B(\mathcal{G}_p) = \{\mathcal{G} : \mathrm{SHD}(\mathcal{G}, \mathcal{G}_p) = 1\}$, containing $O(N^2)$ DAGs. Inside, $w/|A_p| \leqslant E(\mathcal{G}, \mathcal{G}_p; w) \leqslant w/|A_p| + 2(1-w)/N$.

- $\pi(B(\mathcal{G}_p)) \leqslant r\pi(\mathcal{G}_p) \Longrightarrow \beta \geqslant \frac{|A_p|}{w} \log \frac{|B(\mathcal{G}_p)|}{r}$.
- $\pi(B(\mathcal{G}_p)) = p \Longrightarrow \beta \geqslant \frac{|A_p|}{w} \log \frac{|B(\mathcal{G}_p)|}{(1-p)/p}$.

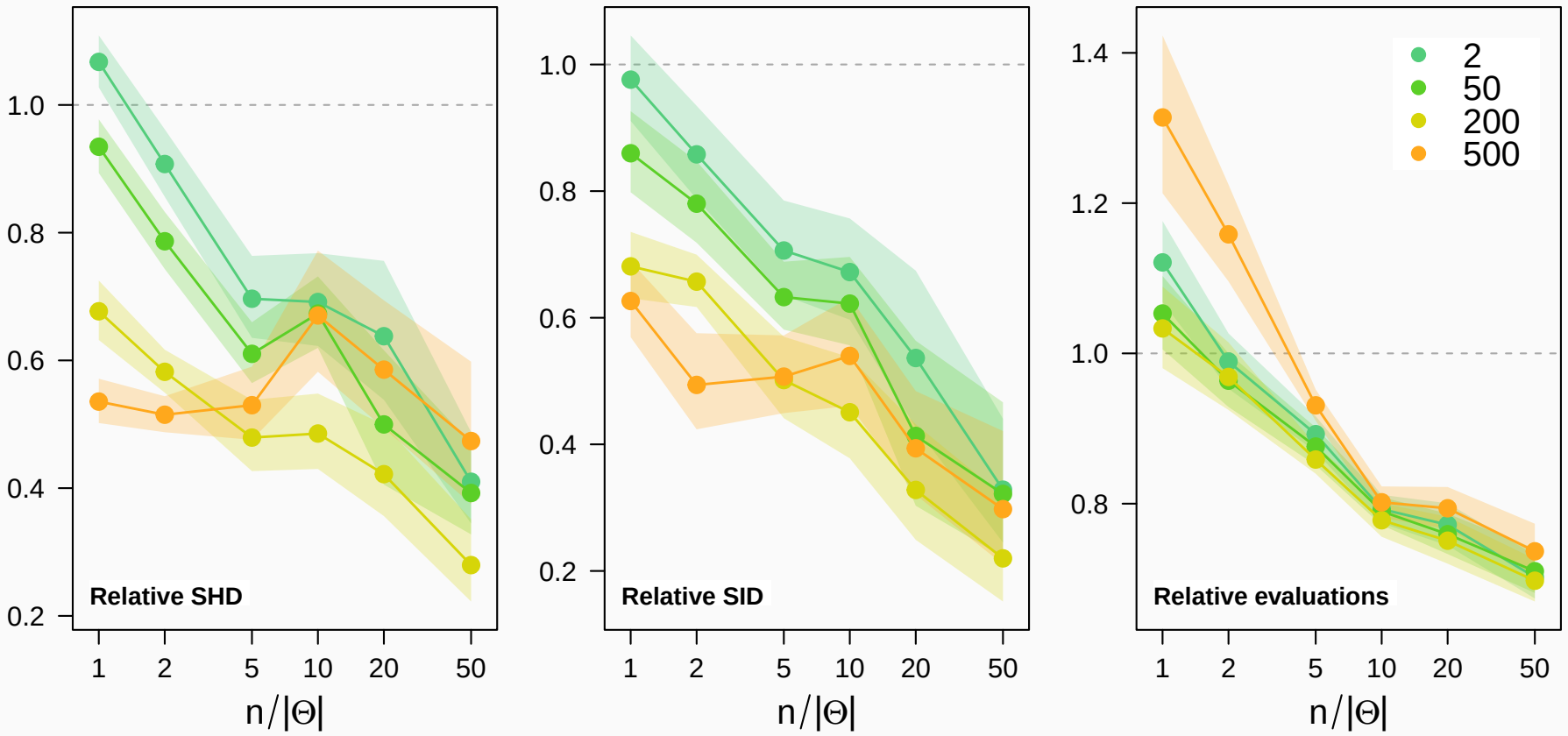
Calibrating w. Balanced: $w^* = \frac{2|A_p|}{N+2|A_p|} = \frac{2\bar{d}}{1+2\bar{d}}$, where $\bar{d} = \frac{|A_p|}{N}$.

Performance: $\mathcal{G}_p$ is Correct



Reference: marginal uniform prior [4], enforcing sparsity by controlling the single-arc inclusion probability.

Performance: $\mathcal{G}_p$ is 50% Wrong



Reference: marginal uniform prior.

Prior corrupted with $0.5|A_p|$ single-arc operations.

✓ Informative Graphical Priors

→ Parameter Estimation with Zero-Inflated Counts

Expectation Conditional Maximisation

The reference paper on zero-inflated count DAGs [1] estimates MLEs with BFGS, gradient and no Hessian, which is numerically unstable.

A better way: expectation-conditional-maximisation (ECM), for a 10x speedup and $10^{-1} \rightarrow 10^{-3}$ relative error in the ML because many numerical optimisations become possible.

1. E-Step: Estimate the structural zero indicators.
2. M-Step:
   - Single-step IRLS: logistic regression for the zero-inflation probability.
   - Single-step IRLS: GLM modelling the count process.

Each step increases Q, so the observed-data likelihood converges to the same stationary point as a full EM like [6].

Hierarchical Estimation

Why estimate the probability of zero-inflation separately for each X_i?

Structural zeros

- appear in blocks along (super-)pathway/group hierarchical lines;
- appear more frequently in some ontology groups than others.

Thus, we can define a pathway-/group-level inactivity prior probability

$$\theta_p \sim Beta(\mu\kappa, (1-\mu)\kappa)$$

and a gene-level inactivity prior probability

$$\zeta_i \sim Beta(\theta_p\kappa, (1-\theta_p)\kappa),$$

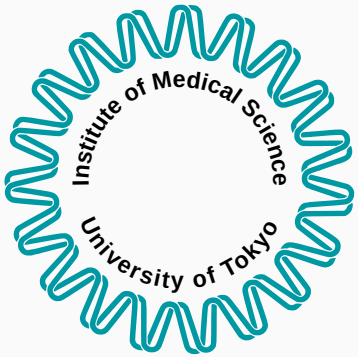
where κ is a precision parameter and θ is a position parameter ($\theta_p \to 0 \Rightarrow \zeta_i = 0, \theta_p \to 1 \Rightarrow \zeta_i = 1$). This shares information across the genes in a pathway: ζ_i shrinks towards θ_p when data are scarce.

Ongoing and Future Work

Many interesting problems that Bayesian statistics is well-suited for!

- Arc types and signs from annotation databases: activation and inhibition should also be in the prior.
- Hierarchical grouping: modelling functional groups and omic types in addition to pathways.
- Data naturally modelled with parametric distributions for zero-inflated count data with over-dispersion.
- Much of the literature still has low scalability and no uncertainty quantification.

Acknowledgements



Noriaki Sato

Seiya Imoto

The Institute of Medical Science, The University of Tokyo

Shuichi Kawano

Faculty of Mathematics, Kyushu University

Rui Yamaguchi

Aichi Cancer Center &

Graduate School of Medicine, Nagoya University

That's all!

Happy to discuss in more detail.

scutari@bnlearn.com

References I

- J. Choi and Y. Ni.
  Model-Based Causal Discovery for Zero-Inflated Count Data.
  Journal of Machine Learning Research, 24(200):1–32, 2023.
- S. Imoto, T. Higuchi, T. Goto, K. Tashiro, S. Kuhara, and S. Miyano.
  Combining Microarrays and Biological Knowledge for Estimating Gene Networks via Bayesian Networks.
  Journal of Bioinformatics and Computational Biology, 2(1):77–98, 2004.
- J. Peters and P. Bühlmann.
  Structural Intervention Distance (SID) for Evaluating Causal Graphs.
  Neural Computation, 27:771–799, 2015.
- M. Scutari.
  An Empirical-Bayes Score for Discrete Bayesian Networks.
  Journal of Machine Learning Research (Proceedings Track, PGM 2016), 52:438–448, 2016.
- I. Tsamardinos, L. E. Brown, and C. F. Aliferis.
  The Max-Min Hill-Climbing Bayesian Network Structure Learning Algorithm.
  Machine Learning, 65(1):31–78, 2006.

References II

- Z. Wang, S. Ma, C. Wang, M. Zappitelli, P. Devarajan, and C. Parikh.
  EM for Regularized Zero-Inflated Regression Models with Applications to Postoperative Morbidity After Cardiac Surgery in Children.
  Statistics in Medicine, 33(29):5192–5208, 2014.