Neyman-Pearson Causal Inference
Supplementary Material

Joni Shaska
University of Southern California
shaska@usc.edu

Urbashi Mitra
University of Southern California
ubli@usc.edu

We present the full proofs of propositions, as well as a time-series example to illustrate the use of the genie bound.

Proposition 1. *Let U be a uniform random variable on $[0,1]$. Then, the optimal detector $\hat{\boldsymbol{\chi}}^*$ for the Neyman-Pearson causal discovery problem is given as follows,*

$$\hat{\boldsymbol{\chi}}^*_{i,j} = \begin{cases} 0, & \frac{dP_{i,j}}{dQ_{i,j}} > \frac{w^-_{i,j}}{w^+_{i,j}}\gamma \\ 0, & \frac{dP_{i,j}}{dQ_{i,j}} = \frac{w^-_{i,j}}{w^+_{i,j}}\gamma \text{ and } U \le \eta \\ 1, & \frac{dP_{i,j}}{dQ_{i,j}} < \frac{w^-_{i,j}}{w^+_{i,j}}\gamma \\ 1, & \frac{dP_{i,j}}{dQ_{i,j}} = \frac{w^-_{i,j}}{w^+_{i,j}}\gamma \text{ and } U > \eta \end{cases} \tag{1}$$

where γ and $\eta \in [0,1]$ are chosen so that $\epsilon^+ = \epsilon$. ■

Proof. We proceed first by showing existence. Let $\gamma \ge 0$ be the largest γ such that

$$\sum_{i,j} w^+_{i,j} P_{i,j}\left(\frac{dP_{i,j}}{dQ_{i,j}} < \gamma\right) \le \epsilon. \tag{2}$$

Then, if the inequality is strict, select η to be

$$\eta = \frac{\epsilon - \sum_{i,j} w^+_{i,j} P_{i,j}\left(\frac{dP_{i,j}}{dQ_{i,j}} < \frac{w^-_{i,j}}{w^+_{i,j}}\gamma\right)}{\sum_{i,j} w^+_{i,j} P_{i,j}\left(\frac{dP_{i,j}}{dQ_{i,j}} = \frac{w^-_{i,j}}{w^+_{i,j}}\gamma\right)} \tag{3}$$

otherwise choose η arbitrarily. Hence, we have that

$$\begin{aligned} \epsilon^+ = \sum_{i,j} w^+_{i,j} P^+_{i,j} = \sum_{i,j} w^+_{i,j}\Bigg(P_{i,j}\left(\frac{dP_{i,j}}{dQ_{i,j}} < \frac{w^-_{i,j}}{w^+_{i,j}}\gamma\right) \\ + \eta P_{i,j}\left(\frac{dP_{i,j}}{dQ_{i,j}} = \frac{w^-_{i,j}}{w^+_{i,j}}\gamma\right)\Bigg) \end{aligned} \tag{4}$$

$$\begin{aligned} = \sum_{i,j} w^+_{i,j} P_{i,j}\left(\frac{dP_{i,j}}{dQ_{i,j}} < \frac{w^-_{i,j}}{w^+_{i,j}}\gamma\right) \\ + \eta \sum_{i,j} w^+_{i,j} P_{i,j}\left(\frac{dP_{i,j}}{dQ_{i,j}} = \frac{w^-_{i,j}}{w^+_{i,j}}\gamma\right) = \epsilon. \end{aligned} \tag{5}$$

We next show that threshold rules are optimal. We use the Lagrange multiplier $\lambda_0 > 0$ (if the optimal estimator is such that $\lambda_0 = 0$ then the constraint is inactive, and so the optimal estimator should always declare $\hat{\boldsymbol{\chi}}_{i,j} = 1$) and seek to minimize $\epsilon^- + \lambda_0\epsilon^+$. Then, we have

$$\begin{aligned} \epsilon^- + \lambda_0\epsilon^+ = \sum_{i,j} w^-_{i,j}\mathbb{P}(\hat{\boldsymbol{\chi}}_{i,j} = 0|\boldsymbol{\chi}_{i,j} = 1) \\ + \lambda_0 \sum_{i,j} w^+_{i,j}\mathbb{P}(\hat{\boldsymbol{\chi}}_{i,j} = 1|\boldsymbol{\chi}_{i,j} = 0) \end{aligned} \tag{6}$$

$$\begin{aligned} = \sum_{i,j}\int_{X_1,\dots,X_n} w^-_{i,j}\mathbb{P}(\hat{\boldsymbol{\chi}}_{i,j} = 0|X_1,\dots,X_n)Q_{i,j}(X_1,..,X_n) \\ + \lambda_0 w^+_{i,j}\mathbb{P}(\hat{\boldsymbol{\chi}}_{i,j} = 1|X_1,\dots,X_n)P_{i,j}(X_1,..,X_n). \end{aligned} \tag{7}$$

Then, to minimize $\epsilon^- + \lambda_0\epsilon^+$, the estimator $\hat{\boldsymbol{\chi}}$ should be such that $\hat{\boldsymbol{\chi}}_{i,j} = 0$ if $w^+_{i,j}Q_{i,j} \le \lambda_0 w^-_{i,j}P_{i,j}(X_1,\dots,X_n)$ and $\hat{\boldsymbol{\chi}}_{i,j} = 1$ otherwise. Since threshold rules are optimal, it remains to be seen how to select the threshold γ. Observe that the probabilities $P_{i,j}\left(\frac{dP_{i,j}}{dQ_{i,j}} < \frac{w^-_{i,j}}{w^+_{i,j}}\gamma\right)$ are increasing in γ for all pairs (i,j), and that the probabilities $Q_{i,j}\left(\frac{dP_{i,j}}{dQ_{i,j}} > \frac{w^-_{i,j}}{w^+_{i,j}}\gamma\right)$ are decreasing in γ. Hence, ϵ^+ is increasing in γ, and ϵ^- is decreasing in γ. Then, to minimize ϵ^-, γ should be chosen to make ϵ^+ as large as possible. □

We repeat some definitions here.

Definition 1. *The Rényi divergence of order λ between two probability measures P and Q is given as*

$$D_\lambda(P||Q) \doteq \frac{1}{\lambda - 1}\log\int_{\mathcal{X}}\left(\frac{dP}{dQ}\right)^\lambda dQ, \tag{8}$$

where $\frac{dP}{dQ}$ is the Radon-Nikodym derivative of P with respect to Q.

Definition 2. *For any two probability measures P and Q, let*

$$D'_\lambda(P||Q) \doteq \int_{\mathcal{X}} F_\lambda(x;P,Q)\log\frac{dP}{dQ} \tag{9}$$

$$D''_\lambda(P||Q) \doteq \int_{\mathcal{X}} F_\lambda(x;P,Q)\bigg(\log\frac{dP}{dQ}\bigg)^2 - \Big(D'_\lambda(P||Q)\Big)^2 \tag{10}$$

$$\text{where} \quad F_\lambda(x;P,Q) \doteq \frac{f(x)^\lambda g(x)^{1-\lambda}}{\int_{\mathcal{X}} f(x)^\lambda g(x)^{1-\lambda}d\mu} \quad \text{for} \quad \lambda \in [0,1] \tag{11}$$

This work has been funded in part by one or more of the following grants: ONR 00014-15-1-2550, NSF CCF-1817200, ARO W911NF1910269, Cisco Foundation 1980393, DOE DE-SC0021417, Swedish Research Council 2018-04359, NSF CCF-2008927, ONR 503400-78050.

Proposition 2. *For the detector $\hat{\boldsymbol{\chi}}^*$ given in Theorem 1, we have for any $\lambda \in [0,1]$*

$$\epsilon^- \leq \sum_{i,j} (w_{i,j}^-)^{1-\lambda}(w_{i,j}^+)^{\lambda}\frac{1}{\gamma^\lambda}e^{-(1-\lambda)D_\lambda(P_{i,j}||Q_{i,j})}, \quad (12)$$

where $D_\lambda(P_{i,j}||Q_{i,j})$ is the the Rényi divergence of order λ between the series of distributions $P_{i,j}$ and $Q_{i,j}$ which is defined in the Appendix.

Proof of Proposition 2. for any i,j, we have

$$\begin{aligned} Q_{i,j}^- &\leq Q_{i,j}\Big(\frac{dP_{i,j}}{dQ_{i,j}} \geq \frac{w_{i,j}^-}{w_{i,j}^+}\gamma\Big) && (13)\\ &= \int_{\mathcal{X}} \mathbb{1}\Big\{\frac{dP_{i,j}}{dQ_{i,j}} \geq \frac{w_{i,j}^-}{w_{i,j}^+}\gamma\Big\}dQ_{i,j} && (14)\\ &\overset{(a)}{\leq} \int_{\mathcal{X}} \Big(\frac{w_{i,j}^+}{w_{i,j}^1}\frac{1}{\gamma}\frac{dP_{i,j}}{dQ_{i,j}}\Big)^\lambda dQ_{i,j} && (15)\\ &= \Big(\frac{w_{i,j}^+}{w_{i,j}^+}\Big)^\lambda \frac{1}{\gamma^\lambda}\int_{\mathcal{X}}\Big(\frac{dP_{i,j}}{dQ_{i,j}}\Big)^\lambda dQ_{i,j}, && (16)\end{aligned}$$

where (a) holds since $\mathbb{1}\{a \geq \gamma\} \leq \left(\frac{a}{\gamma}\right)^\lambda$ for any $a, \gamma, \lambda > 0$. Since, the above holds for any $\lambda > 0$. Multiplying by $w_{i,j}^-$ yields

$$w_{i,j}^- Q_{i,j}^- \leq (w_{i,j}^-)^{1-\lambda}(w_{i,j}^+)^\lambda \frac{1}{\gamma^\lambda}e^{-(1-\lambda)D_\lambda(P_{i,j}||Q_{i,j})}. \quad (17)$$

Summing over i,j completes the proof. □

Proposition 3. *For any detector $\hat{\boldsymbol{\chi}}$ that satisfies $\epsilon^+ \leq \epsilon$, we have that for any $\lambda \in [0,1]$.*

$$\begin{aligned}\epsilon^- \geq \frac{1}{2}\sum_{i,j} w_{i,j}^- \exp\big\{&-(1-\lambda)D_\lambda(P_{i,j}||Q_{i,j})\\ &-\lambda D'_\lambda(P_{i,j}||Q_{i,j}) - \lambda\sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}\big\}\\ &-\epsilon\max_{i,j}\Big\{\frac{w_{i,j}^-}{w_{i,j}^+}\exp\big\{-D'_\lambda(P_{i,j}||Q_{i,j})\\ &+(1-2\lambda)\sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}\big\}\Big\}. \quad (18)\end{aligned}$$

Proof. It is relatively easy to show that for any pair (i,j),

$$\begin{aligned} Q_{i,j}(X_1,..,X_n) = \exp\{&-(1-\lambda)D_\lambda(P_{i,j}||Q_{i,j}) && (19)\\ &-\lambda\log\frac{dP_{i,j}}{dQ_{i,j}}\}F_\lambda(X_1,...,X_n;P_{i,j},Q_{i,j}), && (20)\\ P_{i,j}(X_1,..,X_n) = \exp\{&-(1-\lambda)D_\lambda(P_{i,j}||Q_{i,j}) && (21)\\ &+(1-\lambda)\log\frac{dP_{i,j}}{dQ_{i,j}}\}F_\lambda(X_1,...,X_n;P_{i,j},Q_{i,j}). && (22)\end{aligned}$$

Then, consider the following sets

$$\mathcal{X}_{i,j,\lambda} = \Big\{X_1,...,X_n : |\log\frac{dP_{i,j}}{dQ_{i,j}} - D'_\lambda(P_{i,j}||Q_{i,j})| \leq \sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}\Big\}, \quad (23)$$

$$\mathcal{X}_{i,j}^0 = \Big\{X_1,...,X_n : \hat{\boldsymbol{\chi}}_{i,j} = 0\Big\}, \quad (24)$$

$$\mathcal{X}_{i,j}^1 = \Big\{X_1,...,X_n : \hat{\boldsymbol{\chi}}_{i,j} = 1\Big\}. \quad (25)$$

Hence, for any $X_1,...,X_n \in \mathcal{X}_{i,j,\lambda}$, we have that

$$Q_{i,j}(X_1,..,X_n) \geq \Big(e^{-(1-\lambda)D_\lambda(P_{i,j}||Q_{i,j})-\lambda D'_\lambda(P_{i,j}||Q_{i,j})-\lambda\sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}}\Big) F_\lambda(X_1,...,X_n;P_{i,j},Q_{i,j}), \quad (26)$$

$$P_{i,j}(X_1,..,X_n) \geq \Big(e^{-(1-\lambda)D_\lambda(P_{i,j}||Q_{i,j})+(1-\lambda)D'_\lambda(P_{i,j}||Q_{i,j})-(1-\lambda)\sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}}\Big) F_\lambda(X_1,...,X_n;P_{i,j},Q_{i,j}). \quad (27)$$

Notice that $D'_\lambda(P_{i,j}||Q_{i,j})$ is actually the mean of $\log\frac{dP_{i,j}}{dQ_{i,j}}$ with respect to the distribution F_λ, and $D''_\lambda(P_{i,j}||Q_{i,j})$ is its variance. Then, $\mathcal{X}_{i,j,\lambda}$ is the event that the log-likelihood ratio is within $\sqrt{2}$ standard deviations of its mean, and so from Chebychev's inequality,

$$\int_{\mathcal{X}_{i,j,\lambda}} F_\lambda(X_1,...,X_n;P_{i,j},Q_{i,j}) \geq \frac{1}{2}. \quad (28)$$

Then, from the union bound, we have that

$$\int_{\mathcal{X}_{i,j,\lambda}\cap\mathcal{X}_{i,j}^0} F_\lambda(X_1,...,X_n;P_{i,j},Q_{i,j}) + \int_{\mathcal{X}_{i,j,\lambda}\cap\mathcal{X}_{i,j}^1} F_\lambda(X_1,...,X_n;P_{i,j},Q_{i,j}) \geq \frac{1}{2}. \quad (29)$$

From (27), we have that

$$\begin{aligned}&\int_{\mathcal{X}_{i,j,\lambda}\cap\mathcal{X}_{i,j}^1} F_\lambda(X_1,...,X_n;P_{i,j},Q_{i,j}) && (30)\\ &\leq e^{(1-\lambda)D_\lambda(P_{i,j}||Q_{i,j})-(1-\lambda)D'_\lambda(P_{i,j}||Q_{i,j})+(1-\lambda)\sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}}\int_{\mathcal{X}_{i,j,\lambda}\cap\mathcal{X}_{i,j}^1} P_{i,j}(X_1,...,X_n) && (31)\\ &\leq e^{(1-\lambda)D_\lambda(P_{i,j}||Q_{i,j})-(1-\lambda)D'_\lambda(P_{i,j}||Q_{i,j})+(1-\lambda)\sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}} P_{i,j}^+ && (32)\end{aligned}$$

and so

$$\int_{\mathcal{X}_{i,j,\lambda}\cap\mathcal{X}_{i,j}^0} F_\lambda(X_1,...,X_n;P_{i,j},Q_{i,j}) \geq \frac{1}{2} - e^{(1-\lambda)D_\lambda(P_{i,j}||Q_{i,j})-(1-\lambda)D'_\lambda(P_{i,j}||Q_{i,j})+(1-\lambda)\sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}} P_{i,j}^+. \tag{33}$$

Then, (26) gives

$$Q_{i,j}^- \geq e^{-(1-\lambda)D_\lambda(P_{i,j}||Q_{i,j})-\lambda D'_\lambda(P_{i,j}||Q_{i,j})-\lambda\sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}} \int_{\mathcal{X}_{i,j,\lambda}\cap\mathcal{X}_{i,j}^0} F_\lambda(X_1,...,X_n;P_{i,j},Q_{i,j}), \tag{34}$$

together with (33) we have

$$Q_{i,j}^- \geq \frac{1}{2}e^{-(1-\lambda)D_\lambda(P_{i,j}||Q_{i,j})-\lambda D'_\lambda(P_{i,j}||Q_{i,j})-\lambda\sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}} - e^{-D'_\lambda(P_{i,j}||Q_{i,j})+(1-2)\lambda\sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}} P_{i,j}^+. \tag{35}$$

Multiplying both sides by $w_{i,j}^-$ and summing over all (i,j) pairs gives us

$$\sum_{i,j} w_{i,j}^- Q_{i,j}^- \geq \frac{1}{2}\sum_{i,j} w_{i,j}^- e^{-(1-\lambda)D_\lambda(P_{i,j}||Q_{i,j})-\lambda D'_\lambda(P_{i,j}||Q_{i,j})-\lambda\sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}} - \sum_{i,j} w_{i,j}^+ \frac{w_{i,j}^-}{w_{i,j}^+} e^{-D'_\lambda(P_{i,j}||Q_{i,j})+(1-2)\lambda\sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}} P_{i,j}^+, \tag{36}$$

which yields

$$\sum_{i,j} w_{i,j}^- Q_{i,j}^- \geq \frac{1}{2}\sum_{i,j} w_{i,j}^- e^{-(1-\lambda)D_\lambda(P_{i,j}||Q_{i,j})-\lambda D'_\lambda(P_{i,j}||Q_{i,j})-\lambda\sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}} - \max_{i,j}\left\{\frac{w_{i,j}^-}{w_{i,j}^+} e^{-D'_\lambda(P_{i,j}||Q_{i,j})+(1-2)\lambda\sqrt{2D''_\lambda(P_{i,j}||Q_{i,j})}}\right\} \sum_{i,j} w_{i,j}^+ P_{i,j}^+. \tag{37}$$

Noticing that $\sum_{i,j} w_{i,j}^+ P_{i,j}^+ = \epsilon^+ \leq \epsilon$ completes the proof. $\square$

I. Time-Series Example

As the optimal detector averages over all possible permutations of active edges, it incurs high computational complexity. This complexity can become prohibitive in more complex scenarios. Moving beyond the two-node case, we consider larger graphs and time-series data. Consider the time series problem described by

$$X_{k+1} = \boldsymbol{A}X_k + W_k, \tag{38}$$

where the matrix $\boldsymbol{A}$ has a spectral radius less than one, the exogeneous terms W_k are *i.i.d.* Gaussian random variables with variance σ^2, and X_1 is drawn from the stationary multivariate Gaussian distribution with zero mean and covariance matrix Γ that satisfies $\boldsymbol{\Gamma} = \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^\top + \sigma^2 I$. We let the matrix $\boldsymbol{A}$ have the *dynamic ER prior distribution*: for any $p \in \mathbb{N}$, let $\boldsymbol{R} \in \{-1,1\}^{p\times p}$ consist of independent Rademacher random variables, and so $\boldsymbol{R}$ denotes the signs of potential edges in the graph. For the support of $\boldsymbol{A}$, let $\boldsymbol{\chi} \in \{0,1\}^{p\times p}$ consist of *i.i.d.* Bernoulli random variables each with mean ν. Then, define $\boldsymbol{S} = \boldsymbol{R}\circ\boldsymbol{\chi}$. Finally, let $\sigma(\boldsymbol{S})$ denote the spectral radius of $\boldsymbol{S}$, and so $\boldsymbol{A}$ is defined as

$$\boldsymbol{A} = \begin{cases} r_0\frac{\boldsymbol{S}}{r(\boldsymbol{S})}, & \sigma(\boldsymbol{S}) \neq 0 \\ \boldsymbol{S}, & \sigma(\boldsymbol{S}) = 0, \end{cases} \tag{39}$$

where r_0 is the desired spectral radius of $\boldsymbol{A}$. Note that the dynamic ER distribution is symmetric in the sense that $w_{i,j}^- = w_{i,j}^+ = \frac{1}{p^2}$.

To compute $\hat{\boldsymbol{\chi}}^*$ for this prior, one must compute the likelihood ratio $\frac{dP_{i,j}}{dQ_{i,j}}$, which, as in the previous example, consists of computing multi-dimensional integrals over Gaussian mixtures whose weights are determined by the dynamic ER prior. Hence, even numerical estimation is intractable. While computing $\hat{\boldsymbol{\chi}}^*$ is non-trivial, The genie bound still provides a useful lower bound that can be easily computed. We restate the result.

Assumption 1. *For each (i,j), there exists a random variable S with distribution α_S such that for all X,*

$$P_{i,j}(X) = \sum_s \alpha_s P_{i,j}(X|S=s) = \sum_s \alpha_s P_{i,j,s}, \tag{40}$$

$$Q_{i,j}(X) = \sum_s \alpha_s Q_{i,j}(X|S=s) = \sum_s \alpha_s Q_{i,j,s}, \tag{41}$$

where

$$P_{i,j,s} = P_{i,j}(X|S=s), \tag{42}$$

$$Q_{i,j,s} = Q_{i,j}(X|S=s) \tag{43}$$

$\square$

The assumption suggests a mixture structure of the distributions $P_{i,j}$ and $Q_{i,j}$. We exploit this assumption to determine the following lower bound on the achievable false negative rate.

Proposition 4. *Suppose Assumption 1 holds. Then, for a given tolerance level $\epsilon \in (0,1)$ and any detector $\hat{\boldsymbol{\chi}}$ that satisfies $\epsilon^+ \leq \epsilon$, we have*

$$\sum_{i,j} w_{i,j}^- \sum_s \alpha_s Q_{i,j,s}\left(\frac{dP_{i,j,s}}{dQ_{i,j,s}} > \frac{w_{i,j}^-}{w_{i,j}^+}\gamma\right) \leq \epsilon^-, \tag{44}$$

where γ is chosen so that

$$\sum_{i,j} w_{i,j}^+ \sum_s \alpha_s P_{i,j,s}\left(\frac{dP_{i,j,s}}{dQ_{i,j,s}} < \frac{w_{i,j}^-}{w_{i,j}^+}\gamma\right) = \epsilon. \tag{45}$$

$\square$

To use Proposition 4 we must identify the appropriate genie information. If we let $\boldsymbol{\chi}_{i,j}^-$ denote $\boldsymbol{\chi}$ with the (i,j)th

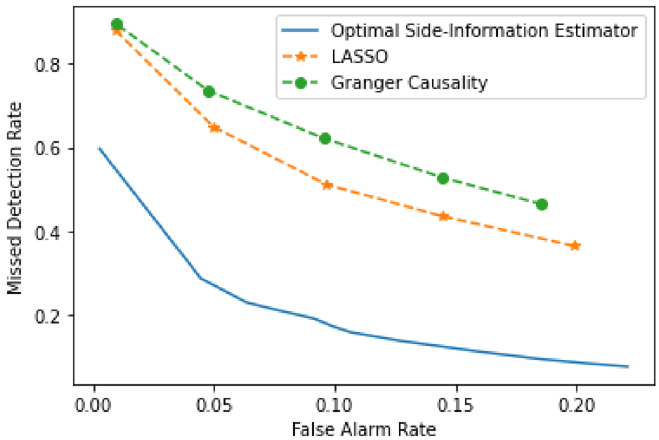

Fig. 1: Performance comparisons between LASSO and Granger causality compared to the genie bound, with $p = 10$, $n = 20$, and $\sigma^2 = 1$.

entry removed, a natural choice of the genie information is $(\boldsymbol{R}, \boldsymbol{\chi}^-_{i,j})$. If we let $\pi_{ER}(\boldsymbol{R}, \boldsymbol{\chi}^-_{i,j})$ be the prior for $\boldsymbol{R}$ and $\boldsymbol{\chi}^-_{i,j}$ given by the dynamic ER distribution, Theorem 4 gives us that for a specified tolerance ϵ,

$$\frac{1}{p^2}\sum_{i,j}\sum_{\boldsymbol{R},\boldsymbol{\chi}^-_{i,j}} \pi_{ER}(\boldsymbol{R}, \boldsymbol{\chi}^-_{i,j}) Q_{i,j,\boldsymbol{R},\boldsymbol{\chi}^-_{i,j}}\left(\frac{dP_{i,j,\boldsymbol{R},\boldsymbol{\chi}^-_{i,j}}}{dQ_{i,j,\boldsymbol{R},\boldsymbol{\chi}^-_{i,j}}} > \gamma\right) \le \epsilon^-, \tag{46}$$

where γ is chosen so that

$$\frac{1}{p^2}\sum_{i,j}\sum_{\boldsymbol{R},\boldsymbol{\chi}^-_{i,j}} \pi_{ER}(\boldsymbol{R}, \boldsymbol{\chi}^-_{i,j}) P_{i,j,\boldsymbol{R},\boldsymbol{\chi}^-_{i,j}}\left(\frac{dP_{i,j,\boldsymbol{R},\boldsymbol{\chi}^-_{i,j}}}{dQ_{i,j,\boldsymbol{R},\boldsymbol{\chi}^-_{i,j}}} < \gamma\right) = \epsilon. \tag{47}$$

Observe that when the genie information $(\boldsymbol{R}, \boldsymbol{\chi}^-_{i,j})$ is given, conditioning on $\boldsymbol{\chi}_{i,j}$ completely specifies the matrix $\boldsymbol{A}$, and so $\frac{dP_{i,j,\boldsymbol{R},\boldsymbol{\chi}^-_{i,j}}}{dQ_{i,j,\boldsymbol{R},\boldsymbol{\chi}^-_{i,j}}}$ is a likelihood ratio between Gaussian distributions, which is easily computed. Moreover, one can observe that the quantities on the left-hand side of equation (46) can be rewritten as

$$\frac{1}{p^2}\sum_{i,j}\mathbb{E}_{\boldsymbol{R},\boldsymbol{\chi}^-_{i,j}}\left[\mathbb{E}_{Q_{i,j,\boldsymbol{R},\boldsymbol{\chi}^-_{i,j}}}\left[\mathbb{1}\left\{\frac{dP_{i,j,\boldsymbol{R},\boldsymbol{\chi}^-_{i,j}}}{dQ_{i,j,\boldsymbol{R},\boldsymbol{\chi}^-_{i,j}}} > \gamma\right\}\right]\right], \tag{48}$$

which can be easily computed via Monte Carlo simulation, similarly for equation (47). We compare the genie bound to the LASSO neighborhood selection method [1] as well as Granger causality [2], which is a commonly used causal discovery algorithm for time-series data.

Briefly, Granger causality consists of a series of model comparisons. For example, consider two time series $\{X_k\}_{k=1}^n$ and $\{Y_k\}_{k=1}^n$. Then, suppose we perform a vector autoregression of $\{Y_k\}_{k=1}^n$ using $\{X_k\}_{k=1}^n$ as the explanatory variable. Then, we say that $\{X_k\}_{k=1}^n$ *Granger causes* $\{Y_k\}_{k=1}^n$ if the variance of the residuals from regressing $\{X_k\}_{k=1}^n$ onto $\{Y_k\}_{k=1}^n$ is significantly lower (compared to some threshold which is a tuning parameter) that the variance of $\{Y_k\}_{k=1}^n$ alone. The results of the comparisons can be seen in Figure 1, which shows the relative tightness of the genie bound.

REFERENCES

[1] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the Lasso," *The Annals of Statistics*, vol. 34, no. 3, pp. 1436 – 1462, 2006. [Online]. Available: https://doi.org/10.1214/009053606000000281

[2] A. B. Barrett, L. Barnett, and A. K. Seth, "Multivariate granger causality and generalized variance," *Physical review E*, vol. 81, no. 4, p. 041907, 2010.