Faculty of Science

Local Independence Graphs

Niels Richard Hansen
Department of Mathematical Sciences

October 11, 2022

Graphical models

Let $G = (\{1, \dots, N\}, E)$ be a graph consisting of vertices $\{1, \dots, N\}$ and an edge set E.

With a notion of separation between subsets of nodes we define an independence model via

$$(A, B \mid C) \in \mathcal{I}_G \Leftrightarrow B \text{ is separated from } A \text{ given } C.$$

Suppose that we also have a probabilistic independence model

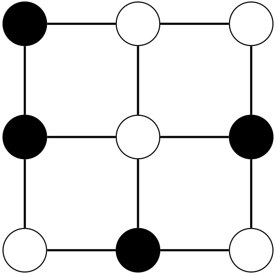
$$(A, B \mid C) \in \mathcal{I}_P \Leftrightarrow X_A \perp\!\!\!\perp_P X_B \mid X_C.$$

We say that (P, G) satisfies the global Markov property if

$$\mathcal{I}_G \subseteq \mathcal{I}_P.$$

The Ising model

$\sigma \in \{-1, 1\}^N$; here $N = 3 \times 3$

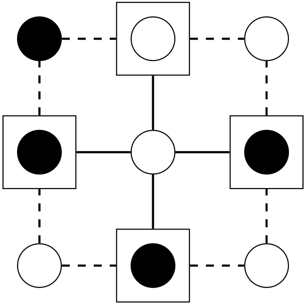


$$\pi(\sigma) \propto e^{-\beta H(\sigma)}, \qquad H(\sigma) = -\frac{1}{2}\sum_{i,j} J_{ij}\sigma_i\sigma_j$$

π factorizes according to G. J is an $N \times N$ symmetric matrix with $J_{ij} \neq 0$ only if $i - j$.

The Ising model – Markov properties

$(\{5\}, \{1,3,7,9\} \mid \{2,4,6,8\}) \in \mathcal{I}_G$



Center node, σ_5, conditionally independent of everything else given its neighbors:

$$\sigma_5 \perp\!\!\!\perp \sigma_1, \sigma_3, \sigma_7, \sigma_9 \mid \sigma_2, \sigma_4, \sigma_6, \sigma_8.$$

Global Markov property: if C separates A and B in the graph,

$$(\sigma_i)_{i \in A} \perp\!\!\!\perp (\sigma_i)_{i \in B} \mid (\sigma_i)_{i \in C}.$$

Glauber dynamics

For $\lambda_i : \{-1,1\}^N \to [0,\infty)$ and $\sigma \in \{-1,1\}^N$ define

$$\lambda_\sigma = -\sum_i \lambda_i(\sigma)$$

$$\sigma(i) = (\sigma_1, \ldots, \sigma_{i-1}, -\sigma_i, \sigma_{i+1}, \ldots, \sigma_N)$$

Consider a Markov process with state space $\{-1,1\}^N$ and intensity matrix $\Lambda = (\lambda_{\sigma,\sigma'})$ with

$$\lambda_{\sigma,\sigma'} = \begin{cases} \lambda_\sigma & \text{if } \sigma' = \sigma \\ \lambda_i(\sigma) & \text{if } \sigma' = \sigma(i) \\ 0 & \text{otherwise} \end{cases}$$

We will consider intensities defined by[1]

$$\lambda_i(\sigma) = e^{-\beta G_i(\sigma)}, \qquad G_i(\sigma) = \sum_j G_{ij}\sigma_i\sigma_j.$$

[1] Glauber. Time-Dependent Statistics of the Ising Model. J. Math. Physics. 1963

Glauber dynamics – detailed balance

π is the invariant distribution if detailed balance holds:

$$\frac{\lambda_i(\sigma(i))}{\lambda_i(\sigma)} = \frac{\pi(\sigma)}{\pi(\sigma(i))}.$$

That is, if

$$\sum_j G_{ij}\sigma_i\sigma_j = \sum_j J_{ij}\sigma_i\sigma_j.$$

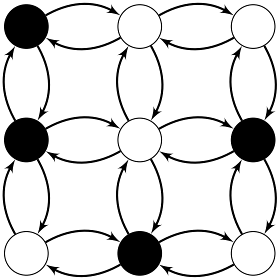
A sufficient condition for detailed balance is then

$$G = J$$

with G thus symmetric.

Detailed balance violated

We can naturally think of G as defining a directed graph

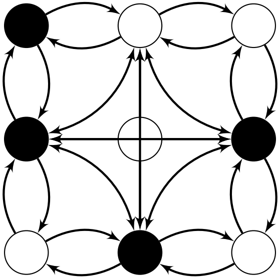


If G is asymmetric it defines a perfectly good Markov process, but detailed balance is violated – the invariant distribution is not an Ising model.

$$\lambda_i(\sigma) = e^{-\beta G_i(\sigma)}, \qquad G_i(\sigma) = \sum_j G_{ij}\sigma_i\sigma_j.$$

Marginalization

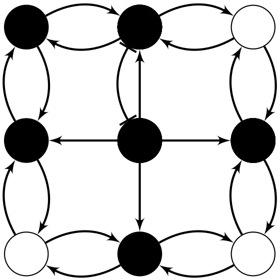
Suppose that the asymmetric G gives the following graph:



We represent marginalization by bidirected edges – but the marginalized process is not a Markov process!

Noise correlation

Glauber dynamics only allow one variable to flip at a time.

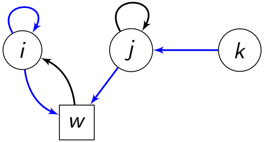


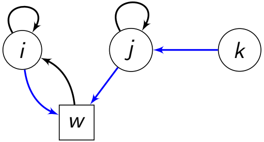
But we could have "correlated noise" so that multiple variables can flip simultaneously.

We represent this by edges with blunt marks.

Directed graphs and μ-separation

In a Directed Graph (DG): B is μ-separated from A given C if there is no μ-open walk from A to B given C.



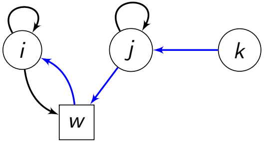


$k \longrightarrow j \longleftarrow w \longleftarrow i$

closed walk given $\{w\}$

$k \longrightarrow j \longleftarrow w \longleftarrow i \longrightarrow i$

open walk given $\{w\}$

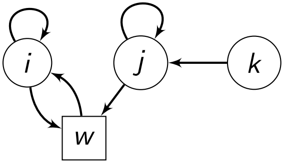


$i \longleftarrow w \longleftarrow j \longleftarrow k$

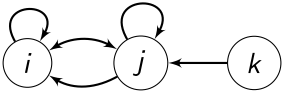
closed walk given $\emptyset$

Graphs and projections

Directed graph G'

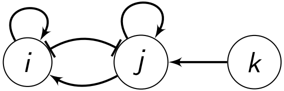


Markov equivalent DMG G



$$\mathcal{I}_{G'}(\{i,j,k\}) = \mathcal{I}_G(\{i,j,k\}) \subset \mathcal{I}_{\tilde{G}}(\{i,j,k\}).$$

Non-equivalent cDG $\tilde{G}$



Independence models for cDMGs

With μ-separation we define an independence model for any directed mixed correlation graphs (cDMGs):

$$G \longmapsto \mathcal{I}_G$$

$$(A, B \mid C) \in \mathcal{I}_G \Leftrightarrow B \text{ is } \mu\text{-sep. from } A \text{ given } C.$$

Two graphs G and G' are Markov equivalent

$$G \sim G'$$

if $\mathcal{I}_G = \mathcal{I}_G$.

The Markov equivalence classes among DGs are singletons:

$$[G] = \{G\}.$$

Main results: Markov and graphoid properties

- Markov equivalent DMGs have a common Markov equivalent supergraph (a greatest element in $[G]$)[1]
- Edge status in the equivalence class is characterized via the directed mixed equivalence graph (DMEG)[1]

- The global Markov property for DGs and conditional local independence is known for event processes[2]
- Abstract graphoid properties for local independence models and their use for proving global Markov properties have been outlined[3]

[1] Mogensen, NRH. Markov equivalence of marginalized local independence graphs, *Annals of Statistics*, 2020.
[2] Didelez. Graphical models for marked point processes based on local independence. JRSS-B 70(1), 2008.
[3] Mogensen, Malinsky, NRH. Causal Learning for Partially Observed Stochastic Dynamical Systems. UAI 2018

Main results: Markov and graphoid properties

- Markov equivalent cDGs do not have a common supergraph. Characterizations of equivalence classes grow superpolynomially in the number of nodes[1]
- Determining Markov equivalence of cDGs is coNP-complete[1]
- The global Markov property holds for cDGs and OU-processes[1]

$$\mathrm{d}X_t = MX_t\mathrm{d}t + D\mathrm{d}W_t$$

Directed edges determined by M, blunt edges determined by DD^T.

[1] Mogensen, NRH. Graphical modeling of stochastic processes driven by correlated errors. *Bernoulli*, 2022

Main results: Learning

- The maximal DMG representing a Markov equivalence class can be constructed from the independence model[1]
- There is a sound and complete FCI-type algorithm that learns the maximal DMG from an oracle assuming faithfulness[2]
- We have some methods and theory for replacing the oracle by conditional local independence testing[3,4]

[1] Mogensen, NRH. Markov equivalence of marginalized local independence graphs, *Annals of Statistics*, 2020.
[2] Mogensen, Malinsky, NRH. Causal Learning for Partially Observed Stochastic Dynamical Systems. UAI 2018
[3] Christgau, Petersen, NRH. Nonparametric conditional local independence testing, arXiv:2203.13550
[4] Thams and NRH. Local Independence Testing for Point Processes *arXiv:2110.12709*, 2022.

Some open problems

- Practical learning
- Global Markov properties
  - Processes not absolutely continuous wrt a product measure.
- Graphical identification
  - Non-parametric identification of interventional effects
  - Identification of parameters
- Constraints imposed by graphs on ...
  - ... the invariant distributions (what is the invariant distribution for non-equilibrium Markov processes)
  - ... other descriptive quantities, e.g. integrated cumulants

Ongoing work: moment equations

The constraints on moments imposed by a linear structural equation and their corresponding graphs have been much studied, partly using (computational) algebra.

One ongoing project (with Mathias et al.) is to understand constraints imposed by local independence graphs, e.g.,

- the Lyapunov equation and higher order cumulant equations arising from cross-sectional distributions of dynamical processes
- higher order integrated cumulants
- invariant distributions of non-equilibrium Markov processes
- one general pattern seems to be that constraints cannot be expressed as (ordinary) conditional independencies of invariant distributions.