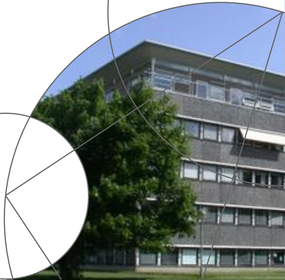
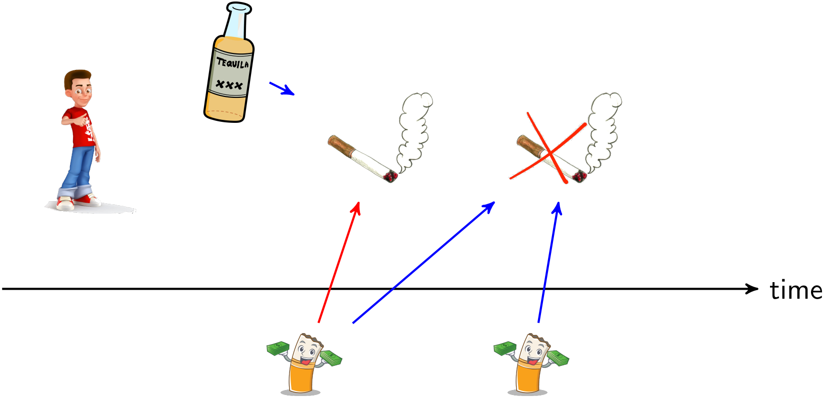
Causal structure learning for partially observed multivariate event processes

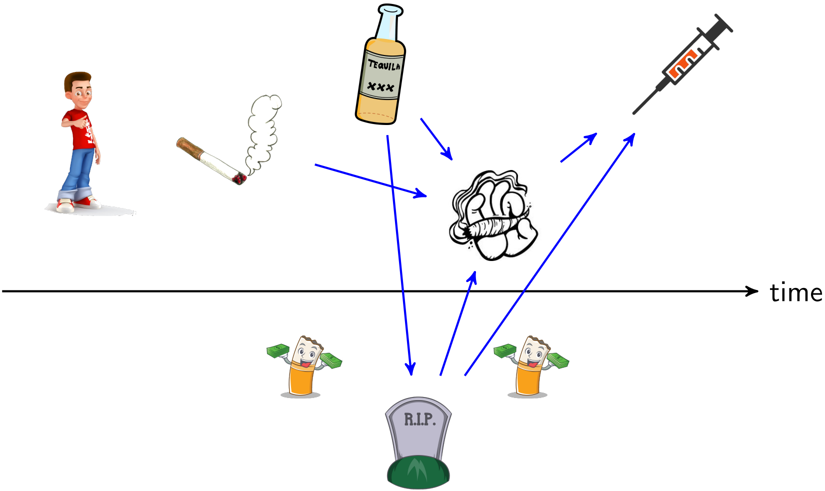
Niels Richard Hansen, Søren Wengel Mogensen, Daniel Malinsky
Department of Mathematical Sciences



The gateway drug theory



The gateway drug theory



Intensities

The k-th event process is modeled in terms of an intensity:

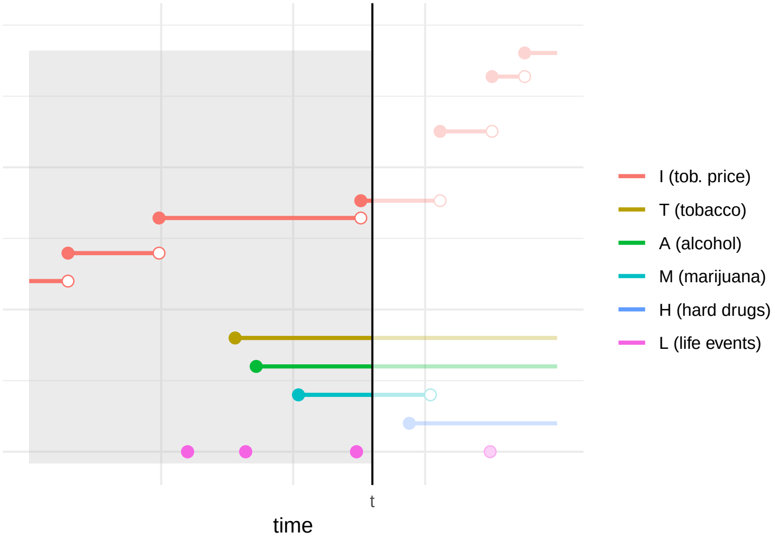
$$P(\text{one } k\text{-event} \in (t, t+\delta] \mid \mathcal{F}_t) \simeq \lambda_t^k \delta, \quad k \in V$$

The $\mathcal{F}_t$ denotes the history of all events up to time t, and λ_t^k depends on $\mathcal{F}_t$.

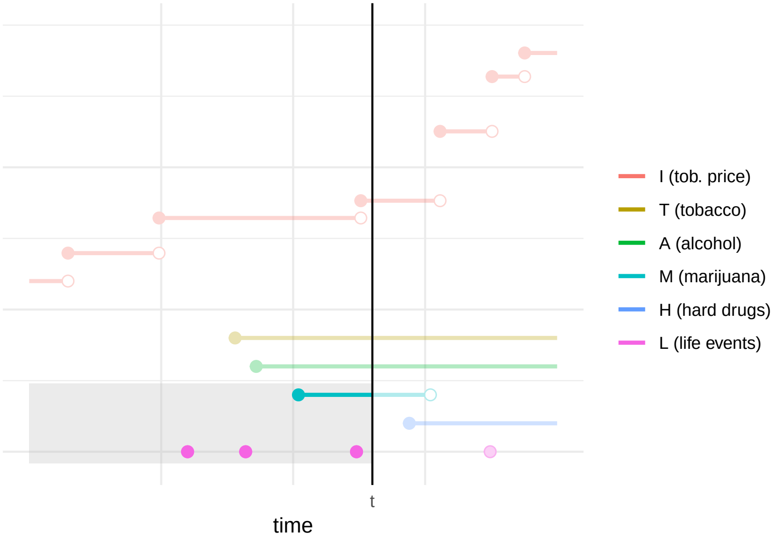
For $C \subseteq V$ define $\mathcal{F}_t^C$ as history of events in C. And

$$\lambda_t^{k,C} = E(\lambda_t^k \mid \mathcal{F}_t^C).$$

The gateway drug event history: $\mathcal{F}_t$



The gateway drug event history: $\mathcal{F}_t^{\{L,M,H\}}$



Local Independence

For $A, B, C \subseteq V$, B is locally independent of A given C,

$$A \not\rightarrow B \mid C \tag{1}$$

if

$$\lambda_t^{k, A \cup C} = \lambda_t^{k, C}$$

for $k \in B$.

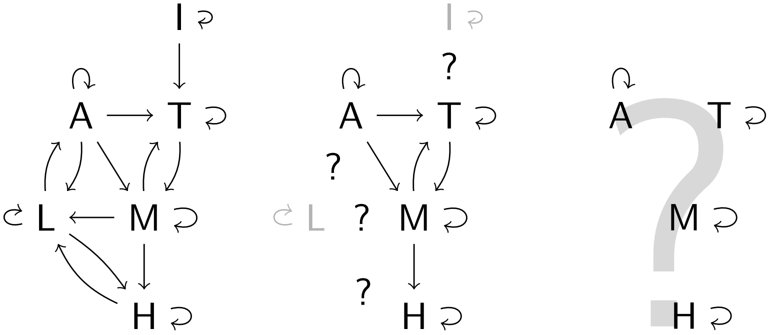
Definition (Local Independence Graph)

A graph $\mathcal{G} = (V, E)$ is a local independence graph if

$$(j, k) \notin E \Longrightarrow j \not\rightarrow k \mid V \backslash \{j\}.$$

Obs: $j \not\rightarrow k \mid V \backslash \{j\}$ if and only if λ_t^k does not depend on j-events.

The gateway drug theory: marginalization



A = Alcohol, T = Tobacco, M = Marijuana, H = Hard drugs
L = Life events, I = Cigarette price.

Abstract independence models

An independence model $\mathcal{I}(V)$ is a ternary relation on subsets of V.

Examples: $A, B, C \subseteq V$

$$\langle A, B \mid C\rangle \in \mathcal{I}_{\mathrm{CI}}(V) \quad \Leftrightarrow \quad X_A \perp\!\!\!\perp X_B \mid X_C$$

$$\langle A, B \mid C\rangle \in \mathcal{I}_{\mathrm{CLI}}(V) \quad \Leftrightarrow \quad \underbrace{A \not\rightarrow B \mid C}_{\text{Cond. Local Ind.}}$$

$\langle A, B \mid C\rangle \in \mathcal{I}_{\mathcal{G}}(V) \quad \Leftrightarrow \quad B$ is graphically separated from A given C in $\mathcal{G} = (V, E)$

Objective: Encode any independence model $\mathcal{I}(V)$ as a graphical independence model $\mathcal{I}_{\mathcal{G}}(V)$ for a suitable graph $\mathcal{G}$ and graphical separation criterion.

Graphical models diagram

Graphs: UG, DG, CG, DAG, AG, MAG, PAG, CMG, DMG, ADMG

Sep. criteria: c-sep, d-sep, m-sep, p-sep, z-sep, δ-sep, μ-sep

$$
\begin{tikzcd}
\hat{\theta}(n) \arrow[r, "P"] & \text{Estimand} & \mathcal{G}' \arrow[l, "\text{ID}"'] \arrow[r, "*\text{-sep}"] \arrow[d, "\text{Projection}"] & \mathcal{I}_{\mathcal{G}'}(V') \arrow[r, equal] \arrow[d, "\text{Restriction}"] & \mathcal{I}_*(V') \arrow[d] & \\
\hat{\theta}(n) \arrow[r, "P"] & \text{Estimand} & \mathcal{G} \arrow[l, "\text{ID}"'] \arrow[r, "*\text{-sep}"] \arrow[dr, "\sim"'] & \mathcal{I}_{\mathcal{G}}(V) \arrow[r, equal] & \mathcal{I}_*(V) \arrow[dl, "\text{Learning}"] & \hat{\mathcal{I}}(n) \arrow[l, "P"'] \arrow[d] \\
\hat{\theta}(n) \arrow[r, "P"] & \text{Estimand} & & {[\mathcal{G}] = \{\mathcal{H} \mid \mathcal{I}_{\mathcal{H}} = \mathcal{I}_{\mathcal{G}}\}} \arrow[ll, "\text{ID}"'] & & {[\hat{\mathcal{G}}(n)]} \arrow[ll, "P"']
\end{tikzcd}
$$

$\mathcal{G}'$ DAG with d-sep:
$$
\begin{tikzcd}
 & & w \arrow[dl] \arrow[dr] & \\
i \arrow[r] & j & & k
\end{tikzcd}
$$

$\mathcal{G}$ AG with m-sep: $i \rightarrow j \leftrightarrow k$

$i \circ\!\!\rightarrow j \leftarrow\!\!\circ k$ $[\mathcal{G}]$ PAG

Directed mixed graphs

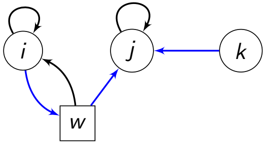
Graphs: DG, DMG

Sep. criteria: δ-sep, μ-sep

$$
\begin{tikzcd}
\color{red}{\mathcal{G}'} \arrow[r, "\color{red}{\mu\text{-sep}}"] \arrow[d, "\color{red}{\text{Projection}}"] & \mathcal{I}_{\mathcal{G}'}(V') \arrow[d] \arrow[drr, phantom, "\text{Restriction}"] & ? & \mathcal{I}_{\mathrm{CLI}}(V') \arrow[d] & \\
\color{red}{\mathcal{G}} \arrow[r, "\color{red}{\mu\text{-sep}}"] \arrow[dr, "\color{red}{\sim}"'] & \mathcal{I}_{\mathcal{G}}(V) & ? & \mathcal{I}_{\mathrm{CLI}}(V) \arrow[dll, "\text{Learning}"] & \hat{\mathcal{I}}(n) \arrow[l, "P"'] \arrow[dl] \\
& \color{red}{[\mathcal{G}] = \{\mathcal{H} \mid \mathcal{I}_{\mathcal{H}} = \mathcal{I}_{\mathcal{G}}\}} & & [\hat{\mathcal{G}}(n)] \arrow[ll, "P"'] &
\end{tikzcd}
$$

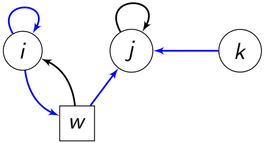
Directed mixed graphs

A directed mixed graph (DMG) $\mathcal{G} = (V, E)$ has directed $\rightarrow$ and bidirected $\leftrightarrow$ edges. Graphical independence is defined via μ-closed walks.



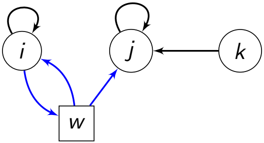
$k \longrightarrow j \longleftarrow w \longleftarrow i$

closed walk given $\{j\}$



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

open walk given $\{j\}$



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

open walk given $\emptyset$

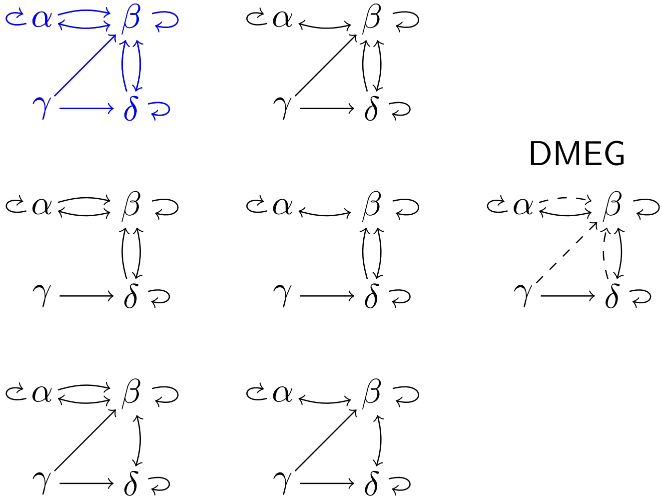
Major results for DMGs

For DMGs[1]:

- A latent projection maps a DMG with vertices V to a DMG with vertices $O \subseteq V$. The μ-separation properties are preserved among observed variables.
- All Markov equivalent DMGs on O have a common Markov equivalent supergraph.
- The maximal DMG representing a Markov equivalence class can be constructed from the independence model.
- Edge status in the equivalence class is characterized via the directed mixed equivalence graph (DMEG).

[1] S. W. Mogensen and NRH. Markov equivalence of marginalized local independence graphs, *Annals of Statistics*, to appear.

Example



DMGs and Markov properties

$$
\begin{tikzcd}
\mathcal{G}' \arrow[r, "\mu\text{-sep}"] \arrow[d] \arrow[dr, phantom, "\text{Projection}"] & \mathcal{I}_{\mathcal{G}'}(V') \arrow[r, phantom, "\subseteq"] \arrow[d] \arrow[dr, phantom, "\text{Restriction}"] & \mathcal{I}_{\mathrm{CLI}}(V') \arrow[d] & \\
\mathcal{G} \arrow[r, "\mu\text{-sep}"] \arrow[dr, "\sim"'] & \mathcal{I}_{\mathcal{G}}(V) \arrow[r, phantom, "\subseteq"] & \mathcal{I}_{\mathrm{CLI}}(V) \arrow[dl, "\text{Learning}"] & \hat{\mathcal{I}}(n) \arrow[l, "P"'] \arrow[dl] \\
& {[\mathcal{G}] = \{\mathcal{H} \mid \mathcal{I}_{\mathcal{H}} = \mathcal{I}_{\mathcal{G}}\}} & \hat{\mathcal{G}}(n) \arrow[l, "P"'] &
\end{tikzcd}
$$

$\mathcal{I}_{\mathcal{G}'}(V') \subseteq \mathcal{I}_{\mathrm{CLI}}(V')$ is the global Markov property.

$\mathcal{I}_{\mathcal{G}'}(V') \supseteq \mathcal{I}_{\mathrm{CLI}}(V')$ is faithfulness.

The global Markov property

Recall that B is locally independent of A given C,

$$A \not\to B \mid C \tag{2}$$

if

$$\lambda_t^{k,A\cup C} = E(\lambda_t^k \mid \mathcal{F}_t^{A\cup C}) = E(\lambda_t^k \mid \mathcal{F}_t^C) = \lambda_t^{k,C}$$

for $k \in B$.

Theorem (V. Didelez[1]; S.W. Mogensen, D. Malinsky & NRH[2])

Under regularity conditions. If C μ-separates A from B in a local independence graph then (2) holds.

[1] Graphical models for marked point processes based on local independence. JRSS-B 70(1), 2008.

[2] Causal Learning for Partially Observed Stochastic Dynamical Systems. UAI 2018

Learning DMEGs

$$
\begin{tikzcd}
\mathcal{G}' \arrow[r, "\mu\text{-sep}"] \arrow[d] \arrow[dr, phantom, "\text{Projection}"] & \mathcal{I}_{\mathcal{G}'}(V') \arrow[r, phantom, "="] \arrow[d] \arrow[dr, phantom, "\text{Restriction}"] & \mathcal{I}_{\mathrm{CLI}}(V') \arrow[d] & \\
\mathcal{G} \arrow[r, "\mu\text{-sep}"] \arrow[dr, "\sim"'] & \mathcal{I}_{\mathcal{G}}(V) \arrow[r, phantom, "="] & \mathcal{I}_{\mathrm{CLI}}(V) \arrow[dl, "\text{Learning}"] & \hat{\mathcal{I}}(n) \arrow[l, "P"'] \arrow[dl] \\
& {[\mathcal{G}] = \{\mathcal{H} \mid \mathcal{I}_{\mathcal{H}} = \mathcal{I}_{\mathcal{G}}\}} & \hat{\mathcal{G}}(n) \arrow[l, "P"'] &
\end{tikzcd}
$$

Learning DMEGs

- Learn $\mathcal{I}_{\mathrm{CLI}}$ and construct the maximal DMG[1] assumming faithfulness.
- Run a sound and complete FCI-type algorithm to construct the maximal DMG[2] assuming faithfulness.
- Test each edge for removal[1] to contruct the DMEG.

For the algorithm to work in practice with data, we need statistical tests of local independence. Ongoing work ...

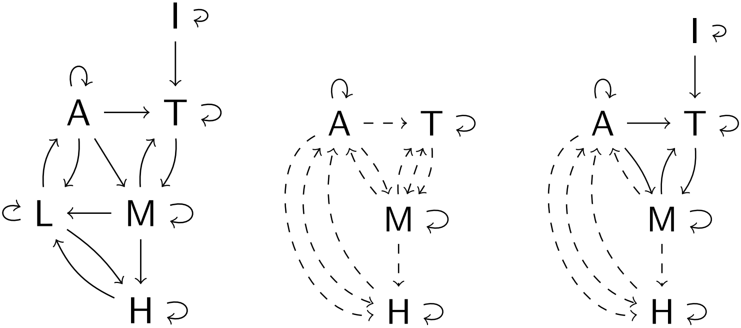
Søren recently showed that testing only

- $\langle \alpha, \beta \mid \beta \rangle$
- $\langle \alpha, \beta \mid \mathrm{pa}(\beta) \backslash \{\alpha\} \rangle$

works surprisingly well in practice for reconstructing the oriented part of the maximal DMG under a condition of ancestral faithfulness.

[1] S. W. Mogensen and NRH. Markov equivalence of marginalized local independence graphs, *Annals of Statistics*, to appear.
[2] Causal Learning for Partially Observed Stochastic Dynamical Systems. UAI 2018

The gateway drug theory DMEGs



A = Alcohol, T = Tobacco, M = Marijuana, H = Hard drugs
L = Life events, I = Cigarette price.

— Thanks for your attention —