Dynamic Structural Causal Models

Philip Boeken, Joris M. Mooij

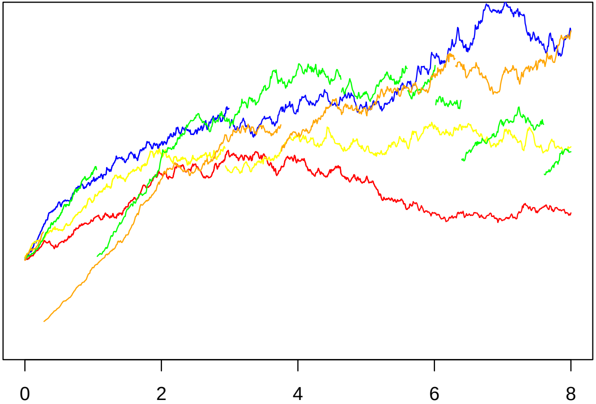
Korteweg-de Vries Institute for Mathematics
Universiteit van Amsterdam

Mercury Machine Learning Lab
Booking.com

Causal Inference for Time Series Data Workshop @ UAI 2024

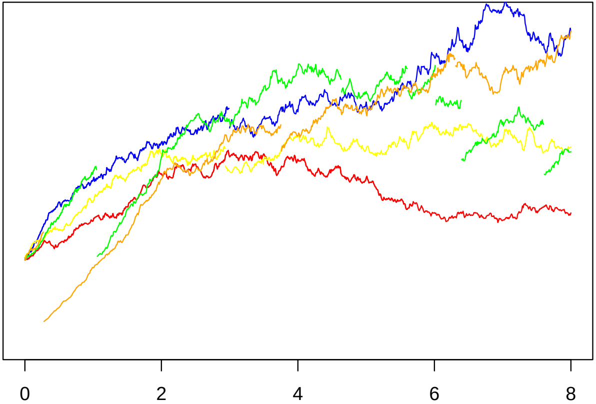
July 19, 2024

Continuous-time Dynamical Systems



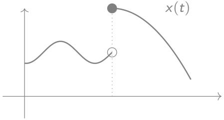
$$X(t) = X^0 + \int_0^t f_1(X(s), Y(s))\mathrm{d}s + \int_0^t f_2(X(s), Y(s))\mathrm{d}W(s) + \int_0^t f_3(X(s-), Y(s-))\mathrm{d}N(s)$$

Continuous-time Dynamical Systems



$$\left.\begin{matrix} x^5 \\ x^4 \\ x^3 \\ x^2 \\ x^1 \end{matrix}\right\} \in D_{\mathcal{T},\mathbb{R}}$$

functions
$x : \mathcal{T} \to \mathbb{R}$
right continuous
with left limits



$$X(t) = X^0 + \int_0^t f_1(X(s), Y(s))\mathrm{d}s + \int_0^t f_2(X(s), Y(s))\mathrm{d}W(s) + \int_0^t f_3(X(s-), Y(s-))\mathrm{d}N(s)$$

Dynamic Structural Causal Models

Structural Causal Model:

- Endogenous V, exogenous W
- Outcome spaces $\mathcal{X}_V$
- Structural equations $X_V = f_v(X_V, X_W)$
- Exogenous distribution $\mathbb{P}(X_W)$

Dynamic Structural Causal Models

Structural Causal Model:

- Endogenous V, exogenous W
- Outcome spaces $\mathcal{X}_V$
- Structural equations $X_V = f_v(X_V, X_W)$
- Exogenous distribution $\mathbb{P}(X_W)$

Main idea: replace static variables by entire sample paths on $\mathcal{T} = [0, T]$.

(Rubenstein et al., 2018)

Dynamic Structural Causal Models

Structural Causal Model:

- Endogenous V, exogenous W
- Outcome spaces $\mathcal{X}_V$
- Structural equations $X_V = f_v(X_V, X_W)$
- Exogenous distribution $\mathbb{P}(X_W)$

Main idea: replace static variables by entire sample paths on $\mathcal{T} = [0, T]$.

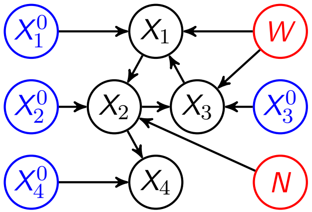
(Rubenstein et al., 2018)

Static SCM	
State space	$\mathcal{X}_v \subseteq \mathbb{R}$
Variable	$X_v \in \mathbb{R}$
Structural equation	$f_v : \mathbb{R}^{m+n} \to \mathbb{R}$

Dynamic SCM	
Function space	$\mathcal{X}_v = D_{\mathcal{T},\mathbb{R}}$
Trajectory	$X_v \in D_{\mathcal{T},\mathbb{R}}$ or $X_v : \mathcal{T} \to \mathbb{R}$
Structural equation	$f_v : D^{m+n}_{\mathcal{T},\mathbb{R}} \to D_{\mathcal{T},\mathbb{R}}$

From a Structural System of SDEs to a DSCM

$$\mathcal{D}: \begin{cases} X_1(t) = X_1^0 + \int_0^t g_1(s-, X_1, X_3)\mathrm{d}W(s) \\ X_2(t) = X_2^0 + \int_0^t g_2(s-, X_1, X_2)\mathrm{d}N(s) \\ X_3(t) = X_3^0 + \int_0^t g_3(s-, X_2, X_3)\mathrm{d}W(s) \\ X_4(t) = X_4^0 + \int_0^t g_4(s-, X_4)\mathrm{d}X_2(s) \end{cases}$$



From a Structural System of SDEs to a DSCM

$$\mathcal{D}: \begin{cases} X_1(t) = X_1^0 + \int_0^t g_1(s-, X_1, X_3)\mathrm{d}W(s) \\ X_2(t) = X_2^0 + \int_0^t g_2(s-, X_1, X_2)\mathrm{d}N(s) \\ X_3(t) = X_3^0 + \int_0^t g_3(s-, X_2, X_3)\mathrm{d}W(s) \\ X_4(t) = X_4^0 + \int_0^t g_4(s-, X_4)\mathrm{d}X_2(s) \end{cases}$$

From a Structural System of SDEs to a DSCM

$$\mathcal{D}: \begin{cases} X_1(t) = X_1^0 + \int_0^t g_1(s-, X_1, X_3)\mathrm{d}W(s) \\ X_2(t) = X_2^0 + \int_0^t g_2(s-, X_1, X_2)\mathrm{d}N(s) \\ X_3(t) = X_3^0 + \int_0^t g_3(s-, X_2, X_3)\mathrm{d}W(s) \\ X_4(t) = X_4^0 + \int_0^t g_4(s-, X_4)\mathrm{d}X_2(s) \end{cases} \xrightarrow{\text{Using Przybyłowicz et al. (2023)}} \mathcal{M}_{\mathcal{D}}: \begin{cases} X_1 = f_1(X_1^0, X_3, W) \\ X_2 = f_2(X_2^0, X_1, N) \\ X_3 = f_3(X_3^0, X_2, W) \\ X_4 = f_4(X_4^0, X_2) \end{cases}$$

From a Structural System of SDEs to a DSCM

Using Przybyłowicz et al. (2023)

$$\mathcal{D}:\begin{cases} X_1(t) = X_1^0 + \int_0^t g_1(s-, X_1, X_3)\mathrm{d}W(s) \\ X_2(t) = X_2^0 + \int_0^t g_2(s-, X_1, X_2)\mathrm{d}N(s) \\ X_3(t) = X_3^0 + \int_0^t g_3(s-, X_2, X_3)\mathrm{d}W(s) \\ X_4(t) = X_4^0 + \int_0^t g_4(s-, X_4)\mathrm{d}X_2(s) \end{cases} \longrightarrow \mathcal{M}_{\mathcal{D}}:\begin{cases} X_1 = f_1(X_1^0, X_3, W) \\ X_2 = f_2(X_2^0, X_1, N) \\ X_3 = f_3(X_3^0, X_2, W) \\ X_4 = f_4(X_4^0, X_2) \end{cases}$$

Description of dynamical evolution of states

Functional dependencies of trajectories

From a Structural System of SDEs to a DSCM

Using Przybyłowicz et al. (2023)

$$\xrightarrow{\hspace{2cm}}$$

$$\mathcal{D}: \begin{cases} X_1(t) = X_1^0 + \int_0^t g_1(s-, X_1, X_3)\mathrm{d}W(s) \\ X_2(t) = X_2^0 + \int_0^t g_2(s-, X_1, X_2)\mathrm{d}N(s) \\ X_3(t) = X_3^0 + \int_0^t g_3(s-, X_2, X_3)\mathrm{d}W(s) \\ X_4(t) = X_4^0 + \int_0^t g_4(s-, X_4)\mathrm{d}X_2(s) \end{cases}$$

$$\mathcal{M}_{\mathcal{D}}: \begin{cases} X_1 = f_1(X_1^0, X_3, W) \\ X_2 = f_2(X_2^0, X_1, N) \\ X_3 = f_3(X_3^0, X_2, W) \\ X_4 = f_4(X_4^0, X_2) \end{cases}$$

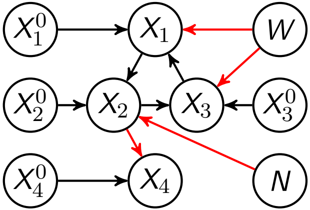
From a Structural System of SDEs to a DSCM

$$
\begin{tikzcd}
\mathcal{D}: \begin{cases} X_1(t) = X_1^0 + \int_0^t g_1(s-, X_1, X_3)\,\mathrm{d}W(s) \\ X_2(t) = X_2^0 + \int_0^t g_2(s-, X_1, X_2)\,\mathrm{d}N(s) \\ X_3(t) = X_3^0 + \int_0^t g_3(s-, X_2, X_3)\,\mathrm{d}W(s) \\ X_4(t) = X_4^0 + \int_0^t g_4(s-, X_4)\,\mathrm{d}X_2(s) \end{cases} \arrow[r, "\text{Using Przybyłowicz et al. (2023)}"] \arrow[d] & \mathcal{M}_{\mathcal{D}}: \begin{cases} X_1 = f_1(X_1^0, X_3, W) \\ X_2 = f_2(X_2^0, X_1, N) \\ X_3 = f_3(X_3^0, X_2, W) \\ X_4 = f_4(X_4^0, X_2) \end{cases} \arrow[d] \\
\mathcal{D}_{\mathrm{do}(X_3 = x_3)} \arrow[r] & \begin{array}{c} (\mathcal{M}_{\mathcal{D}})_{\mathrm{do}(X_3 = x_3)} \\ = \\ \mathcal{M}_{(\mathcal{D}_{\mathrm{do}(X_3 = x_3)})} \end{array}
\end{tikzcd}
$$

DSCM Markov Property

A DSCM is a genuine (cyclic) SCM as in Bongers et al. (2021).

Graph $G(\mathcal{M})$:



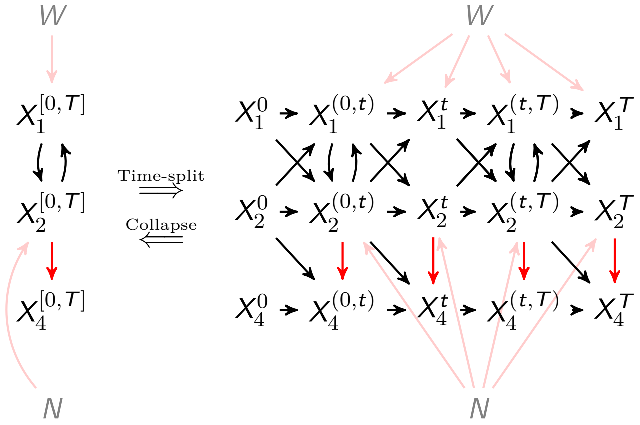
Markov property (Forré and Mooij, 2017):

$$X_1 \overset{\sigma}{\underset{G}{\perp}} X_4 \mid X_2 \implies X_1 \underset{\mathbb{P}(X_V)}{\perp\!\!\!\perp} X_4 \mid X_2$$

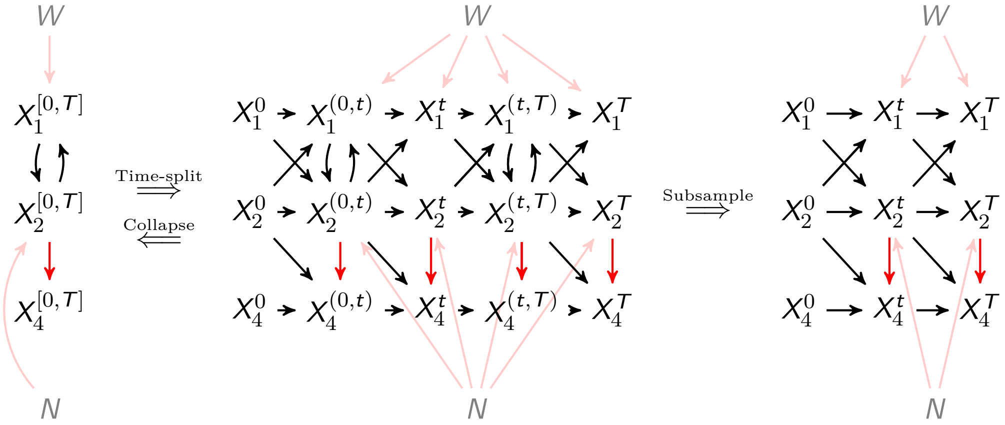
Time-splitting, Subsampling

$$
\begin{tikzcd}
W \arrow[d] \\
X_1^{[0,T]} \arrow[d, shift right] \\
X_2^{[0,T]} \arrow[u, shift right] \arrow[d] \\
X_4^{[0,T]} \\
N \arrow[uu, bend left]
\end{tikzcd}
$$

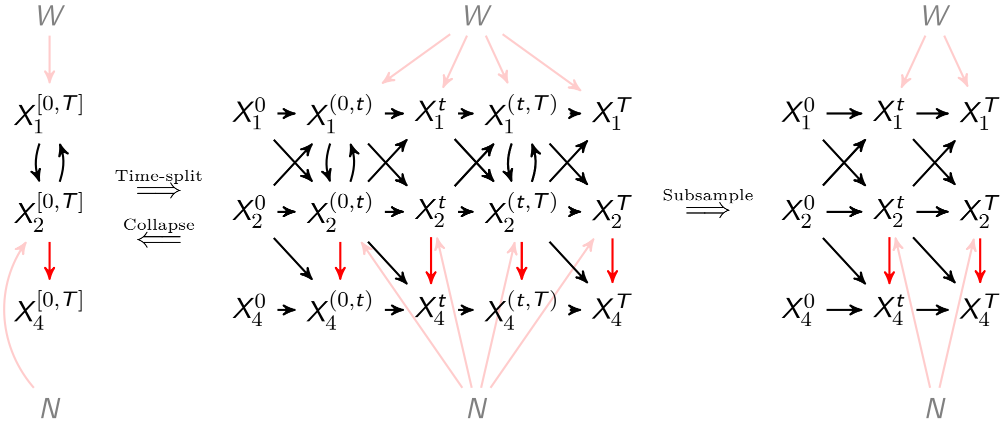
Time-splitting, Subsampling



Time-splitting, Subsampling



Time-splitting, Subsampling



Time-splitting is invertible, but subsampling is not!

Results for DSCMs

We can apply existing ‘static’ SCM methods to dynamical systems:

Cyclic causal effect identification:

- Do-calculus (Forré and Mooij, 2020)
- Generalised adjustment formulae (Forré and Mooij, 2020)
- ID algorithm (Forré and Mooij, 2023)

Cyclic constraint-based causal discovery:

- LCD (Cooper, 1997; Forré and Mooij, 2023)
- Y-structures (Mani, 2006; Forré and Mooij, 2023)
- FCI (Spirtes et al., 1995; Forré and Mooij, 2023)

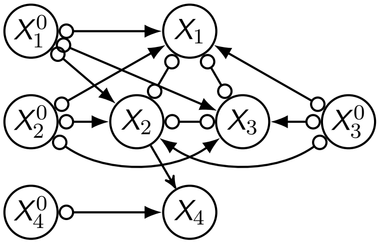
Causal Effect Identification: Backdoor Adjustment

$$\begin{tikzcd}
& W \arrow[dl] \arrow[dr] & \\
X_1^{[0,T]} \arrow[rr, bend left] & & X_2^{[0,T]} \arrow[ll, bend left]
\end{tikzcd}$$

$$\mathbb{P}(X_2|\,\mathrm{do}(X_1 = x)) = \int_{D_{\mathcal{T},\mathbb{R}}} \mathbb{P}(X_2|X_1 = x, W = w)\mathrm{d}\mathbb{P}(w)$$

Constraint-based Causal Discovery: FCI

Output of FCI with independence oracle:



In practice, require conditional independence test for e.g. $X_1 \perp\!\!\!\perp X_4 \mid X_2$ (Lundborg et al., 2022; Manten et al., 2024).

Local Conditional Independence

Let $X_A \not\to X_B \mid X_C$ denote local conditional independence of X_B from X_A given X_C.

Theorem

For a collapsed DSCM $\mathcal{M}$ with no instantaneous effects and independent integrators, $G(\mathcal{M})$ is a local independence graph and

$$X_A \underset{G(\mathcal{M})}{\overset{\sigma}{\perp}} X_B | X_C \implies X_A \not\to X_B \mid X_C.$$

References I

Bongers, S., Forré, P., Peters, J., and Mooij, J. M. (2021). Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics*, 49(5).

Cooper, G. F. (1997). A Simple Constraint-Based Algorithm for Efficiently Mining Observational Databases for Causal Relationships. *Data Mining and Knowledge Discovery*.

Forré, P. and Mooij, J. M. (2017). Markov Properties for Graphical Models with Cycles and Latent Variables.

Forré, P. and Mooij, J. M. (2020). Causal Calculus in the Presence of Cycles, Latent Confounders and Selection Bias. In *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, pages 71–80. PMLR.

Forré, P. and Mooij, J. M. (2023). A Mathematical Introduction to Causality.

Lundborg, A. R., Shah, R. D., and Peters, J. (2022). Conditional Independence Testing in Hilbert Spaces with Applications to Functional Data Analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(5):1821–1850.

Mani, S. (2006). *A Bayesian Local Causal Discovery Framework*. University of Pittsburgh ETD, University of Pittsburgh.

Manten, G., Casolo, C., Ferrucci, E., Mogensen, S., Salvi, C., and Kilbertus, N. (2024). Signature Kernel Conditional Independence Tests in Causal Discovery for Stochastic Processes.

Rubenstein, P. K., Bongers, S., Schölkopf, B., and Mooij, J. M. (2018). From deterministic ODEs to dynamic structural causal models. In *Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-18)*, pages 114–123.

Spirtes, P. L., Meek, C., and Richardson, T. S. (1995). Causal Inference in the Presence of Latent Variables and Selection Bias.