A Markov Property for Sample Paths of Stochastic Processes

Philip Boeken
(Joint work with Joris Mooij)

Korteweg-de Vries Institute for Mathematics
Universiteit van Amsterdam

Mercury Machine Learning Lab
Booking.com

Amsterdam Causality Meeting

February 29, 2024

Introduction

Mathematical Modelling of Stochastic Dynamical Systems

- Dynamic Bayesian Networks
- Ordinary Differential Equations
- Random Differential Equations
- Stochastic Differential Equations
  - Itô SDEs (continuous, Markov)
  - Jump-diffusion SDEs (Discontinuous, Markov)
  - Semimartingale SDEs (Discontinuous, Non-Markov)

Causal Modelling of Stochastic Dynamical Systems

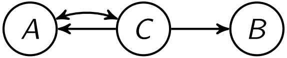
- Structural Causal Models ‘on standard Borel spaces’
- Refining ‘Dynamical Structural Causal Models’
- Modelling interventions on SDEs

The importance of a graphical Markov property I I

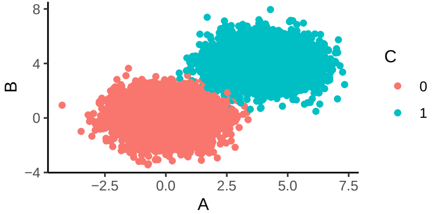
For an acyclic SCM M with endogenous variables V, graph (ADMG) $G(M)$, observational distribution $\mathbb{P}_M(X_V)$, we have the global Markov property:

$$A \underset{G}{\overset{d}{\perp}} B \mid C \implies X_A \underset{\mathbb{P}(X_V)}{\perp\!\!\!\perp} X_B \mid X_C$$

for all subsets $A, B, C \subseteq V$.



$\implies$



$$A \underset{G}{\overset{d}{\perp}} B \mid C \qquad\qquad X_A \underset{\mathbb{P}}{\perp\!\!\!\perp} X_B \mid X_C$$

The importance of a graphical Markov property II

A Markov property for a causal model M and a notion of intervention $M_{\mathrm{do}(X)}$ imply:

1. Transportability of statistical relations: $Y \perp^d_G S \mid X \implies \mathbb{E}[Y|X] = \mathbb{E}[Y|X, S=1]$.
2. Identification of causal effects:
   - Do-calculus, e.g. $Y \perp^d_{G_{\mathrm{do}(I_X)}} X|I_X, Z \implies \mathbb{P}(Y|\,\mathrm{do}(X), Z) = \mathbb{P}(Y|X, Z)$
   - Adjustment formulae, e.g. $Y \perp^d_{G_{\mathrm{do}(I_X)}} X|I_X, Z$ and $Z \perp^d_{G_{\mathrm{do}(I_X)}} I_X$ imply

$$\mathbb{P}(Y|\,\mathrm{do}(X=x)) = \sum_z \mathbb{P}(Y|X=x, Z=z)\mathbb{P}(Z=z).$$

   - ID algorithm
3. Constraint-based causal discovery, under the faithfulness assumption

$$A \underset{G}{\overset{d}{\not\perp}} B \mid C \implies X_A \underset{\mathbb{P}(X_V)}{\not\perp\!\!\!\perp} X_B \mid X_C$$

   and with availability of a statistical CI test.

For an overview, see lecture notes for the MasterMath course 'Causality' (Forré and Mooij, 2023).

Dynamical Systems I

Often, SCMs are used to model for a single item, a single measurement per variable. When data is drawn i.i.d., statistical methods can be used for inference.

Consider the following data of multiple *mosquitofish*:

Fish	Weight (g)	Age (d)	VI	Water temp.
1	1.2	30	0	16.2
2	0.8	35	1	17.0
3	0.67	29	0	17.9
4	1.12	25	0	15.4
⋮				

Dynamical systems II

It can be that for each item, the variables are measured repeatedly over time. (Often referred to as multidimensional time series, or panel data.)

Fish	Weight (g)	Age (d)	VI	Water temp.
1	0.41	1	0	16.2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	1.23	40	1	15.0
2	0.37	1	0	15.7
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
2	1.45	40	1	18.3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Dynamical systems III

In certain domains (e.g. physics, chemistry, biology, neurology), specific differential equation models are known which appropriately describe such dynamical systems.

- Ordinary Differential Equations
- Random Differential Equations
- Stochastic Differential Equations

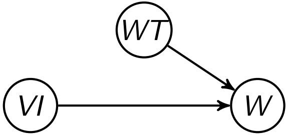
However, clasically these models are merely descriptive, and are not equipped with causal semantics.

Some existing work

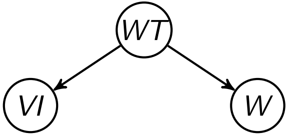
- Granger causality for DBNs (Eichler, 2012)
- Local Independence Graphs (SDEs, no causal calculus) (Didelez, 2008; Mogensen and Hansen, 2020)
- Causal constraints models: Blom et al. (2021)
  - Uses Simon's causal ordering algorithm to solve sets of equations
- Equilibration of perfectly adaptive systems (RDEs) (Blom and Mooij, 2023)
- Dynamic Structural Causal Models (Rubenstein et al., 2018)
  - Models trajectories as a whole
  - We extend their definition from ODEs to SDEs, and prove a Markov property.
- Structural Dynamical Causal Models (Bongers et al., 2022)
  - Models trajectories of RDE solutions as a whole, proves a Markov property.
  - Includes derivative processes as endogenous variables (impossible for SDEs!)

Our goal

In this work, we derive a Markov property for entire sample paths of discrete time and continuous time stochastic processes, which allows for causal reasoning, inference, and discovery, on the level of entire sample paths.



??



Introduction

Mathematical Modelling of Stochastic Dynamical Systems

- Dynamic Bayesian Networks
- Ordinary Differential Equations
- Random Differential Equations
- Stochastic Differential Equations
  - Itô SDEs (continuous, Markov)
  - Jump-diffusion SDEs (Discontinuous, Markov)
  - Semimartingale SDEs (Discontinuous, Non-Markov)

Causal Modelling of Stochastic Dynamical Systems

- Structural Causal Models 'on standard Borel spaces'
- Refining 'Dynamical Structural Causal Models'
- Modelling interventions on SDEs

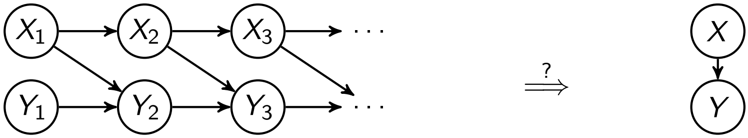
Dynamic Bayesian Networks

Definition (DBN, Dean and Kanazawa (1989); Murphy (2002))

Let V be a finite set of endogenous variables, and let $\mathcal{T} = [1, T] \cap \mathbb{N}_0$ for some $T \in \mathbb{N}$. A set of conditional distributions

$$\mathcal{D}^b : \left\{ \mathbb{P}(X_v(t) | \{X_V(s) : s < t\}) \quad \text{for all } v \in V \right. \tag{1}$$

is referred to as a *Dynamic Bayesian Network*.



The class of DBNs includes ARMA, Hidden Markov Models, MDP, POMDPs, etc.

Ordinary Differential Equations

Let V be a finite set of endogenous variables, W a finite set of exogenous variables. Let $\mathcal{T} = [0, T]$ for some $T \in \mathbb{N}$.

Definition (ODE)

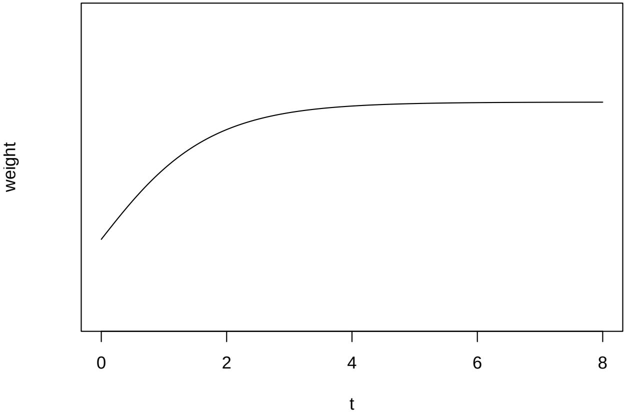
For $\xi_V \in \mathbb{R}^{|V|}$, $e_W \in C(\mathcal{T}, \mathbb{R})^{|W|}$ and $g_V : \mathcal{T} \times \mathbb{R}^{|V|} \times \mathbb{R}^{|W|} \to \mathbb{R}^{|V|}$, we refer to

$$\mathcal{D}^o : \begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} x_v(t) = g_v(t, x_V(t), e_W(t)) \\ x_v(0) = \xi_v \end{cases} \qquad \text{for all } v \in V$$

as an *ordinary differential equation*.

This can easily be extended to higher (finite) order ODE's.

ODE: Example



Gompertz ODE: $\frac{\mathrm{d}}{\mathrm{d}t}x(t) = x(t)(a - b\ln(x(t))), \quad x(0) = c$

Random Differential Equations

The growth of different fish can be governed by different dynamics. Growth rate of mosquitofish depends on genetic factors, and whether it got a certain viral infection as larvae.
So, for each individual the dynamics are deterministic, over the population the dynamics are stochastic.

Definition (RDE)

For probability space $(\Omega, \mathcal{F}, \mathbb{P})$, random variables $\xi_V : \Omega \to \mathbb{R}^{|V|}$, $E_W : \Omega \to C(\mathcal{T}, \mathbb{R})^{|W|}$ and $g_V : \mathcal{T} \times \mathbb{R}^{|V|} \times \mathbb{R}^{|W|} \to \mathbb{R}^{|V|}$, we refer to

$$\mathcal{D}^r : \begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} X_v(t) = g_v(t, X_V(t), E_W(t)) \\ X_v(0) = \xi_v \end{cases} \qquad \text{for all } v \in V$$

as a *random differential equation*. A stochastic process $X_V : \Omega \to C^1(\mathcal{T}, \mathbb{R})^{|V|}$ is called a *solution* of $\mathcal{D}^r$ if for $\mathbb{P}$-almost all ω it satisfies the ODE

$$\mathcal{D}^o(\omega) : \begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} X_v(\omega, t) = g_v(t, X_V(\omega, t), E_W(\omega, t)) \\ X_v(\omega, 0) = \xi_v(\omega) \end{cases} \qquad \text{for all } v \in V.$$

Integral representation of ODE/RDE

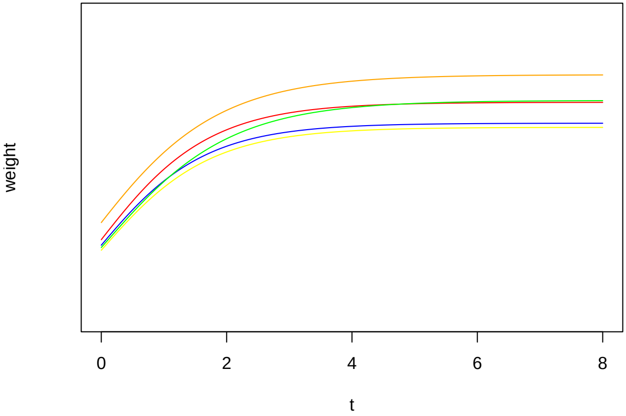
An ordinary/random differential equation

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} X_v(t) = g_v(t, X_V(t), E_W(t)) \\ X_v(0) = \xi_v \end{cases}$$

can equivalently be represented by the integral equation

$$X_v(t) = \xi_v + \int_0^t g_v(s, X_V(s), E_W(s)) \mathrm{d}s.$$

RDE: Example



$$X(t) = C + \int_0^t X(s)(A - B\ln(X(s)))\mathrm{d}s$$

Random flutuations during lifetime.

When random factors influence the dynamics throughout the life of an individual, this is often modelled with an SDE. This allows for non-differentiability, unbounded variation, and even jumps.

Definition (Brownian motion, Thiele, Bachelier, Wiener)

An $\mathbb{R}$-valued stochastic process B with time index $[0, \infty)$ is called a *Brownian motion* if

1. $B(0) = 0$ a.s.
2. $B \in C(\mathcal{T}, \mathbb{R})$ a.s.
3. The increments of B are independent
4. $B(t) - B(s) \sim \mathcal{N}(0, t - s)$ for all $0 \leq s < t < \infty$.

The Brownian motion is nowhere differentiable and has unbounded variation, i.e. $\lim_n \sum_{i=0}^{n-1} |B(\omega, t_{i+1}^n) - B(\omega, t_i^n)| = \infty$ for almost all ω, where $t_{i+1}^n := t\frac{i}{n}$.

Stochastic Integration II

Recall the definition of the *Riemann integral*

$$\int_0^t X(s)\mathrm{d}s := \lim_{n\to\infty} \sum_{i=0}^{n-1} X(t_i^n)(t_{i+1}^n - t_i^n),$$

and recall for $A : \mathcal{T} \to \mathbb{R}$ of bounded variation the *Riemann-Stieltjes integral*

$$\int_0^t X(s)\mathrm{d}A(s) := \lim_{n\to\infty} \sum_{i=0}^{n-1} X(t_i^n)(A(t_{i+1}^n) - A(t_i^n)).$$

Definition (Stochastic integral, Itô)

Given Brownian motion B, stochastic process X, then (under measurability and integrability conditions) the Itô integral of X w.r.t. B is a random variable defined as the limit

$$\int_0^t X(s)\mathrm{d}B(s) := \operatorname*{plim}_{n\to\infty} \sum_{i=0}^{n-1} X(t_i^n)(B(t_{i+1}^n) - B(t_i^n)).$$

Stochastic Differential Equations (Diffusion, Itô)

Definition (Itô/diffusion SDE)

Given a Brownian motion B and functions $g_1, g_2 : \mathbb{R} \to \mathbb{R}$, we refer to

$$X(t) = X(0) + \int_0^t g_1(X(s))\mathrm{d}s + \int_0^t g_2(X(s))\mathrm{d}B(s) \tag{2}$$

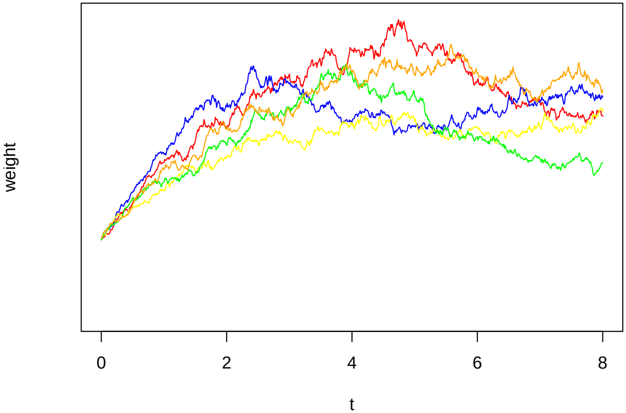
as a *diffusion SDE*.

Theorem (Markov solutions, Protter (2004), Theorem V.6.32)

Any solution X of (2) is strong Markov and continuous (a.k.a. a diffusion).

Note that SDEs strictly generalise Random Differential Equations.

Diffusion SDE: Example



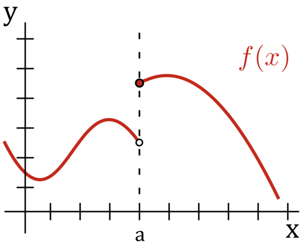
$$X(t) = C + \int_0^t X(s)(A - B\ln(X(s)))\mathrm{d}s + \int_0^t D\mathrm{d}B(s)$$

Random jumps during lifetime

A fish's tail might get bitten off, causing a downward jump in its weight. To this end, we go beyond continuous functions $C(\mathcal{T}, \mathbb{R})$, and consider the space of càdlàg functions.

Definition (Càdlàg functions)

A function $f : \mathcal{T} \to \mathbb{R}$ is called *càdlàg* (continu à droit, limité à gauche) if it is right-continuous and has left-limits. Let $D(\mathcal{T}, \mathbb{R}) := \{f : \mathcal{T} \to \mathbb{R} \,|f \text{ is càdlàg}\}$.



Stochastic Differential Equations (Jump-diffusion)

Definition (Jump-diffusion SDE)

Let B be a Brownian motion, N a jump process, then we refer to

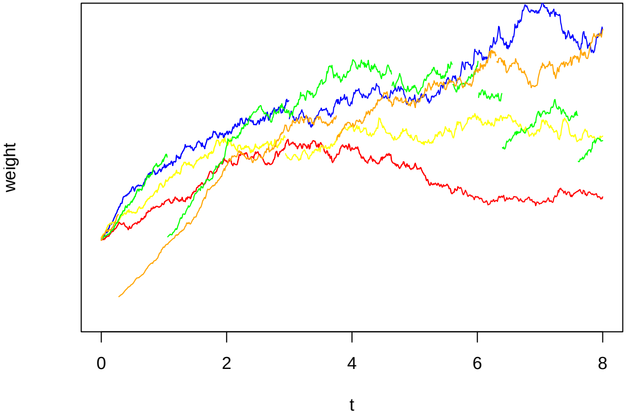
$$X(t) = \xi + \int_0^t g_1(X(s))\mathrm{d}s + \int_0^t g_2(X(s))\mathrm{d}B(s) + \int_0^t g_3(X(s-))\mathrm{d}N(s) \qquad (3)$$

as a *jump-diffusion SDE*, where $X(t-) := \lim_{s\uparrow t} X(s)$.

Theorem (Protter (2004) Theorem V.6.32)

Any solution X of (3) is strong Markov and càdlàg.

Jump-diffusion SDE: Example



$$X(t) = C + \int_0^t X(s)(A - B\ln(X(s)))\mathrm{d}s + \int_0^t D\mathrm{d}B(s) - \int_0^t \mathrm{d}N(s)$$

Discontinuous, non-Markov stochastic processes (semimartingales)

The most general class of stochastic processes Z for which a stochastic integral $\int X \mathrm{d}Z$ can be defined, is the class of *semimartingales* (Protter, 2004, Theorem III.1.1).

Important properties of semimartingales are:

- semimartingales are not necessarily Markov;
- all semimartingales are càdlàg.

Stochastic Differential Equations (Semimartingale)

Definition (Semimartingale SDE)

For finite index sets V, W, a finite random variable ξ_w, an $\mathbb{R}$-valued semimartingale Z_w for every $w \in W$, and function $g_{v,w} : \mathcal{T} \times \mathbb{R}^{|V|} \to \mathbb{R}$ for all $v \in V, w \in W$, we refer to

$$\mathcal{D}^s : X_v(t) = \xi_v + \sum_{w \in W} \int_0^t g_{v,w}(s, X_V(s-))\mathrm{d}Z_w(s) \quad \text{for all } v \in V \tag{4}$$

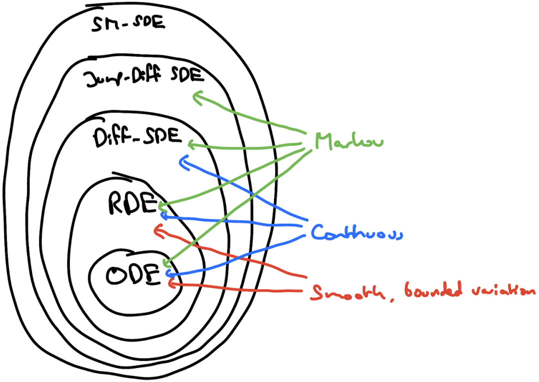
as a *semimartingale SDE*.

Theorem (Protter (2004) Theorem V.3.7)

If $g_{v,w}$ is Lipschitz for all $v \in V, w \in W$, there exists a unique semimartingale X_V that is a solution of $\mathcal{D}^s$.

Comparison of modelling frameworks

- Discrete time: DBN
- Continuous time:



Introduction

Mathematical Modelling of Stochastic Dynamical Systems

- Dynamic Bayesian Networks
- Ordinary Differential Equations
- Random Differential Equations
- Stochastic Differential Equations
  - Itô SDEs (continuous, Markov)
  - Jump-diffusion SDEs (Discontinuous, Markov)
  - Semimartingale SDEs (Discontinuous, Non-Markov)

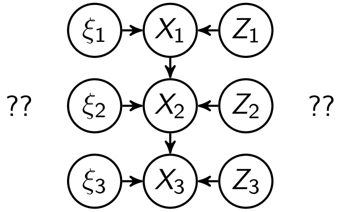
Causal Modelling of Stochastic Dynamical Systems

- Structural Causal Models 'on standard Borel spaces'
- Refining 'Dynamical Structural Causal Models'
- Modelling interventions on SDEs

Given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, independent random variables $\xi_1, \xi_2, \xi_3 : \Omega \to \mathbb{R}$ and semimartingales Z_1, Z_2, Z_3 taking values in $\mathbb{R}$, consider the SDE

$$\mathcal{D} : \begin{cases} X_1(t) = \xi_1 + \int_0^t g_1(s, X_1(s-))\mathrm{d}Z_1(s) \\ X_2(t) = \xi_2 + \int_0^t g_2(s, X_1(s-), X_2(s-))\mathrm{d}Z_2(s) \\ X_3(t) = \xi_3 + \int_0^t g_3(s, X_2(s-), X_3(s-))\mathrm{d}Z_3(s) \end{cases}$$

What can be said about the causal relations among the variables $\xi_1, \xi_2, \xi_3, Z_1, Z_2, Z_3$ and solutions X_1, X_2, X_3? (Provided they exist).



Structural Causal Models 'on Standard Borel Spaces' I

Definition (Standard Borel Space)

A measurable space $(\mathcal{X}, \Sigma)$ is called a *standard Borel space* if there exists a metric d on $\mathcal{X}$ such that $(\mathcal{X}, d)$ is a complete metric space with a countable dense subset (a.k.a. a *Polish* space), for which Σ is the induced Borel σ-algebra.

Theorem (Kuratowski, Kechris (1995), Theorem 15.6)

Every standard Borel space is measurably isomorphic to one of the following measurable spaces:

- *a countable set $N \subseteq \mathbb{N}$ with the power set σ-algebra 2^N.*
- *the real line $\mathbb{R}$ with the Borel σ-algebra.*

Structural Causal Models 'on Standard Borel Spaces' II

Definition (Structural causal model, Bongers et al. (2021))

Formally, a *structural causal model* is a tuple

$$\mathcal{M} = \langle V, W, \mathcal{X}, \mathcal{E}, f, \mathbb{P}_{\mathcal{E}} \rangle$$

where

1. V, W are disjoint finite index sets of *endogenous variables* and *exogenous variables* respectively,
2. the *endogenous domain* $\mathcal{X} = \prod_{v \in V} \mathcal{X}_v$ and *exogenous domain* $\mathcal{E} = \prod_{w \in W} \mathcal{E}_w$ are products of *standard Borel spaces* $\mathcal{X}_v, \mathcal{E}_w$,
3. the *causal mechanism* $f : \mathcal{X} \times \mathcal{E} \to \mathcal{X}$ is a measurable function,
4. the *exogenous distribution* $\mathbb{P}(E_W) = \bigotimes_{w \in W} \mathbb{P}(E_w)$ is a product of probability measures.

Structural Causal Models ‘on Standard Borel Spaces’ III

Theorem (Markov property, Forré and Mooij (2017); Bongers et al. (2021))

Let $\mathcal{M}$ be an acyclic SCM with graph G, then its observational distribution $\mathbb{P}(X_V)$ satisfies the d-separation Global Markov Property, i.e.

$$A \underset{\mathcal{G}}{\overset{d}{\perp}} B \mid C \implies X_A \underset{\mathbb{P}(X_V)}{\perp\!\!\!\perp} X_B \mid X_C$$

for all subsets $A, B, C \subseteq V$.

Structural Causal Models ‘on Standard Borel Spaces’ IV

As a consequence, we get causal effect identification:

- Do-calculus, (Pearl, 2009; Forré and Mooij, 2020)
- Generalised adjustment formulae (Pearl, 2009; Forré and Mooij, 2020)
- ID algorithm (Pearl, 2009; Forré and Mooij, 2023)

and constrain-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)

(provided a suitable CI test is available for the standard Borel spaces $\mathcal{X}_v$).

From SDEs to SCMs I

Theorem (Skorokhod (1956))

On the space of càdlàg functions $D(\mathcal{T}, \mathbb{R})$ *there exists a topology* J_1 *such that* $(D(\mathcal{T}, \mathbb{R}), \sigma(J_1))$ *is a standard Borel space, where* $\sigma(J_1)$ *denotes the (Borel)* σ*-algebra generated by* J_1*.*

From SDEs to SCMs II

Theorem (Solution functions of SDEs)

Given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, random variables $\xi_1, \xi_2 : \Omega \to \mathbb{R}$, semimartingales Z_1, Z_2 taking values in $\mathbb{R}$, and Lipschitz g_1, g_2 of appropriate dimension, for the SDE

$$\mathcal{D} : \begin{cases} X_1(t) = \xi_1 + \int_0^t g_1(s, X_1(s-))\mathrm{d}Z_1(s) \\ X_2(t) = \xi_2 + \int_0^t g_2(s, X_1(s-), X_2(s-))\mathrm{d}Z_2(s) \end{cases}$$

there exists **measurable solution functions**

$$\begin{aligned} f_1 &: \mathbb{R} \times D(\mathcal{T}, \mathbb{R}) \to D(\mathcal{T}, \mathbb{R}) \\ f_2 &: \mathbb{R} \times D(\mathcal{T}, \mathbb{R})^2 \to D(\mathcal{T}, \mathbb{R}) \end{aligned} \quad \text{such that} \quad \begin{aligned} X_1 &= f_1(\xi_1, Z_1) \\ X_2 &= f_2(\xi_2, X_1, Z_2) \end{aligned}$$

almost surely, where (X_1, X_2) is the solution of $\mathcal{D}$.

Dynamic Structural Causal Models I

We 'overload' the definition of DSCMs by Rubenstein et al. (2018):

Definition (Dynamic Structural Causal Model)

Given a time index $\mathcal{T} = [0, T)$ or $\mathcal{T} = [1, T) \cap \mathbb{N}$ for $T \in \mathbb{R} \cup \{\infty\}$, a *Dynamic Structural Causal Model* is an SCM

$$\mathcal{M} = \langle V_0 \cup V_p, W_0 \cup W_p, \mathcal{X}, \mathcal{E}, f, \mathbb{P}_{\mathcal{E}} \rangle \,.$$

- Endogenous initial conditions/parameters V_0, endogenous processes V_p
- Exogenous initial conditions/parameters W_0, exogenous processes W_p
- Standard borel spaces $\mathcal{X} = \mathbb{R}^{|V_0|} \times D(\mathcal{T}, \mathbb{R})^{|V_p|}$ and $\mathcal{E} = \mathbb{R}^{|W_0|} \times D(\mathcal{T}, \mathbb{R})^{|W_p|}$
- Structural equations f (that can contain solution functions of an SDE)
- Exogenous distribution $\mathbb{P}_{\mathcal{E}} = \mathbb{P}(E_{W_0}) \otimes \mathbb{P}(Z_{W_p})$ (that factorizes over $W_0 \cup W_p$)

Dynamic Structural Causal Models II

For the semimartingale SDE with g_1, g_2, g_3 Lipschitz

$$\mathcal{D}: \begin{cases} X_1(t) = \xi_1 + \int_0^t g_1(s, X_1(s-))\mathrm{d}Z_1(s) \\ X_2(t) = \xi_2 + \int_0^t g_2(s, X_1(s-), X_2(s-))\mathrm{d}Z_2(s) \\ X_3(t) = \xi_3 + \int_0^t g_3(s, X_2(s-), X_3(s-))\mathrm{d}Z_3(s) \end{cases}$$

there exist solution functions f_1, f_2, f_3, from which we can construct DSCM $\mathcal{M}_{\mathcal{D}}$:

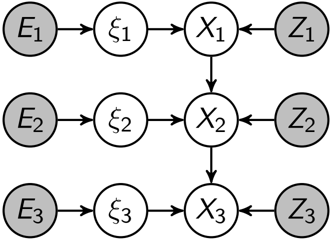
- endogenous $V_0 = \{\xi_1, \xi_2, \xi_3\}$, $V_p = \{X_1, X_2, X_3\}$
- exogenous $W_0 = \{E_1, E_2, E_3\}$, $W_p = \{Z_1, Z_2, Z_3\}$
- exogenous distribution $\mathbb{P}(E_1) \otimes \mathbb{P}(E_2) \otimes \mathbb{P}(E_3) \otimes \mathbb{P}(Z_1) \otimes \mathbb{P}(Z_2) \otimes \mathbb{P}(Z_3)$
- structural equations:

$$\begin{aligned} \xi_1 &= E_1, & X_1 &= f_1(\xi_1, Z_1) \\ \xi_2 &= E_2, & X_2 &= f_2(\xi_2, X_1, Z_2) \\ \xi_3 &= E_3, & X_3 &= f_3(\xi_3, X_2, Z_3) \end{aligned}$$

DSCM Markov property and do-calculus

$$\mathcal{D}: \begin{cases} X_1(t) = \xi_1 + \int_0^t g_1(s, X_1(s-))\mathrm{d}Z_1(s) \\ X_2(t) = \xi_2 + \int_0^t g_2(s, X_1(s-), X_2(s-))\mathrm{d}Z_2(s) \\ X_3(t) = \xi_3 + \int_0^t g_3(s, X_2(s-), X_3(s-))\mathrm{d}Z_3(s) \end{cases}$$

$$\mathcal{M}_{\mathcal{D}}: \begin{cases} \xi_1 = E_1, \\ \xi_2 = E_2, \\ \xi_3 = E_3, \\ X_1 = f_1(\xi_1, Z_1) \\ X_2 = f_2(\xi_2, X_1, Z_2) \\ X_3 = f_3(\xi_3, X_2, Z_3) \end{cases}$$



Corollary

We now have a Markov property $X_1 \perp_G^d X_3 | X_2 \implies X_1 \perp\!\!\!\perp_{\mathbb{P}} X_3 | X_2$, *we can reason about* $\mathbb{P}(X_3 | \operatorname{do}(\xi_1))$, *etc.*

General procedure $\mathcal{D} \to \mathcal{M}_{\mathcal{D}}$

For an arbitrary SDE $\mathcal{D}$:

$$X_{V_p}(t) = \xi_{V_p} + \sum_{w \in W_p} \int_0^t g_{V_p,w}(s, X_{V_p}(s-))\mathrm{d}Z_w(s)$$

1. Find topological ordering of the equations
2. Merge variables in a strongly connected component into a multidimensional variable
3. Solve the equations along the topological ordering
4. Specify the DSCM $\mathcal{M}_{\mathcal{D}}$.

This forces the DSCM to be acyclic.

This is yet to be formalised using Simon's causal ordering algorithm, following Blom et al. (2021).

Interventions on $\mathcal{D}$ and $\mathcal{M}$ I

Similar to Hansen and Sokol (2014) and Rubenstein et al. (2018), we define hard interventions on SDEs:

Definition (Perfect intervention on $\mathcal{D}$)

For intervention target $I \subseteq V_p$, intervention value $x_I \in D(\mathcal{T}, \mathbb{R})^{|I|}$ and all $v \in V_p$:

$$\mathcal{D}_{\mathrm{do}(X_I = x_I)} : \begin{cases} X_v(t) = X_0 + \sum_{w \in W_p} \int_0^t g_{v,w}(s, X(s-))\mathrm{d}Z_w(s) & \text{if } v \notin I \\ X_v(t) = x_v(t) & \text{if } v \in I. \end{cases}$$

Definition (Perfect intervention on $\mathcal{M}$)

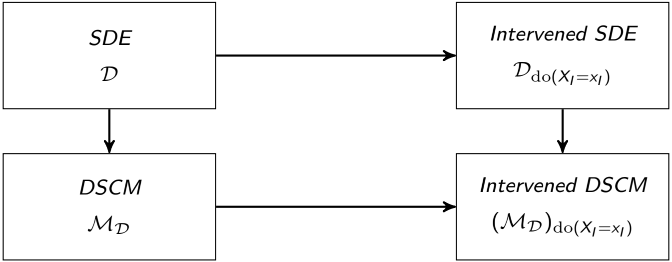
For intervention target $I \subseteq V_p$, intervention value $x_I \in \mathcal{X}_I$ and all $v \in V_p$:

$$\mathcal{M}_{\mathrm{do}(X_I = x_I)} : \begin{cases} X_v = f_v(X_V, E_W) & \text{if } v \notin I \\ X_v = x_v & \text{if } v \in I. \end{cases}$$

Interventions on $\mathcal{D}$ and $\mathcal{M}$ II

Theorem

For $I \subseteq V_p$ a strongly connected component of $\mathcal{D}$, the following diagram commutes:



so we have

$$\mathcal{M}_{\mathcal{D}_{\mathrm{do}(X_I = x_I)}} = (\mathcal{M}_{\mathcal{D}})_{\mathrm{do}(X_I = x_I)}.$$

From DBNs to DSCMs

Theorem (Solution functions of DBNs)

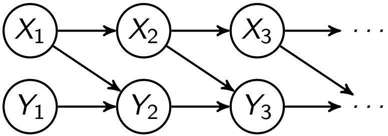
Given the following Dynamic Bayesian Network:

$$\mathcal{D} : \begin{cases} \mathbb{P}(X(t) | \{X(s) : s < t\}) \\ \mathbb{P}(Y(t) | \{Y(s), X(s) : s < t\}) \end{cases}$$

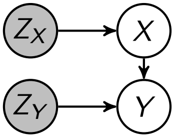
there exists a DSCM

$$\mathcal{M}_{\mathcal{D}} : \begin{cases} X = f_X(Z_X) \\ Y = f_Y(X, Z_Y) \\ \mathbb{P}(Z_X) \otimes \mathbb{P}(Z_Y). \end{cases}$$

$G(\mathcal{D})$:



$\Longrightarrow$ $G(\mathcal{M}_{\mathcal{D}})$:



Discussion

Refinements compared with Rubenstein et al. (2018):

- Allow for stochasticity: instead of ODE $\mathcal{D}^o$, consider SDE $\mathcal{D}^s$.
- Trajectory spaces are standard Borel spaces, structural equations are measurable.
- DSCMs are now proper SCMs, allowing for causal reasoning:
  - Graphical Markov property
  - Do-calculus
  - Causal discovery (CI Testing with functional data: Lundborg et al. (2022)).

Possible improvements:

- Use Simon's causal ordering algorithm for automatically solving $\mathcal{D}$, following Blom et al. (2021).
- Let SDEs be structural equations for allow for interventions $\mathcal{M}_{\mathrm{do}(X_i=x_i)}$ *within* a strongly connected component. Requires a pathwise stochastic integration theory, like rough path theory, or Itô-Föllmer integration.

References I

Blom, T. and Mooij, J. M. (2023). Causality and independence in perfectly adapted dynamical systems. *Journal of Causal Inference*, 11(1).

Blom, T., Van Diepen, M. M., and Mooij, J. M. (2021). Conditional independences and causal relations implied by sets of equations. *The Journal of Machine Learning Research*, 22(1):178:8044–178:8105.

Bongers, S., Blom, T., and Mooij, J. M. (2022). Causal Modeling of Dynamical Systems.

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*.

Dean, T. and Kanazawa, K. (1989). A model for reasoning about persistence and causation. *Computational Intelligence*, 5(2):142–150.

References II

Didelez, V. (2008). Graphical Models for Marked Point Processes Based on Local Independence. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(1):245–264.

Eichler, M. (2012). Graphical modelling of multivariate time series. *Probability Theory and Related Fields*, 153(1):233–268.

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.

Hansen, N. and Sokol, A. (2014). Causal interpretation of stochastic differential equations. *Electronic Journal of Probability*, 19(none):1–24.

References III

Kechris, A. S. (1995). *Classical Descriptive Set Theory*, volume 156 of *Graduate Texts in Mathematics*. Springer New York, New York, NY.

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.

Mogensen, S. W. and Hansen, N. R. (2020). Markov equivalence of marginalized local independence graphs. *The Annals of Statistics*, 48(1).

Murphy, K. (2002). *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis.

Pearl, J. (2009). *Causality*. Cambridge University Press.

References IV

Protter, P. E. (2004). *Stochastic Integration and Differential Equations*. Number 21 in Applications of Mathematics. Springer, Berlin ; New York, 2nd ed edition.

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.

Skorokhod, A. V. (1956). Limit Theorems for Stochastic Processes. *Theory of Probability & Its Applications*, 1(3):261–290.

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