Causal models for dynamical systems

by Jonas Peters, Stefan Bauer, Niklas Pfister

Presenter: Jiamei Zhang

UC Irvine

March 2023

Outline

Summary of the paper:

- **Style**: Not a regular methodology/engineering paper with much technical details; Combination of a book chapter/letter/review
- **Goal**: Model the full time evolution of dynamic system and data measured at different time points
- **Contribution**: Propose a modeling framework for structural causal models (SCM) in dynamic systems: two ways to introduce randomness
- **Prerequisite** Understanding observational distribution and intervention distribution; Modular structure of SCM.

Two versions of SCMs: 1) with measurement noise

A deterministic SCM over d variables $x_1, ..., x_d$ is a collection of d assignments

$$x^k := f^k\left(x^{\mathbf{PA}_k}\right), \quad k = 1, \ldots, d \tag{1}$$

If now obtain noisy observations of system, we have

$$X^k := x^k + \varepsilon^k \tag{2}$$

- Assume the system (1) is uniquely solvable
- Each SCM induce a state of system characterized by a point in $\mathbb{R}^d$
- In the noisy observations, ϵ^1, ϵ^d are jointly independent r.v. So (2) induces a joint distribution instead of a single point over observed r.v. $X^1, ..., X^d$

A stochastic SCM over d variables $x_1, ..., x_d$ is a collection of d assignments

$$X^k := f^k \left(X^{\mathbf{PA}_k}, \varepsilon^k \right), \quad k = 1, \ldots, d \tag{3}$$

- Difference is randomness enters inside the structural assignment
- Similarly,assume the system (3) is uniquely solvable
- Assume noise components are jointly independent

The intervention induces a new state of the system:

1. Replace the assignment with $do(x^j := \tilde{f}^j\left(X^{\tilde{\mathbf{PA}}_j}\right))$ in model-1 and $do(x^j := \tilde{f}^j\left(X^{\tilde{\mathbf{PA}}_j}, \tilde{\varepsilon}^k\right)$ in model-2
2. Under interventions on fixed $j \neq k$, $X^k \mid X^{\mathbf{PA}_k} = x$ keeps invariant

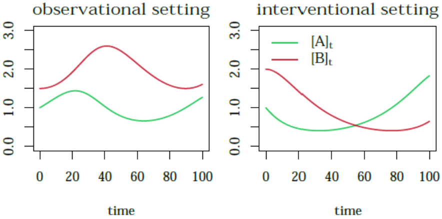


Figure: Left with observational setting; Right with intervening on initial concentration of B and rate of A k_1

This work focus on continuous time systems that governed by ordinary differential equations (ODE), such as bioprocessing, genetics, economic, neuroscience, etc. Here present two typical examples

Figure: **Lotka-Volterra model**: illustrate dynamics between prey and predators

$$A \xrightarrow{k_1} 2A$$
$$A + B \xrightarrow{k_2} 2B$$
$$B \xrightarrow{k_3} \emptyset,$$
$$\tfrac{d}{dt}[A] = k_1[A] - k_2[A][B]$$
$$\tfrac{d}{dt}[B] = k_2[A][B] - k_3[B].$$

Figure: **Michaelis-Menten kinetics**: enzyme-involved reactions resulting non-linear ODE

$$E + S \xrightarrow{k_1} ES$$
$$ES \xrightarrow{k_2} E + S$$
$$ES \xrightarrow{k_3} E + P, \quad \Rightarrow \quad \tfrac{d}{dt}[P] = c_1 \frac{[S]}{c_2 + [S]},$$

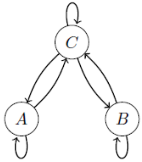
$$\frac{d}{dt}[S] = -k_1[S][E] + k_2[ES]$$
$$\frac{d}{dt}[E] = -k_1[S][E] + k_2[ES] + k_3[ES]$$
$$\frac{d}{dt}[P] = k_3[ES]$$
$$\frac{d}{dt}[ES] = k_1[S][E] - k_2[ES] - k_3[ES]$$

Connecting ODE and causal graph

ODE representations

$$A \xrightarrow{k_1} 2A$$
$$B \xrightarrow{k_2} 2B$$
$$A + C \xrightarrow{k_3} 2C$$
$$B + C \xrightarrow{k_4} 2C$$
$$C \xrightarrow{k_5} \varnothing$$

$$\frac{\mathrm{d}}{\mathrm{dt}}[A] = k_1[A] - k_3[A][C]$$
$$\frac{\mathrm{d}}{\mathrm{dt}}[B] = k_2[B] - k_4[B][C]$$
$$\frac{\mathrm{d}}{\mathrm{dt}}[C] = k_3[A][C] + k_4[B][C] - k_5[C]$$



A *deterministic causal kinetic model* over processes $x := (x_t)_t := (x_t^1, ..., x_t^d)_t$ is a collection of d ODEs and initial value assignments are

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} x_t^1 &:= f^1\left(x_t^{\mathbf{PA}_1}\right), \qquad & x_0^1 &:= \xi_0^1, \\
\frac{\mathrm{d}}{\mathrm{d}t} x_t^2 &:= f^2\left(x_t^{\mathbf{PA}_2}\right), \qquad & x_0^2 &:= \xi_0^2, \\
&\vdots \\
\frac{\mathrm{d}}{\mathrm{d}t} x_t^d &:= f^d\left(x_t^{\mathbf{PA}_d}\right), \qquad & x_0^d &:= \xi_0^d.
\end{aligned}
\tag{4}
$$

- For amy $k \in \{1, ..., d\}$, $\frac{\mathrm{d}}{\mathrm{d}t} x_t^k$ denotes the time derivative of component x^k at time t
- Similarly, it requires initial values is uniquely solvable; If we consider the initial values as r.v, it induces a distribution over $x = (x_t)_t$.
- With measurement noise $\mathbf{X}_t = \mathbf{x}_t + \varepsilon_t$, assuming noise component ϵ_t is iid, it induces a distribution over $\mathbf{X} = (\mathbf{X}_t)_t$

A *stochastic causal kinetic model* over processes $\mathbf{X} := (\mathbf{X}_t)_t = (X_t^1, ..., X_t^d)$ is a collection of d SDEs and initial value assignments are

$$\mathrm{d}X_t^k := f^k\left(X_t^{\mathbf{PA}_k}\right)\mathrm{d}t + h^k\left(X_t^{\mathbf{PA}_k}\right)\mathrm{d}W_t^k, \quad X_0^k := \xi_0^k, \qquad (5)$$

- dW_t^k are independent white noise processes. For example, $W_t = \int_0^t dW_s$ is a Brownian motion.
- The functions f^k are called drift coefficient and h^k are called diffusion coefficients.
- In most basic setting, h^k is assume constant so that

$$X_t^k := \int_0^t f^k\left(X_s^{\mathbf{PA}_k}\right)\mathrm{d}s + W_t^k \qquad (6)$$

Interventions can change the dynamics of the process x^k, the initial values or both at the same time.

- A deterministic causal kinetic model over a process $(x_t)_t$, an intervention on x^k are denoted by

$$do\left(x_0^k := \xi\right) \quad \text{and} \quad do\left(\frac{\mathrm{d}}{\mathrm{d}t}x_t^k := g\left(x_t^{\mathbf{PA}}\right)\right), \qquad (7)$$

- For a stochastic causal kinetic model, we analogously define the interventions

$$do\left(X_0^k := \xi\right) \quad \text{and} \quad do\left(\mathrm{d}X_t^k := g\left(X_t^{\mathrm{PA}}\right) + j\left(X_t^{\mathrm{PA}}\right)\mathrm{d}W_t^k\right) \qquad (8)$$

Intervention representations: Continued

- For example, Lotka-Voltera model, changing the rate of reaction k_1, and initial [B]
- If we want change x^k to a constant value, we could also achieved by: 1) introducing a forcing term on x^k, or 2) changing the strength of dependence of $\frac{\mathrm{d}}{\mathrm{d}t}x_t^k$ on x^k, or 3) changing the parent set
- For a system described well by ODEs, the most natural way to formulate intervention is to formulate it as differential equations, too.
    - For example, for a differentiable ζ that $x_t := \zeta(t)$, $do(x_t^k := \zeta(t))$ is realized by $do(\frac{\mathrm{d}}{\mathrm{d}t}x_t^k := \frac{\mathrm{d}}{\mathrm{d}t}\zeta(t)$ and $do(x_0^k := \zeta(0))$

Lots of questions remain unsolved for dynamical systems:

- In the deterministic settings, solving a standard algebraic equation is easier than solving an algebraic equation involving differentials.
- The fitting process is more challenging when noise variables are not additive.
- It is not clear whether conditional independence (d-seperation) is the right notion
- How to involve hidden variables, do-calculus, adjustments on observed distributions are not as clear as iid cases.

Conceptual procedures: exploit the invariances induced by the underlying causal kinetic model for causal discovery

1. Assume a target process $y := x_1$, for which the parents are unknown and of particular interest;
2. If the measurements stem from an underlying causal kinetic model under different interventional settings, in none of which the variable y has been intervened on;
3. Suppose each repetition is part of an environment or experimental condition, and the assignment under each condition is known;
4. Then n repetitions are generated by the model $\frac{\mathrm{d}}{\mathrm{d}t} y_t = f\left(x_t^{\mathbf{PA}_y}\right)$ for a fixed f^y (or possible with additional measurement noise
5. The whole process should ouput a ranking of models (or variables) by trading off predictability and invariance.

Summary: causal kinetic model VS iid case

- We discussed SCM governed by differential equations: with measurement noise or driving noise
- The two model classes are a start point of more questions. Many aspects are not fully understood or proved as we have for iid data
- Difference to SCM: not only Markov equivalence in a single a single graph; mild interventions may carry lots of information on causal structure
- Share in common: causal kinetic models exhibit the same modularity as structural causal models.

Question: Briefly explain the difference of representing 'intervention' between iid causal model and kinetic causal model
Email: jiameiz@uci.edu