Intervention in Ornstein-Uhlenbeck SDEs

Alexander Sokol

Joint work with Niels Richard Hansen

18th EYSM, 29 August, 2013

Agenda

1. Review of structural equation models
2. SDEs as models of causality
3. The Ornstein-Uhlenbeck case
4. Identifiability of postintervention distributions
5. Perspectives

Review of structural equation models

Consider a collection of variables $X_1, \ldots, X_p$ (a gene expression network, concentrations of chemicals, et cetera).

X_2

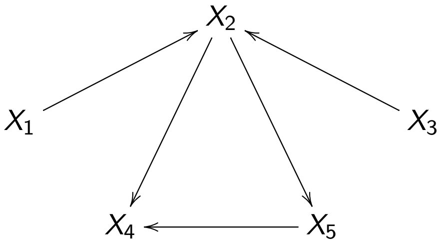
X_1 X_3

X_4 X_5

In most situations, we will only be able to observe the distribution of the variables $X_1, \ldots, X_p$.

Review of structural equation models

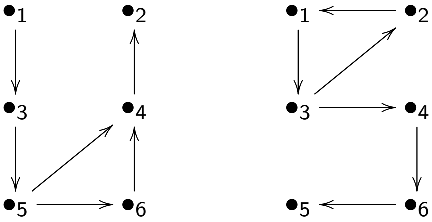
Consider a collection of variables $X_1, \ldots, X_p$ (a gene expression network, concentrations of chemicals, et cetera).



However, we will often be interested in the **causal** relationships between the variables. Such relationships may be expressed by means of a graph. Causal relationships allow us to reason about the effects of interventions.

Review of structural equation models

Definition. A directed acyclic graph (DAG) G on a set of vertices V is a directed graph with no cycles.



Above: The left graph is a DAG, while the right graph is not a DAG.

In a DAG, there is a natural notion of "parents", "children", "ancestors" and "descendants".

Review of structural equation models

A model for causality can be obtained through the notion of a structural equation model (SEM).

Definition. A set of variables $X_1, \ldots, X_p$ is said to satisfy a set of structural equations with DAG G on $\{1, \ldots, p\}$ if there exists noise variables $\varepsilon_1, \ldots, \varepsilon_p$ such that

$$X_i = f_i(X_{\mathrm{pa}(i)}, \varepsilon_i)$$

This allows for a notion of "intervention": The result of making the intervention $X_i := x_i$ is defined to be the variables resulting from removing the i'th equation, substituting x_i for X_i in the remaining equations, and removing all edges in G pointing towards X_i.

Review of structural equation models

A recurring problem in causal inference is that for a set of variables $X_1, \ldots, X_p$ satisfying a set of structural equations, the DAG G is not uniquely determined from the distribution of $X_1, \ldots, X_p$.

This makes it difficult to estimate the effects of interventions.

It is reasonable to believe that causality is easier to handle when observations are time-dependent. In the following, we will:

1. Introduce a notion of intervention for SDEs.
2. Calculate intervention effects in Ornstein-Uhlenbeck SDEs.
3. Argue that postintervention distributions often are uniquely determined from observational distributions.
4. Relate our results to classical causal inference.

SDEs as models of causality

The stochastic integral $\int_0^t H_s \,\mathrm{d}X_s$ can be defined for integrand processes $(H_t)_{t\geq 0}$ and integrators $(X_t)_{t\geq 0}$ satisfying suitable regularity properties, and behaves "like an integral" in the sense that for example

$$\lim_n \sum_{k=1}^{2^n} H_{t(k-1)/2^n}(X_{tk/2^n} - X_{t(k-1)/2^n}) = \int_0^t H_s \,\mathrm{d}X_s,$$

where the limit is in probability.

The class of "well-behaved integrators" is the space of **semimartingales**.

SDEs as models of causality

Using the stochastic integral, we may consider equations in X such as

$$X_t = 1 + \int_0^t X_{s-} \, \mathrm{d}Z_s,$$

where Z is some semimartingale. This equation can be written more suggestively in "differential form" as

$$\mathrm{d}X_t = X_{t-} \, \mathrm{d}Z_t.$$

We refer to such equations as **stochastic differential equations** (SDEs). The solution to this particular equation, for example, is

$$X_t = \exp\left(Z_t - \frac{1}{2}[Z^c]_t\right) \prod_{0<s\leq t} (1 + \Delta Z_s) \exp\left(-\Delta Z_s\right).$$

SDEs as models of causality

In general, we are most interested in multivariate SDEs. Consider:

- A p-dimensional initial condition X_0.
- A d-dimensional semimartingale Z.
- A continuous function $a : \mathbb{R}^p \to \mathbb{M}(p, d)$,

where $\mathbb{M}(p, d)$ is the space of real $p \times d$ matrices. We then consider the equation

$$X_t^i = X_0^i + \sum_{j=1}^{d} \int_0^t a_{ij}(X_{s-}) \, \mathrm{d}Z_s^j \qquad \text{for } i \leq p \tag{$\dagger$}$$

which has a unique solution whenever a is Lipschitz. ($\dagger$) can be written more compactly using differential and matrix notation as

$$\mathrm{d}X_t = a(X_{t-}) \, \mathrm{d}Z_t.$$

SDEs as models of causality

As the argument t in X_t frequently denotes time, it is often natural to interpret the solution (X_t) to $(\dagger)$ in terms of causality: We may use the formula

$$\mathrm{d}X_t = a(X_{t-})\,\mathrm{d}Z_t$$

to make the approximation

$$X_{t+\Delta} = X_t + a(X_t)(Z_{t+\Delta} - Z_t)$$

and consider this as describing how interventions in X_t will influence $X_{t+\Delta}$.

In the following, we formalize a notion of intervention in SDEs and show that this notion essentially is equivalent to a version of the above idea.

SDEs as models of causality

Definition. Consider some $m \leq p$ and $\zeta \in \mathbb{R}$. The postintervention SDE arising from making the intervention $X^m := \zeta$ in the SDE (†) is the p-dimensional SDE

$$Y_t^i = Y_0^i + \sum_{j=1}^{d} \int_0^t b_{ij}(Y_{s-})\,\mathrm{d}Z_s^j \qquad \text{for } i \leq p \tag{††}$$

where $b : \mathbb{R}^p \to \mathbb{M}(p,d)$, $b_{ij}(y) = a_{ij}(y)$ for $i \neq m$ and $b_{mj}(y) = 0$, $Y_0^i = X_0^i$ for $i \neq m$ and $Y_0^m = \zeta$. In differential form, the intervention takes us from the p-dimensional SDE

$$\mathrm{d}X_t = a(X_{t-})\,\mathrm{d}Z_t$$

to the p-dimensional SDE

$$\mathrm{d}Y_t = b(Y_{t-})\,\mathrm{d}Z_t.$$

SDEs as models of causality

Our first objective is to give an interpretation of what this definition means. We will do this using the notion of Euler schemes.

Definition. Fix $T > 0$ and consider $\Delta > 0$ such that $N = T/\Delta$ is natural. The Euler scheme $(X^{\Delta}_{k\Delta})_{k \leq N}$ for $(\dagger)$ over $[0, T]$ with step size Δ is the set of p-dimensional variables defined by putting $X^{\Delta}_0 = X_0$ and recursively defining

$$X^{\Delta}_{k\Delta} = X^{\Delta}_{(k-1)\Delta} + a(X^{\Delta}_{(k-1)\Delta})(Z_{k\Delta} - Z_{(k-1)\Delta}).$$

Theorem (Protter, Émery). Assume that a is Lipschitz. Connecting the variables $(X^{\Delta}_{k\Delta})$ into a function $(X^{\Delta}_t)_{t \leq T}$ in a suitable way and letting Δ tend to zero, it holds that $\sup_{t \leq T} \mid X_t - X^{\Delta}_t \mid$ tends to zero in probability.

SDEs as models of causality

A structural equation model is implicit in the recursive definition of the Euler scheme: $(X^\Delta_{k\Delta})_{k \le N}$ constitutes a SEM of $p(N+1)$ one-dimensional variables with:

- Functional relationships:

$$(X^\Delta_{k\Delta})^i = (X^\Delta_{(k-1)\Delta})^i + \sum_{j=1}^{d} a_{ij}(X^\Delta_{(k-1)\Delta})(Z^j_{k\Delta} - Z^j_{(k-1)\Delta}).$$

- Noise variables $(Z_{k\Delta} - Z_{(k-1)\Delta})_{k \le N}$.
- DAG G defined by letting $((i_1, j_1), (i_2, j_2))$ be an edge if and only if $i_2 = i_1 + 1$ and $j_1 = j_2$ or $j_1 \neq j_2$ and $a_{j_1 \cdot}$ is not independent of the j_2'th coordinate.

Note that given the mapping a, this DAG is not learned: Rather, for fixed a, the above constitutes a rule for which DAG we think of as the "true" DAG for the SEM, and in practice it is a which is to be estimated.

SDEs as models of causality

Below, an illustration of the DAG:

$$
\begin{tikzcd}
(X^{\Delta})^1_0 \arrow[r] \arrow[dr] & (X^{\Delta})^1_{\Delta} \arrow[r] \arrow[dr] & (X^{\Delta})^1_{2\Delta} \arrow[r] \arrow[dr] & (X^{\Delta})^1_{3\Delta} \arrow[r, dashed] \arrow[dr, dashed] & {} \\
(X^{\Delta})^2_0 \arrow[r] & (X^{\Delta})^2_{\Delta} \arrow[r] & (X^{\Delta})^2_{2\Delta} \arrow[r] & (X^{\Delta})^2_{3\Delta} \arrow[r, dashed] & {} \\
(X^{\Delta})^3_0 \arrow[r] \arrow[ur] & (X^{\Delta})^3_{\Delta} \arrow[r] \arrow[ur] & (X^{\Delta})^3_{2\Delta} \arrow[r] \arrow[ur] & (X^{\Delta})^3_{3\Delta} \arrow[r, dashed] \arrow[ur, dashed] & {} \\
 & Z_{\Delta} - Z_0 \arrow[u, bend left] \arrow[uu, bend left] \arrow[uuu, bend left] & Z_{2\Delta} - Z_{\Delta} \arrow[u, bend left] \arrow[uu, bend left] \arrow[uuu, bend left] & Z_{3\Delta} - Z_{2\Delta} \arrow[u, bend left] \arrow[uu, bend left] \arrow[uuu, bend left] &
\end{tikzcd}
$$

SDEs as models of causality

The following result is then immediate:

Proposition. Fix $T > 0$ and $\Delta > 0$ such that $N = T/\Delta$ is natural. The Euler scheme SEM for the SDE (††) is equal to the postintervention SEM obtained by making the intervention $(X^{\Delta})^m_{k\Delta} := \zeta$ for all $k \leq N$ in the Euler scheme SEM for the SDE (†).

Essentially, this means that the following diagram commutes:

$$\begin{tikzcd}
\boxed{\text{Observational Euler SEM}} \arrow[r] \arrow[d] & \boxed{\text{Observational SDE}} \arrow[d] \\
\boxed{\text{Postintervention Euler SEM}} \arrow[r] & \boxed{\text{Postintervention SDE}}
\end{tikzcd}$$

The Ornstein-Uhlenbeck case

Consider now the p-dimensional Ornstein-Uhlenbeck SDE

$$\mathrm{d}X_t = B(X_t - A)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t,$$

Let Y be the process obtained from making the intervention $X^m := \zeta$. With Y^{-m} denoting Y minus the m'th coordinate, we then obtain that Y^{-m} satisfies the Ornstein-Uhlenbeck SDE

$$\mathrm{d}Y_t^{-m} = \tilde{B}(Y_s^{-m} - \tilde{A})\,\mathrm{d}t + \tilde{\sigma}\,\mathrm{d}W_t,$$

where

$$\begin{aligned} \tilde{B} &= B \text{ minus the } m\text{'th row and column} \\ \tilde{\sigma} &= \sigma \text{ minus the } m\text{'th row} \\ \tilde{A} &= \alpha - \tilde{B}^{-1}\beta, \end{aligned}$$

and $\alpha, \beta \in \mathbb{R}^{p-1}$ with $\alpha_i = A_i$, $\beta_i = B_{im}(\zeta - A_m)$.

The Ornstein-Uhlenbeck case

For the Ornstein-Uhlenbeck process, the zeroes of the mean reversion speed matrix B measures causal dependency.

In special cases, the stationary mean and variance of postintervention OU processes are computable and comparable with the observational OU process. Consider the case $\sigma = I_3$ and

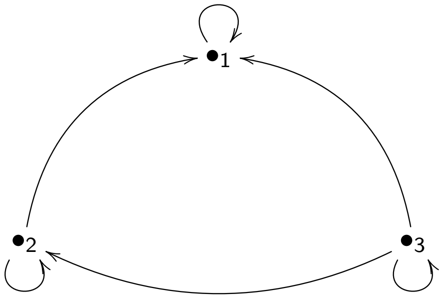
$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ 0 & b_{22} & b_{23} \\ 0 & 0 & b_{33} \end{bmatrix},$$

where the diagonal entries are assumed negative to ensure existence of the stationary distribution.

We consider the stationary distribution corresponding to the interventions $X^2 := \zeta$ and $X^3 := \zeta$.

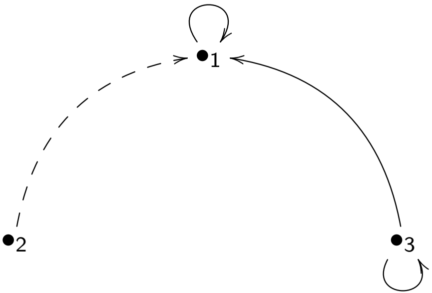
The Ornstein-Uhlenbeck case

Some intuition. The observational process has causality structure:



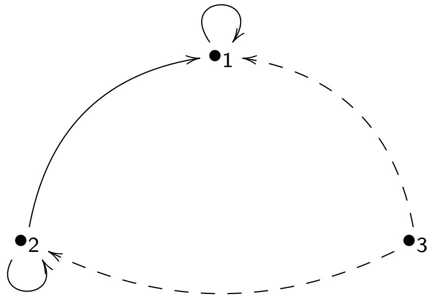
The Ornstein-Uhlenbeck case

Some intuition. The result of $X^2 := \zeta$ has causality structure:



The Ornstein-Uhlenbeck case

Some intuition. The result of $X^3 := \zeta$ has causality structure:



The Ornstein-Uhlenbeck case

The stationary means are:

$$\text{No intervention:} \quad \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix}$$

$$X^2 := \zeta \text{ intervention:} \quad \begin{bmatrix} A_1 - \frac{B_{12}}{B_{11}}(\zeta - A_2) \\ \zeta \\ A_3 \end{bmatrix}$$

$$X^3 := \zeta \text{ intervention:} \quad \begin{bmatrix} A_1 - \left(\frac{B_{13}}{B_{11}} - \frac{B_{12}B_{23}}{B_{11}B_{22}}\right)(\zeta - A_3) \\ A_2 - \frac{B_{23}}{B_{22}}(\zeta - A_3) \\ \zeta \end{bmatrix}$$

Expressions for the stationary variance can also be obtained. The results agree with heuristic reasoning.

Identifiability of intervention distributions

Recall that in the classical (non-SDE) case, the DAG and thus the postintervention distributions are indeterminate from the observational distribution. In the SDE case, this is **not** typically the case.

Example. For the Ornstein-Uhlenbeck SDE

$$\mathrm{d}X_t = BX_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t,$$

the transition probabilities of moving from state x in time t is

$$P_t(x,\cdot) = \mathcal{N}\left(\exp(tB)x, \int_0^t \exp(sB)\sigma\sigma^t \exp(sB^t)\,\mathrm{d}s\right).$$

Thus, only $\sigma\sigma^t$ and not σ is determined from the distribution of X. However, explicit calculations show that intervention distributions **also** are determined only from $\sigma\sigma^t$ and not σ.

Identifiability of intervention distributions

This example can be extended to a more general context:

Theorem. Consider the two SDEs

$$\mathrm{d}X_t = a(X_{t-})\,\mathrm{d}Z_t$$

and

$$\mathrm{d}X_t = \tilde{a}(X_{t-})\,\mathrm{d}\tilde{Z}_t,$$

where Z and $\tilde{Z}$ are Lévy processes of dimension d and $\tilde{d}$, respectively, and a and $\tilde{a}$ are Lipschitz and bounded. Then:

1. For each of the two SDEs, all solutions (as the initial distribution varies) are Feller processes with the same semigroup.
2. If the semigroups for the two SDEs are equal, the postintervention distributions are the same when the initial distributions are the same.

Identifiability of intervention distributions

Conclusions from the theorem:

- For the case of SDEs driven by Lévy processes, postintervention distributions are uniquely determined by the semigroup.
- As the semigroup essentially is determined by the distribution (for example in the case where the solution process is irreducible), this in practice means that **postintervention distributions are identifiable from observational distributions**.
- As the zeroes of the coefficient a determines the causal relationship between the variables in the SDE, and a in general is not identifiable from the observational distribution, we also find that **postintervention distributions are identifiable even when the causal network is unidentifiable**.

Perspectives

- Solve control-theory type problems for optimal choice of interventions
- Understand **why** classical causal calculus works (Causal structure of SDE does not carry over to DAG structure of stationary distributions)
- For Ornstein-Uhlenbeck processes, apply penalized estimation methods to obtain sparse estimates of B and thus of the causal structure (Recall that zeroes identify the causal structure)

Thank you for your attention!