Causal Graphical Models

Table of Contents

Recap

- Causality is the study of the underlying structure of a system;
- Through studying causality and causal models, we can:
  - Answer questions about how a system responds to **changes** in its mechanisms;
  - Specify causal relationships that are more **stable** and more fundamental than generic statistical relationships.
- A model assigns truth values to statements about the system; a causal model assigns truth values to causal statements/queries

Types of queries

1. **Associations**: $p(Y|X = x)$

Types of queries

1. **Associations**: $p(Y|X = x)$
2. **Intervention**: $p(Y|do(X = x), Z = z)$ or $p(Y_{\{X=x\}}|Z = z)$

Types of queries

1. **Associations**: $p(Y|X = x)$
2. **Intervention**: $p(Y|do(X = x), Z = z)$ or $p(Y_{\{X=x\}}|Z = z)$
3. **Counterfactual**: $p(Y_{\{X=x'\}}|Y = y, X = x)$

Table of Contents

1. Statistical Models

Model: Set of random variables $(X_1, X_2, ..., X_N)$ follow a **joint probability distribution** $P(X_1, X_2, ..., X_N)$.

Queries: For two disjoint subsets $S, S' \subset \{1, ..., N\}$, we have

$$p(X_S | X_{S'}) = \frac{\sum_{i \notin (S \cup S')} p(X_1, X_2, ..., X_N)}{\sum_{i \notin S'} p(X_1, X_2, ..., X_N)} = \frac{p(X_{(S \cup S')})}{p(X_{S'})}$$

Representation: To represent a generic distribution over N binary variables, we need $O(2^N)$ space. The situation is even worse for continuous distributions (which don't have a parametric form like e.g. Multivariate Gaussian)...

1. Statistical Models

Conditional Independence

For a given probability distribution p, X is **conditionally independent** of Y given Z if:

$$p(x|y,z) = p(x|z) \text{ whenever } p(y,z) > 0$$

and this is written $X \perp\!\!\!\perp Y|Z$.

Conditional independences allow us to restrict the space of possible probability distributions in a structured way.

1. Statistical Models

Bayesian networks

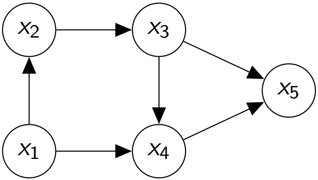
Bayesian Networks (BNs) are a compact means for representing a probability distribution, consisting of:

1. Statistical Models

Bayesian networks

Bayesian Networks (BNs) are a compact means for representing a probability distribution, consisting of:

- A directed acyclic graph (DAG) G:



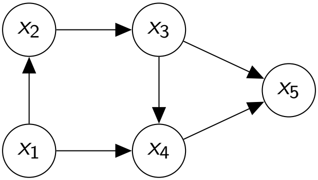
- A probability distribution $P(x|pa(x))$ for every node x in the DAG

1. Statistical Models

Bayesian networks

Bayesian Networks (BNs) are a compact means for representing a probability distribution, consisting of:

- A directed acyclic graph (DAG) G:



- A probability distribution $P(x|pa(x))$ for every node x in the DAG

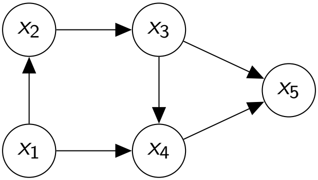
The joint probability distribution is then $P(x_1, ..., x_N) = \prod_i P(x_i|pa(x_i))$.

1. Statistical Models

Bayesian networks

Bayesian Networks (BNs) are a compact means for representing a probability distribution, consisting of:

- A directed acyclic graph (DAG) G:



- A probability distribution $P(x|pa(x))$ for every node x in the DAG

The joint probability distribution is then $P(x_1, ..., x_N) = \prod_i P(x_i|pa(x_i))$.

If a joint probability distribution q can be factorized in this way, we say that q is **Markov** relative to G.

1. Statistical Models

Simple DAGs

DAG	$X \perp\!\!\!\perp Y$	$X \perp\!\!\!\perp Y \mid Z$

1. Statistical Models

Simple DAGs

DAG	$X \perp\!\!\!\perp Y$	$X \perp\!\!\!\perp Y \mid Z$
$x \rightarrow z \rightarrow y$	No	Yes

1. Statistical Models

Simple DAGs

DAG	$X \perp\!\!\!\perp Y$	$X \perp\!\!\!\perp Y \mid Z$
$x \rightarrow z \rightarrow y$	No	Yes
$x \leftarrow z \rightarrow y$	No	Yes

1. Statistical Models

Simple DAGs

DAG	$X \perp\!\!\!\perp Y$	$X \perp\!\!\!\perp Y \mid Z$
$x \rightarrow z \rightarrow y$	No	Yes
$x \leftarrow z \rightarrow y$	No	Yes
$x \rightarrow z \leftarrow y$	Yes	No

1. Statistical Models

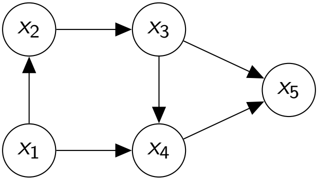
d-separation

We can characterize the restrictions a DAG makes on the joint distribution by the **conditional independences** it implies. **d-separation** is a graphical criterion for determining these conditional independences.

$X \perp\!\!\!\perp Y | Z$ is true in every distribution p Markov to G if:

all paths between X and Y are blocked by Z.

Example: $x_2 \perp\!\!\!\perp x_5 | \{x_3, x_4\}$



Inference

How do we compute quantities like $P(x_3|x_5)$ in a Bayesian network?

Inference

How do we compute quantities like $P(x_3|x_5)$ in a Bayesian network? In general inference in Bayesian networks is NP-complete.

- **Exact methods**: "Enumeration", junction-tree algorithm, cut-set conditioning
- **Approximate methods**: Approximate message passing, MCMC/Gibbs sampling, HMC, variational methods

1. Statistical Models

Summary

- **Bayesian networks** allow us to represent joint probability distributions efficiently, in the sense of:
  1. Less space required;
  2. Ability to specify assumptions/restrictions, which may aid learning;
  3. Faster inference (query answering) methods
- **d-separation** specifies the CIs implied by a DAG using the idea of "open" and "blocked" paths
- **Observational equivalence**: Two distinct DAGs may imply the same set of conditional independences. As a result, we can never distinguish between them using data alone.

2. Causal Models

Model: Set of **interventional** probability distributions
$\boldsymbol{P*} = \{P_{X=x}(v) : X \subseteq V\}$.

Queries: Any interventional probability, e.g. $p(Y|do(X), Z)$

Representation: Clearly, naïvely this is even more difficult to represent than the single joint probability distribution. However, DAGs provide a convenient way to handle this...

2. Causal Models

Causal Bayesian Networks

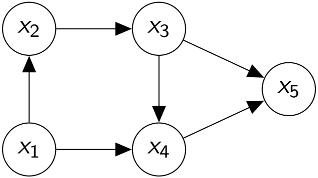
Causal Bayesian Networks are a means for representing a set of interventional distributions, consisting of:

2. Causal Models

Causal Bayesian Networks

Causal Bayesian Networks are a means for representing a set of interventional distributions, consisting of:

- A directed acyclic graph (DAG) G:

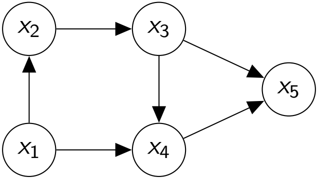


2. Causal Models

Causal Bayesian Networks

Causal Bayesian Networks are a means for representing a set of interventional distributions, consisting of:

- A directed acyclic graph (DAG) G:



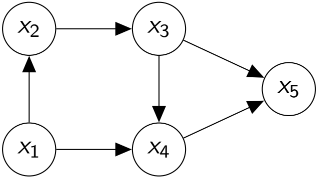
- A probability distribution $P(x|pa(x))$ for every node x in the DAG

2. Causal Models

Causal Bayesian Networks

Causal Bayesian Networks are a means for representing a set of interventional distributions, consisting of:

- A directed acyclic graph (DAG) G:



- A probability distribution $P(x|pa(x))$ for every node x in the DAG

The joint probability distribution is then $P(x_1, ..., x_N) = \prod_i P(x_i|pa(x_i))$.

2. Causal Models

Principle of Independent Causal Mechanisms

The **ICM principle** states that the individual causal mechanisms of a systems' causal generative process do not:

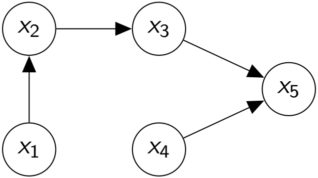
2. Causal Models

Principle of Independent Causal Mechanisms

The **ICM principle** states that the individual causal mechanisms of a systems' causal generative process do not:

- *inform* each other
- *influence* each other

To mimic this in our causal model, we derive interventional distributions by making the necessary intervention and leaving all other mechanisms.
Example: $P_{X_4=x'}(x_1, ..., x_5)$



$$P_{X_4=x'}(x_1, ..., x_5) = P(x_2|x_1)P(x_3|x_2)\mathbb{1}_{x_4=x'}P(x_5|x_3, x_4)$$

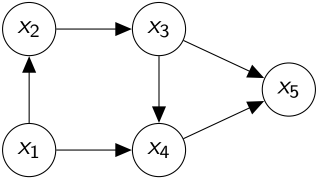
2. Causal Models

Inference

How do we compute $P_{X=x}(Y|Z)$ generally?

do-calculus: A set of rules for algebraically manipulating interventional expressions to obtain an formula which only uses probabilistic expressions involving observed variables.

Example: Backdoor adjustment $P_{X_4=x'}(x_5)$



Want a set of variables Z that block all backdoor paths from x_4 to x_5

$$P_{X_4=x'}(x_5) = \sum_{x_3} P(x_5|x_3, x')P(x_3)$$

2. Causal Models

Inference?

3. Functional Causal Model

Model: Modelled **generative process**

Queries: Any counterfactual probability, e.g. $p(Y_{X=x'}|Y=y)$

Representation: See next slide...

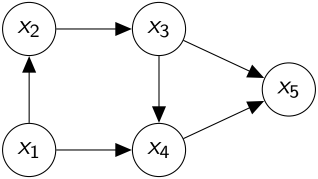
Structural Causal Models (SCMs) are a means for representing a generative process, consisting of:

3. Functional Causal Model

Structural Causal Models

Structural Causal Models (SCMs) are a means for representing a generative process, consisting of:

- A directed acyclic graph (DAG) G:

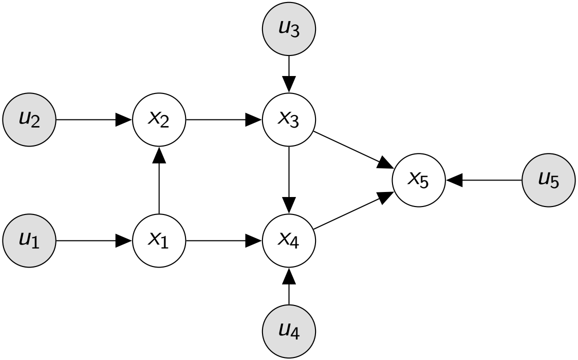


3. Functional Causal Model

Structural Causal Models

Structural Causal Models (SCMs) are a means for representing a generative process, consisting of:

- A directed acyclic graph (DAG) G:

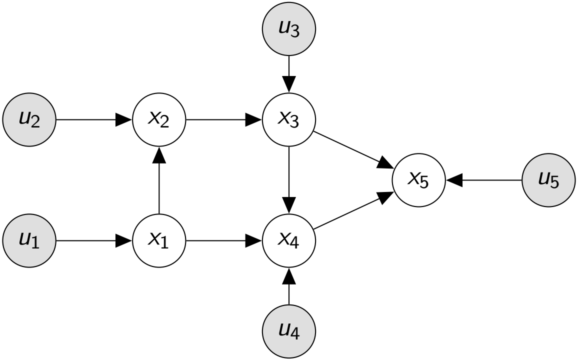


3. Functional Causal Model

Structural Causal Models

Structural Causal Models (SCMs) are a means for representing a generative process, consisting of:

- A directed acyclic graph (DAG) G:



- A functional equation $x_i = f_i(pa(x_i), u_i)$ for every node x_i;
- A distribution $P(u)$ over the "unobserved"/"noise" variables

How to compute $P(Y_{X=x'}|Z=z)$?
We write $< M, P(u) >$ to represent an SCM, where M is the functional equation model and $P(u)$ is the distribution over u.

1. **Abduction**: Compute $P(u|Z=z)$
2. **Action**: Modify the SCM by replacing the structural equation for $X = f_X(pa(X), u_X)$ with $X = x$
3. **Prediction**: Compute the conditional probability $P(Y|Z=z)$ **in the new SCM** $< M_{X=x}, P(u|Z=z) >$

Table of Contents

Learning

What might we want to learn?

- **Structure Learning**: Learning causal graphs based on observed data (and perhaps assumptions)
- **Learning conditional distributions**: $P(x|pa(x))$
- **Learning functional relatiomships**: $f(x|pa(x), u), P(u)$

Structure Learning

Recall: From data alone we cannot distinguish between a class of observationally equivalent DAGs.

PC algorithm

Assumptions:

- **Causal Sufficiency:** No hidden/latent variables
- **Causal Faithfulness**: If a conditional independence holds in the distribution, then we do have the corresponding d-separation (i.e. "simplest/most-restrictive DAG")

Steps:

1. **Learn the skeleton**
2. **Learn the v-structures**
3. **Direct the remaining edges**

Other structure learning algorithms

- Greedy equivalence search
- MMHC
- LINGAM ("gaussianness" of variables)
- BACKSHIFT