What Causality Is (stats for mathematicians)

Andrew Critch

UC Berkeley

August 31, 2011

Foreword: The value of examples

- With any hard question, it helps to start with **simple, concrete versions** of the question first.
- Another reason to focus on **concrete examples** is that they can be important in our everyday lives.
- I personally find "deep conversations" are more productive when both parties **try** and **insist on** having **concrete examples** to illustrate what they mean.

Foreword: The value of examples

- With any hard question, it helps to start with **simple, concrete versions** of the question first.
- Another reason to focus on **concrete examples** is that they can be important in our everyday lives.
- I personally find "deep conversations" are more productive when both parties **try** and **insist on** having **concrete examples** to illustrate what they mean.

Foreword: The value of examples

- With any hard question, it helps to start with **simple, concrete versions** of the question first.
- Another reason to focus on **concrete examples** is that they can be important in our everyday lives.
- I personally find “deep conversations” are more productive when both parties **try** and **insist on** having **concrete examples** to illustrate what they mean.

Foreword: The value of examples

- With any hard question, it helps to start with **simple, concrete versions** of the question first.
- Another reason to focus on **concrete examples** is that they can be important in our everyday lives.
- I personally find "deep conversations" are more productive when both parties **try** and **insist on** having **concrete examples** to illustrate what they mean.

Some causal questions

- Does smoking **cause** cancer? How much?
- Can lack of sleep **cause** obesity?
- How much does electricity reliability **affect** water transportation in California?
- Does religion **make** people happier?
- Is the fridge **keeping** my beer cold?

Some causal questions

- Does smoking **cause** cancer? How much?
- Can lack of sleep **cause** obesity?
- How much does electricity reliability **affect** water transportation in California?
- Does religion **make** people happier?
- Is the fridge **keeping** my beer cold?

Some causal questions

- Does smoking **cause** cancer? How much?
- Can lack of sleep **cause** obesity?
- How much does electricity reliability **affect** water transportation in California?
- Does religion **make** people happier?
- Is the fridge **keeping** my beer cold?

Some causal questions

- Does smoking **cause** cancer? How much?
- Can lack of sleep **cause** obesity?
- How much does electricity reliability **affect** water transportation in California?
- Does religion **make** people happier?
- Is the fridge **keeping** my beer cold?

Some causal questions

- Does smoking **cause** cancer? How much?
- Can lack of sleep **cause** obesity?
- How much does electricity reliability **affect** water transportation in California?
- Does religion **make** people happier?
- Is the fridge **keeping** my beer cold?

Some causal questions

- Does smoking **cause** cancer? How much?
- Can lack of sleep **cause** obesity?
- How much does electricity reliability **affect** water transportation in California?
- Does religion **make** people happier?
- Is the fridge **keeping** my beer cold?

Some causal questions

- Does smoking **cause** cancer? How much?
- Can lack of sleep **cause** obesity?
- How much does electricity reliability **affect** water transportation in California?
- Does religion **make** people happier?
- Is the fridge **keeping** my beer cold?

Outline

Coin-biasing games

Consider a game consisting of coin flips where earlier coin outcomes affect the biases of later coins in a prescribed way.

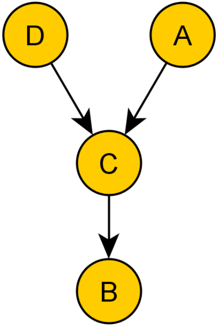
(Imagine I have some clear, heavy plastic that I can stick to the later coins to give them any bias I want, on the fly.)

Coin-biasing games

Consider a game consisting of coin flips where earlier coin outcomes affect the biases of later coins in a prescribed way.

(Imagine I have some clear, heavy plastic that I can stick to the later coins to give them any bias I want, on the fly.)

Example: "DACB"

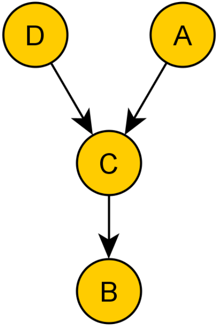


This diagram represents a coin-biasing game with 4 flips.

The first two coin flips are from fair coins, and the outcomes (0 or 1) are labelled D and A. (I'm using the letters out of sequence on purpose.)

Based on the outcome DA, a bias is chosen for another coin, which we flip and label its outcome C. Similarly C determines a bias for the B coin.

Example: "DACB"

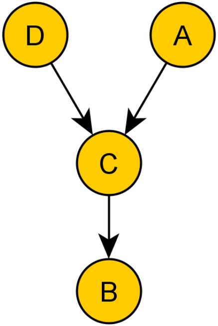


This diagram represents a coin-biasing game with 4 flips.

The first two coin flips are from fair coins, and the outcomes (0 or 1) are labelled D and A. (I'm using the letters out of sequence on purpose.)

Based on the outcome DA, a bias is chosen for another coin, which we flip and label its outcome C. Similarly C determines a bias for the B coin.

Example: “DACB”

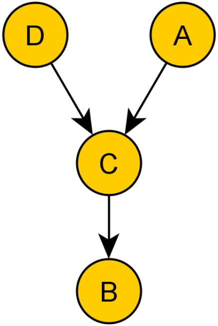


This diagram represents a coin-biasing game with 4 flips.

The first two coin flips are from fair coins, and the outcomes (0 or 1) are labelled D and A. (I’m using the letters out of sequence on purpose.)

Based on the outcome DA, a bias is chosen for another coin, which we flip and label its outcome C. Similarly C determines a bias for the B coin.

Example: "DACB"

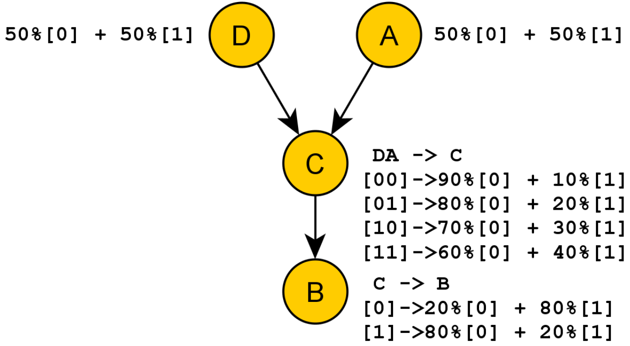


This diagram represents a coin-biasing game with 4 flips.

The first two coin flips are from fair coins, and the outcomes (0 or 1) are labelled D and A. (I'm using the letters out of sequence on purpose.)

Based on the outcome DA, a bias is chosen for another coin, which we flip and label its outcome C. Similarly C determines a bias for the B coin.

To fully specify the biasing game, we must augment our diagram with a list of biases:



Thus, a **coin-biasing game** is specified by data (V, G, Θ), where:

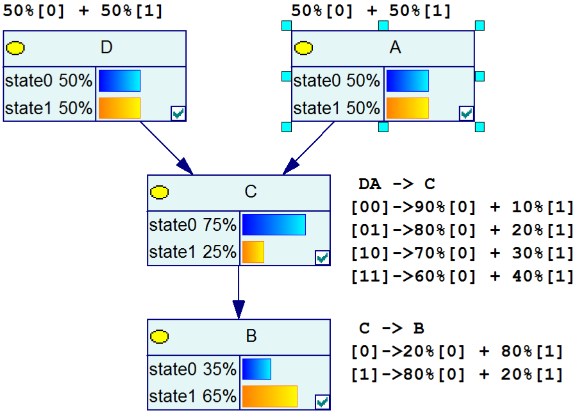
- V is a set of **binary variables**,
- G is a **directed acyclic graph** (DAG) called the **structure**, whose vertices are the variables, and
- Θ is a **conditional probability table** (CPT) specifying the values

$$P(V_i = v \mid \overline{parents}(V_i) = \overline{w})$$

for all i, v, and $\overline{w}$

Note: without the binarity restriction, this is the definition of a **Bayesian network** or **Bayes net** [J. Pearl, 1985].

If our “DACB” game were repeated many times, each time generating D and A with fair coins, and then C and B with the biases as prescribed above, the following marginal probabilities result:



Now suppose the "DACB" game is running inside a box, but we **don't know** its structure graph G or the CPT parameters Θ. Each time it runs, it prints us out a receipt showing the value of the variables A, B, C, and D, **in that order**, but nothing else:

1100
1000
1100
0100
1101
...

Say we got 50,000 such receipts, from which we estimate a **probability table** for the 16 possible outcomes...

Now suppose the "DACB" game is running inside a box, but we **don't know** its structure graph G or the CPT parameters Θ. Each time it runs, it prints us out a receipt showing the value of the variables A, B, C, and D, **in that order**, but nothing else:

1100
1000
1100
0100
1101
...

Say we got 50,000 such receipts, from which we estimate a **probability table** for the 16 possible outcomes...

$$
\begin{aligned}
P(0000) &= 0.0449, & P(1000) &= 0.0393,\\
P(0001) &= 0.0343, & P(1001) &= 0.0301,\\
P(0010) &= 0.0199, & P(1010) &= 0.0395,\\
P(0011) &= 0.0610, & P(1111) &= 0.0803,\\
P(0100) &= 0.1808, & P(1100) &= 0.1574,\\
P(0101) &= 0.1426, & P(1101) &= 0.1195,\\
P(0110) &= 0.0048, & P(1110) &= 0.0106,\\
P(0111) &= 0.0153, & P(1111) &= 0.0199
\end{aligned}
$$

From this **probability table** we can infer any **correlational** relationships we want. How about **causality**? Stats 101 quiz:

From the probabilities alone, can we infer G, the **causal structure** of the game? What extra information is needed?

$$
\begin{aligned}
P(0000) &= 0.0449, & P(1000) &= 0.0393,\\
P(0001) &= 0.0343, & P(1001) &= 0.0301,\\
P(0010) &= 0.0199, & P(1010) &= 0.0395,\\
P(0011) &= 0.0610, & P(1111) &= 0.0803,\\
P(0100) &= 0.1808, & P(1100) &= 0.1574,\\
P(0101) &= 0.1426, & P(1101) &= 0.1195,\\
P(0110) &= 0.0048, & P(1110) &= 0.0106,\\
P(0111) &= 0.0153, & P(1111) &= 0.0199
\end{aligned}
$$

From this **probability table** we can infer any **correlational** relationships we want. How about **causality**? Stats 101 quiz:

From the probabilities alone, can we infer G, the **causal structure** of the game? What extra information is needed?

$$P(0000) = 0.0449, \quad P(1000) = 0.0393,$$
$$P(0001) = 0.0343, \quad P(1001) = 0.0301,$$
$$P(0010) = 0.0199, \quad P(1010) = 0.0395,$$
$$P(0011) = 0.0610, \quad P(1111) = 0.0803,$$
$$P(0100) = 0.1808, \quad P(1100) = 0.1574,$$
$$P(0101) = 0.1426, \quad P(1101) = 0.1195,$$
$$P(0110) = 0.0048, \quad P(1110) = 0.0106,$$
$$P(0111) = 0.0153, \quad P(1111) = 0.0199$$

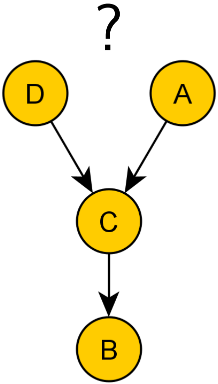
From this **probability table** we can infer any **correlational** relationships we want. How about **causality**? Stats 101 quiz:

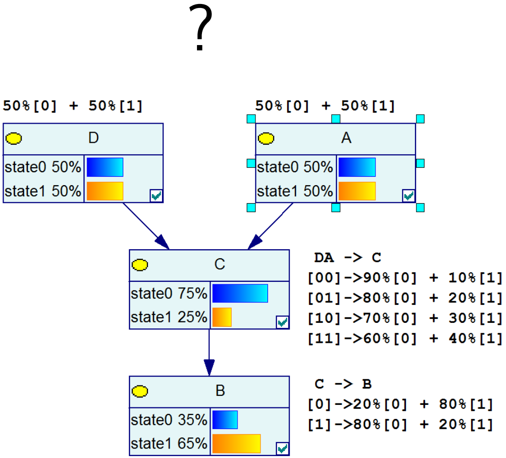
From the probabilities alone, can we infer G, the **causal structure** of the game? What extra information is needed?

$$
\begin{aligned}
P(0000) &= 0.0449, & P(1000) &= 0.0393,\\
P(0001) &= 0.0343, & P(1001) &= 0.0301,\\
P(0010) &= 0.0199, & P(1010) &= 0.0395,\\
P(0011) &= 0.0610, & P(1111) &= 0.0803,\\
P(0100) &= 0.1808, & P(1100) &= 0.1574,\\
P(0101) &= 0.1426, & P(1101) &= 0.1195,\\
P(0110) &= 0.0048, & P(1110) &= 0.0106,\\
P(0111) &= 0.0153, & P(1111) &= 0.0199
\end{aligned}
$$

From this **probability table** we can infer any **correlational** relationships we want. How about **causality**? Stats 101 quiz:

From the probabilities alone, can we infer G, the **causal structure** of the game? What extra information is needed?


?
D
A
C
B


?
50%[0] + 50%[1]
D
state0 50%
state1 50%
50%[0] + 50%[1]
A
state0 50%
state1 50%
C
state0 75%
state1 25%
DA -> C
[00]->90%[0] + 10%[1]
[01]->80%[0] + 20%[1]
[10]->70%[0] + 30%[1]
[11]->60%[0] + 40%[1]
B
state0 35%
state1 65%
C -> B
[0]->20%[0] + 80%[1]
[1]->80%[0] + 20%[1]

Yes We Can

The **probability data** alone is enough information to reliably infer the causal structure G of the "DACB" game.

The reason is that, by arising from G, the 16 **probabilities** $p_{000}, p_{0001}, \ldots p_{1111}$ are forced to satisfy a system of 13 **polynomial equations** $f_j = 0$ encoding **conditional independence** properties readable from the graph, which do not depend on the CPT Θ. [Pistone, Riccomagno, Wynn, 2001]

These equations are **almost never** satisfied by coin-biasing games arising from other graphs that aren't subgraphs of G, and coin-biasing games arising from G **almost never** satisfy conditional independence properties of its proper subgraphs.

Yes We Can

The **probability data** alone is enough information to reliably infer the causal structure G of the "DACB" game.

The reason is that, by arising from G, the 16 **probabilities** $p_{000}, p_{0001}, \ldots p_{1111}$ are forced to satisfy a system of 13 **polynomial equations** $f_j = 0$ encoding **conditional independence** properties readable from the graph, which do not depend on the CPT Θ. [Pistone, Riccomagno, Wynn, 2001]

These equations are **almost never** satisfied by coin-biasing games arising from other graphs that aren't subgraphs of G, and coin-biasing games arising from G **almost never** satisfy conditional independence properties of its proper subgraphs.

Yes We Can

The **probability data** alone is enough information to reliably infer the causal structure G of the "DACB" game.

The reason is that, by arising from G, the 16 **probabilities** $p_{000}, p_{0001}, \ldots p_{1111}$ are forced to satisfy a system of 13 **polynomial equations** $f_j = 0$ encoding **conditional independence** properties readable from the graph, which do not depend on the CPT Θ. [Pistone, Riccomagno, Wynn, 2001]

These equations are **almost never** satisfied by coin-biasing games arising from other graphs that aren't subgraphs of G, and coin-biasing games arising from G **almost never** satisfy conditional independence properties of its proper subgraphs.

Yes We Can

The **probability data** alone is enough information to reliably infer the causal structure G of the "DACB" game.

The reason is that, by arising from G, the 16 **probabilities** $p_{000}, p_{0001}, \ldots p_{1111}$ are forced to satisfy a system of 13 **polynomial equations** $f_j = 0$ encoding **conditional independence** properties readable from the graph, which do not depend on the CPT Θ. [Pistone, Riccomagno, Wynn, 2001]

These equations are **almost never** satisfied by coin-biasing games arising from other graphs that aren't subgraphs of G, and coin-biasing games arising from G **almost never** satisfy conditional independence properties of its proper subgraphs.

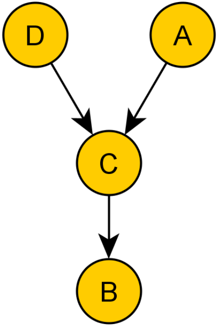
Yes We Can

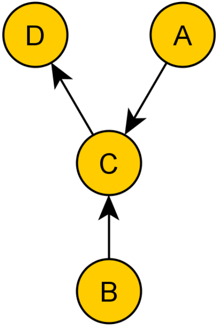
The **probability data** alone is enough information to reliably infer the causal structure G of the "DACB" game.

The reason is that, by arising from G, the 16 **probabilities** $p_{000}, p_{0001}, \ldots p_{1111}$ are forced to satisfy a system of 13 **polynomial equations** $f_j = 0$ encoding **conditional independence** properties readable from the graph, which do not depend on the CPT Θ. [Pistone, Riccomagno, Wynn, 2001]

These equations are **almost never** satisfied by coin-biasing games arising from other graphs that aren't subgraphs of G, and coin-biasing games arising from G **almost never** satisfy conditional independence properties of its proper subgraphs.

Yes We Can





$\Rightarrow A \perp\!\!\!\perp D, \quad AD \perp\!\!\!\perp B \mid C$

$\Rightarrow A \perp\!\!\!\perp B, \quad AB \perp\!\!\!\perp D \mid C$

Yes We Can

Moreover, except for some subtleties I'll address soon, this fortunate situation *almost* generalizes to other coin-biasing and even **die-biasing** games (i.e. the binarity assumption on the variables is not needed).

Andrew Critch, UC Berkeley
critch@math.berkeley.edu

So what is causality?

In short, it's the extent to which we can employ directed graphical models to predict and control real-world phenomena. I.e. it's how well we can **pretend nature is a die-biasing game**.

Definition [J. Pearl, 2000]

A (fully specified) **causal theory** is defined by an ordered triple (V, G, Θ): a set of variables, a DAG on the variables, and a compatible CPT. If not all of V, often a subset $O \subset V$ of **observed variables** is also specified, and the others are called **hidden variables**.

(Note: This formal framework is enough to discuss any other notion of causality I've seen, including all those listed on the Wikipedia and Stanford Encyclopedia of Philosophy entries.)

So what is causality?

In short, it's the extent to which we can employ directed graphical models to predict and control real-world phenomena. I.e. it's how well we can **pretend nature is a die-biasing game**.

Definition [J. Pearl, 2000]

A (fully specified) **causal theory** is defined by an ordered triple (V, G, Θ): a set of variables, a DAG on the variables, and a compatible CPT. If not all of V, often a subset $O \subset V$ of **observed variables** is also specified, and the others are called **hidden variables**.

(Note: This formal framework is enough to discuss any other notion of causality I've seen, including all those listed on the Wikipedia and Stanford Encyclopedia of Philosophy entries.)

So what is causality?

In short, it's the extent to which we can employ directed graphical models to predict and control real-world phenomena. I.e. it's how well we can **pretend nature is a die-biasing game**.

Definition [J. Pearl, 2000]

A (fully specified) **causal theory** is defined by an ordered triple (V, G, Θ): a set of variables, a DAG on the variables, and a compatible CPT. If not all of V, often a subset $O \subset V$ of **observed variables** is also specified, and the others are called **hidden variables**.

(Note: This formal framework is enough to discuss any other notion of causality I've seen, including all those listed on the Wikipedia and Stanford Encyclopedia of Philosophy entries.)

So what is causality?

In short, it's the extent to which we can employ directed graphical models to predict and control real-world phenomena. I.e. it's how well we can **pretend nature is a die-biasing game**.

Definition [J. Pearl, 2000]

A (fully specified) **causal theory** is defined by an ordered triple (V, G, Θ): a set of variables, a DAG on the variables, and a compatible CPT. If not all of V, often a subset $O \subset V$ of **observed variables** is also specified, and the others are called **hidden variables**.

(Note: This formal framework is enough to discuss any other notion of causality I've seen, including all those listed on the Wikipedia and Stanford Encyclopedia of Philosophy entries.)

So what is causality?

A **joint probability distribution** P on the variables V is generated by (G, Θ) in the obvious way (like a die-biasing game),

$$P(v_1 \dots v_n) = \prod_i P\left(v_i \,|\, parents(v_i)\right)$$

With this framework in place, we can say that

- **causal inference** is the problem of recovering (G, Θ) from the probabilities P or other partial information, and
- **causal hypotheses** are partial specifications of causal theories. For example, perhaps only (V, G) is described, or only part of G.

Causal hypotheses are used to make two kinds of predictions:

So what is causality?

A **joint probability distribution** P on the variables V is generated by (G, Θ) in the obvious way (like a die-biasing game),

$$P(v_1 \ldots v_n) = \prod_i P\left(v_i \mid parents(v_i)\right)$$

With this framework in place, we can say that

- **causal inference** is the problem of recovering (G, Θ) from the probabilities P or other partial information, and
- **causal hypotheses** are partial specifications of causal theories. For example, perhaps only (V, G) is described, or only part of G.

Causal hypotheses are used to make two kinds of predictions:

So what is causality?

A **joint probability distribution** P on the variables V is generated by (G, Θ) in the obvious way (like a die-biasing game),

$$P(v_1 \ldots v_n) = \prod_i P\left(v_i \,|\, parents(v_i)\right)$$

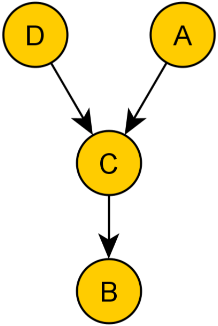
With this framework in place, we can say that

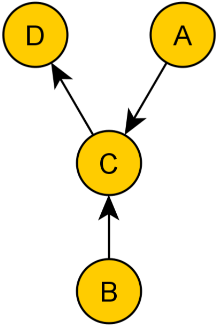
- **causal inference** is the problem of recovering (G, Θ) from the probabilities P or other partial information, and
- **causal hypotheses** are partial specifications of causal theories. For example, perhaps only (V, G) is described, or only part of G.

Causal hypotheses are used to make two kinds of predictions:

So what is causality?

1. **Observational predictions**, in the form of **conditional independence statements** implied by the graph G. E.g.,



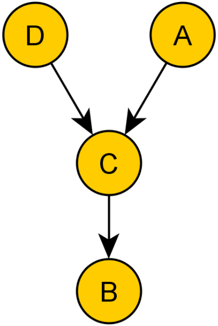


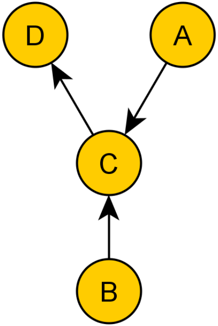
$\Rightarrow A \perp\!\!\!\perp D, \quad AD \perp\!\!\!\perp B \mid C$

$\Rightarrow A \perp\!\!\!\perp B, \quad AB \perp\!\!\!\perp D \mid C$

So what is causality?

1. **Observational predictions**, in the form of **conditional independence statements** implied by the graph G. E.g.,





$\Rightarrow A \perp\!\!\!\perp D, \quad AD \perp\!\!\!\perp B \mid C$

$\Rightarrow A \perp\!\!\!\perp B, \quad AB \perp\!\!\!\perp D \mid C$

So what is causality?

2. **Interventional predictions**: When we already believe in a context where we can **intervene** and fix one of the variables in mid-process, the graph G predicts which variables will respond, and the CPT Θ predicts *how*.

For example, in "DACB", if we catch coin C before it lands and set it to 1, then B will land 0 a lot more often, but A and D will remain fair coins.

Andrew Critch, UC Berkeley
critch@math.berkeley.edu

So what is causality?

2. **Interventional predictions**: When we already believe in a context where we can **intervene** and fix one of the variables in mid-process, the graph G predicts which variables will respond, and the CPT Θ predicts *how*.

   For example, in "DACB", if we catch coin C before it lands and set it to 1, then B will land 0 a lot more often, but A and D will remain fair coins.

Subtleties

Many mathematical subtleties arise in trying to infer the underlying graph of a causal process. Even die-biasing games on 3 and 4 variables have causally important mathematical properties that are highly non-intuitive.

- At least 3 variables are required to test any **causal** relationship, observationally. (I.e., on two variables/dice, only the DAG with no edge can be recovered, so $A \to B$ is indistinguishable from $B \to A$.)
- Not every causal structure G can be recovered uniquely from the outputs of a die-biasing game on it. Instead, DAGs come in small equivalence classes with other DAGs that are **"observationally indistinguishable"** from them.

Subtleties

Many mathematical subtleties arise in trying to infer the underlying graph of a causal process. Even die-biasing games on 3 and 4 variables have causally important mathematical properties that are highly non-intuitive.

- At least 3 variables are required to test any **causal** relationship, observationally. (I.e., on two variables/dice, only the DAG with no edge can be recovered, so $A \to B$ is indistinguishable from $B \to A$.)
- Not every causal structure G can be recovered uniquely from the outputs of a die-biasing game on it. Instead, DAGs come in small equivalence classes with other DAGs that are **"observationally indistinguishable"** from them.

Subtleties

Many mathematical subtleties arise in trying to infer the underlying graph of a causal process. Even die-biasing games on 3 and 4 variables have causally important mathematical properties that are highly non-intuitive.

- At least 3 variables are required to test any **causal** relationship, observationally. (I.e., on two variables/dice, only the DAG with no edge can be recovered, so $A \to B$ is indistinguishable from $B \to A$.)
- Not every causal structure G can be recovered uniquely from the outputs of a die-biasing game on it. Instead, DAGs come in small equivalence classes with other DAGs that are **"observationally indistinguishable"** from them.

Subtleties

- The **observational** equivalence class of a DAG is determined by its "collision structure", i.e. the occurrence of induced subgraphs of the form $A \to B \leftarrow C$.
- If there are variables whose outcomes we never observe, we might not notice they're there or how many states they have (although sometimes we can).

(Perhaps such combinatorial subtleties are the reason philosophers have been confused about causality for it for so long?)

Subtleties

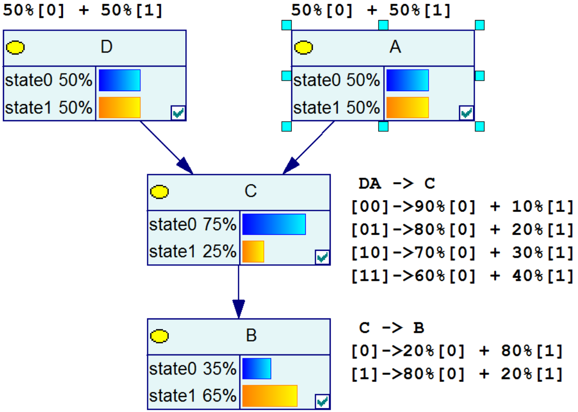
- The **observational** equivalence class of a DAG is determined by its "collision structure", i.e. the occurrence of induced subgraphs of the form $A \to B \leftarrow C$.
- If there are variables whose outcomes we never observe, we might not notice they're there or how many states they have (although sometimes we can).

(Perhaps such combinatorial subtleties are the reason philosophers have been confused about causality for it for so long?)

Subtleties

- The **observational** equivalence class of a DAG is determined by its "collision structure", i.e. the occurrence of induced subgraphs of the form $A \to B \leftarrow C$.
- If there are variables whose outcomes we never observe, we might not notice they're there or how many states they have (although sometimes we can).

(Perhaps such combinatorial subtleties are the reason philosophers have been confused about causality for it for so long?)


50%[0] + 50%[1]
D
state0 50%
state1 50%
50%[0] + 50%[1]
A
state0 50%
state1 50%
C
state0 75%
state1 25%
DA -> C
[00]->90%[0] + 10%[1]
[01]->80%[0] + 20%[1]
[10]->70%[0] + 30%[1]
[11]->60%[0] + 40%[1]
B
state0 35%
state1 65%
C -> B
[0]->20%[0] + 80%[1]
[1]->80%[0] + 20%[1]

Correlation $\neq$ causation, but...

Causal hypotheses:

- can still be made mathematically precise;
- imply testable predictions in form of **conditional independences** and **interventions**; and
- under the right circumstances can be reliably inferred from **probabilities** observed without **interventions** (controlled experiments).

Are there any objections to these statements?

Correlation $\neq$ causation, but...

Causal hypotheses:

- can still be made mathematically precise;
- imply testable predictions in form of **conditional independences** and **interventions**; and
- under the right circumstances can be reliably inferred from **probabilities** observed without **interventions** (controlled experiments).

Are there any objections to these statements?

Correlation $\neq$ causation, but...

Causal hypotheses:

- can still be made mathematically precise;
- imply testable predictions in form of **conditional independences** and **interventions**; and
- under the right circumstances can be reliably inferred from **probabilities** observed without **interventions** (controlled experiments).

Are there any objections to these statements?

Correlation $\neq$ causation, but...

Causal hypotheses:

- can still be made mathematically precise;
- imply testable predictions in form of **conditional independences** and **interventions**; and
- under the right circumstances can be reliably inferred from **probabilities** observed without **interventions** (controlled experiments).

Are there any objections to these statements?

Correlation $\neq$ causation, but...

Causal hypotheses:

- can still be made mathematically precise;
- imply testable predictions in form of **conditional independences** and **interventions**; and
- under the right circumstances can be reliably inferred from **probabilities** observed without **interventions** (controlled experiments).

Are there any objections to these statements?

Alert

When possible, **controlled experiments** remain an invaluable tool for testing a causal theory G: they can *distinguish graphs* that are observationally equivalent, they *test the additional hypothesis* that we can **intervene** in a way that respects the hypothesized **causal** structure, and can identify many observationally indistinguishable causal relations in the presence of **hidden variables**.

Remark

In practice, it is difficult to **falsify** a causal theory G from **probability observation** alone, because we usually have high prior confidence in the existence of **hidden variables**. But observed CI relations can still serve as strong evidence **in favor** of the theory, because they almost never occur unless the induced subgraph on the observed variables implies them.

Alert

When possible, **controlled experiments** remain an invaluable tool for testing a causal theory G: they can *distinguish graphs* that are observationally equivalent, they *test the additional hypothesis* that we can **intervene** in a way that respects the hypothesized **causal** structure, and can identify many observationally indistinguishable causal relations in the presence of **hidden variables**.

Remark

In practice, it is difficult to **falsify** a causal theory G from **probability observation** alone, because we usually have high prior confidence in the existence of **hidden variables**. But observed CI relations can still serve as strong evidence **in favor** of the theory, because they almost never occur unless the induced subgraph on the observed variables implies them.

Perhaps the best feature of graphical models:

Such claims can be made precise by writing down the graphs, and disputed accordingly (which is great for mathematicians, who usually hate talking about subjects that lack precise notation.)

A hope for observational sciences

I personally hope that such discoveries can inform public policy and medical decisions on questions previously considered unanswerable.

Perhaps the best feature of graphical models:

Such claims can be made precise by writing down the graphs, and disputed accordingly (which is great for mathematicians, who usually hate talking about subjects that lack precise notation.)

A hope for observational sciences

I personally hope that such discoveries can inform public policy and medical decisions on questions previously considered unanswerable.

Why didn't anybody tell me?

Probably because they didn't know either!

Graphical modeling is by far the most structured and rigorous framework for understanding causality to date, and it's not very old. Major advancements occurred in the late 1980s, 90s and 2000s, and there's still a lot of work to be done...

Why didn't anybody tell me?

Probably because they didn't know either!

Graphical modeling is by far the most structured and rigorous framework for understanding causality to date, and it's not very old. Major advancements occurred in the late 1980s, 90s and 2000s, and there's still a lot of work to be done...

Some history / seminal texts

- Pearl (1988) *Probabilistic Reasoning in Intelligent Systems* (Graphs becoming popular in CS for **subjective belief propagation** networks, or Bayes nets.)
- Lauritzen (1996) *Graphical Models*. (Beginning to view graphs as **generative processes** underlying statistical theories.)
- Pearl (2000) *Causality: Models, Reasoning, and Inference*. (Pearl advocating understanding graphical models, as a **framework** for stating statistical theories, by essentially all scientists and medical professionals.)
- Pistone, Riccomango, Wynn (2001) *Algebraic Statistics*. (**Computational commutative algebra** being recognized as a tool for studying the structure of graphical model predictions.)

Some history / seminal texts

- Pearl (1988) *Probabilistic Reasoning in Intelligent Systems* (Graphs becoming popular in CS for **subjective belief propagation** networks, or Bayes nets.)
- Lauritzen (1996) *Graphical Models*. (Beginning to view graphs as **generative processes** underlying statistical theories.)
- Pearl (2000) *Causality: Models, Reasoning, and Inference*. (Pearl advocating understanding graphical models, as a **framework** for stating statistical theories, by essentially all scientists and medical professionals.)
- Pistone, Riccomango, Wynn (2001) *Algebraic Statistics*. (**Computational commutative algebra** being recognized as a tool for studying the structure of graphical model predictions.)

Some history / seminal texts

- Pearl (1988) *Probabilistic Reasoning in Intelligent Systems* (Graphs becoming popular in CS for **subjective belief propagation** networks, or Bayes nets.)
- Lauritzen (1996) *Graphical Models.* (Beginning to view graphs as **generative processes** underlying statistical theories.)
- Pearl (2000) *Causality: Models, Reasoning, and Inference.* (Pearl advocating understanding graphical models, as a **framework** for stating statistical theories, by essentially all scientists and medical professionals.)
- Pistone, Riccomango, Wynn (2001) *Algebraic Statistics.* (**Computational commutative algebra** being recognized as a tool for studying the structure of graphical model predictions.)

Some history / seminal texts

- Pearl (1988) *Probabilistic Reasoning in Intelligent Systems* (Graphs becoming popular in CS for **subjective belief propagation** networks, or Bayes nets.)
- Lauritzen (1996) *Graphical Models.* (Beginning to view graphs as **generative processes** underlying statistical theories.)
- Pearl (2000) *Causality: Models, Reasoning, and Inference*. (Pearl advocating understanding graphical models, as a **framework** for stating statistical theories, by essentially all scientists and medical professionals.)
- Pistone, Riccomango, Wynn (2001) *Algebraic Statistics.* (**Computational commutative algebra** being recognized as a tool for studying the structure of graphical model predictions.)

Some history / seminal texts

- Pearl (1988) *Probabilistic Reasoning in Intelligent Systems* (Graphs becoming popular in CS for **subjective belief propagation** networks, or Bayes nets.)
- Lauritzen (1996) *Graphical Models*. (Beginning to view graphs as **generative processes** underlying statistical theories.)
- Pearl (2000) *Causality: Models, Reasoning, and Inference*. (Pearl advocating understanding graphical models, as a **framework** for stating statistical theories, by essentially all scientists and medical professionals.)
- Pistone, Riccomango, Wynn (2001) *Algebraic Statistics*. (**Computational commutative algebra** being recognized as a tool for studying the structure of graphical model predictions.)

Some history / seminal texts

- Pachter, Sturmfels, (2005) *Algebraic Statistics for Computational Biology* Cambridge University Press.
- Drton, Sturmfels, Sullivant (2009) *Lectures on Algebraic Statistics*, Springer.

Some history / seminal texts

- Pachter, Sturmfels, (2005) *Algebraic Statistics for Computational Biology* Cambridge University Press.

- Drton, Sturmfels, Sullivant (2009) *Lectures on Algebraic Statistics*, Springer.

It's time to see some math in action!