A New Causal Power Theory

Univ of Kent
2008

Kevin Korb[1] Erik Nyberg[2]

[1]Clayton School of IT
Monash University

[2]History and Philosophy of Science
University of Melbourne

Contents

Causal Power Theories

- S Wright (1934) The method of path coefficients. *Ann Math Stat, 5*, 161-215.
- IJ Good (1961) A causal calculus. *BJPS, 11,* 305-318.
- P Cheng (1997) From covariation to causation. *Psych Rev, 104,* 367-405.
- C Glymour & P Cheng (1998) Causal mechanism and probability. Oaksford and Chater (Eds.) *Rational models of cognition.* Oxford.
- C Glymour (2001) *The Mind's Arrows*. MIT Press.
- C Hitchcock (2001) The intransitivity of causation. *JP, 98,* 273-299.
- E Hiddleston (2005) Causal powers. *BJPS, 56,* 27-59.
- L Hope and K Korb (2005) An Information-theoretic causal power theory. *Australian AI Conference,* pp. 805-811. Springer.

Causal Power

What is *causal power*?

The power of some event to bring about (prevent) another event.

Examples

- The power of anticoagulants to prevent death from heart attack.
- The power of exercise to prevent heart attacks.
- The power of a doctor's advice to exercise to bring about exercise.
- The power of a doctor's advice to exercise to prevent heart attacks.

Causal Power

What is *causal power*?

> *The power of some event to bring about (prevent) another event.*

Examples

- The power of anticoagulants to prevent death from heart attack.
- The power of exercise to prevent heart attacks.
- The power of a doctor's advice to exercise to bring about exercise.
- The power of a doctor's advice to exercise to prevent heart attacks.

Causal Power

What is *causal power*?

> *The power of some event to bring about (prevent) another event.*

Examples

- The power of anticoagulants to prevent death from heart attack.
- The power of exercise to prevent heart attacks.
- The power of a doctor's advice to exercise to bring about exercise.
- The power of a doctor's advice to exercise to prevent heart attacks.

Causal Power

What is *causal power*?

> *The power of some event to bring about (prevent) another event.*

Examples

- The power of anticoagulants to prevent death from heart attack.
- The power of exercise to prevent heart attacks.
- The power of a doctor's advice to exercise to bring about exercise.
- The power of a doctor's advice to exercise to prevent heart attacks.

Causal Power

What is *causal power*?

> *The power of some event to bring about (prevent) another event.*

Examples

- The power of anticoagulants to prevent death from heart attack.
- The power of exercise to prevent heart attacks.
- The power of a doctor's advice to exercise to bring about exercise.
- The power of a doctor's advice to exercise to prevent heart attacks.

Causal Power

What is *causal power*?

> *The power of some event to bring about (prevent) another event.*

Examples

- The power of anticoagulants to prevent death from heart attack.
- The power of exercise to prevent heart attacks.
- The power of a doctor's advice to exercise to bring about exercise.
- The power of a doctor's advice to exercise to prevent heart attacks.

Causal Power

One important point:

> *Causal power is always relative to a reference class.*

- The power of the pill to prevent pregnancy
  - Amongst women
  - Amongst men
- The power of extra exercise to prevent heart attacks.
  - Amongst middle-aged couch potatoes
  - Amongst athletes
  - Amongst teenagers

Usually the reference class is implicit.

Causal Power

One important point:

Causal power is always relative to a reference class.

- The power of the pill to prevent pregnancy
  - Amongst women
  - Amongst men
- The power of extra exercise to prevent heart attacks.
  - Amongst middle-aged couch potatoes
  - Amongst athletes
  - Amongst teenagers

Usually the reference class is implicit.

Causal Power

One important point:

> *Causal power is always relative to a reference class.*

- The power of the pill to prevent pregnancy
  - Amongst women
  - Amongst men
- The power of extra exercise to prevent heart attacks.
  - Amongst middle-aged couch potatoes
  - Amongst athletes
  - Amongst teenagers

Usually the reference class is implicit.

Causal Power

One important point:

Causal power is always relative to a reference class.

- The power of the pill to prevent pregnancy
  - Amongst women
  - Amongst men
- The power of extra exercise to prevent heart attacks.
  - Amongst middle-aged couch potatoes
  - Amongst athletes
  - Amongst teenagers

Usually the reference class is implicit.

Causal Power

One important point:

> *Causal power is always relative to a reference class.*

- The power of the pill to prevent pregnancy
  - Amongst women
  - Amongst men
- The power of extra exercise to prevent heart attacks.
  - Amongst middle-aged couch potatoes
  - Amongst athletes
  - Amongst teenagers

Usually the reference class is implicit.

Causal Power

One important point:

Causal power is always relative to a reference class.

- The power of the pill to prevent pregnancy
  - Amongst women
  - Amongst men
- The power of extra exercise to prevent heart attacks.
  - Amongst middle-aged couch potatoes
  - Amongst athletes
  - Amongst teenagers

Usually the reference class is implicit.

Causal Power

One important point:

Causal power is always relative to a reference class.

- The power of the pill to prevent pregnancy
  - Amongst women
  - Amongst men
- The power of extra exercise to prevent heart attacks.
  - Amongst middle-aged couch potatoes
  - Amongst athletes
  - Amongst teenagers

Usually the reference class is implicit.

Causal Power

One important point:

> *Causal power is always relative to a reference class.*

- The power of the pill to prevent pregnancy
  - Amongst women
  - Amongst men
- The power of extra exercise to prevent heart attacks.
  - Amongst middle-aged couch potatoes
  - Amongst athletes
  - Amongst teenagers

Usually the reference class is implicit.

Causal Power

One important point:

> *Causal power is always relative to a reference class.*

- The power of the pill to prevent pregnancy
  - Amongst women
  - Amongst men
- The power of extra exercise to prevent heart attacks.
  - Amongst middle-aged couch potatoes
  - Amongst athletes
  - Amongst teenagers

Usually the reference class is implicit.

Causal Power

We should like to develop an explicit quantitative measure of causal power, generalizing (improving on) our intuitive judgments.

- Stochastic causality comes in degrees ("effect size" in medicine)
- Potentially allowing for precise judgments of causal attribution
  - hence, the interest of cog psych
- Clarifying the explanatory import of causal Bayesian networks

Causal Power

We should like to develop an explicit quantitative measure of causal power, generalizing (improving on) our intuitive judgments.

- Stochastic causality comes in degrees ("effect size" in medicine)
- Potentially allowing for precise judgments of causal attribution
  - hence, the interest of cog psych
- Clarifying the explanatory import of causal Bayesian networks

Causal Power

We should like to develop an explicit quantitative measure of causal power, generalizing (improving on) our intuitive judgments.

- Stochastic causality comes in degrees (“effect size” in medicine)
- Potentially allowing for precise judgments of causal attribution
    - hence, the interest of cog psych
- Clarifying the explanatory import of causal Bayesian networks

Causal Power

We should like to develop an explicit quantitative measure of causal power, generalizing (improving on) our intuitive judgments.

- Stochastic causality comes in degrees ("effect size" in medicine)
- Potentially allowing for precise judgments of causal attribution
  - hence, the interest of cog psych
- Clarifying the explanatory import of causal Bayesian networks

Causal Power

We should like to develop an explicit quantitative measure of causal power, generalizing (improving on) our intuitive judgments.

- Stochastic causality comes in degrees ("effect size" in medicine)
- Potentially allowing for precise judgments of causal attribution
  - hence, the interest of cog psych
- Clarifying the explanatory import of causal Bayesian networks

Path Models

Most theories of causal power are based on binary networks (Cheng, Glymour, Hiddleston).

The first theory, Wright (1934), uses standardized linear Gaussian models: path models.

Desideratum 1

Causal power theory should apply to **any** kind of causal Bayesian network – linear, binomial, multinomial.

Path Models

Most theories of causal power are based on binary networks (Cheng, Glymour, Hiddleston).

The first theory, Wright (1934), uses standardized linear Gaussian models: path models.

Desideratum 1

Causal power theory should apply to **any** kind of causal Bayesian network – linear, binomial, multinomial.

Path Models

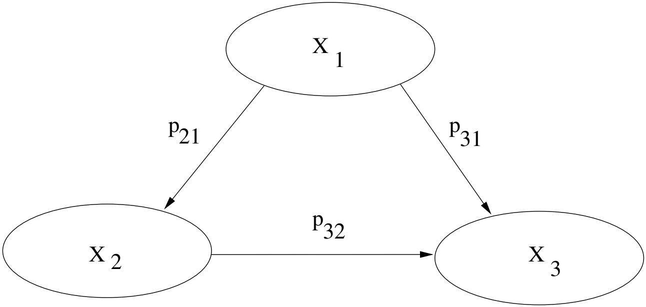
Most theories of causal power are based on binary networks (Cheng, Glymour, Hiddleston).

The first theory, Wright (1934), uses standardized linear Gaussian models: path models.

Desideratum 1

Causal power theory should apply to **any** kind of causal Bayesian network – linear, binomial, multinomial.

Path Models



r	1	2	3
1	1		
2	r_{12}	1	
3	r_{13}	r_{23}	1

Path Models

Theorem (Explained Variation)

Path coefficients are equal to the square root of the variation in the child variable attributable to the parent.

I.e.,

$$\sum_i p_{ji}^2 = 1$$

- As a consequence of standardization
- Requires a residual term U with coefficient p_{ju}

Path Models

Theorem (Explained Variation)

Path coefficients are equal to the square root of the variation in the child variable attributable to the parent.

I.e.,

$$\sum_i p_{ji}^2 = 1$$

- As a consequence of standardization
- Requires a residual term U with coefficient p_{ju}

Path Models

Theorem (Explained Variation)

Path coefficients are equal to the square root of the variation in the child variable attributable to the parent.

I.e.,

$$\sum_i p_{ji}^2 = 1$$

- As a consequence of standardization
- Requires a residual term U with coefficient p_{ju}

Path Models

Theorem (Explained Variation)

Path coefficients are equal to the square root of the variation in the child variable attributable to the parent.

I.e.,

$$\sum_i p_{ji}^2 = 1$$

- As a consequence of standardization
- Requires a residual term U with coefficient p_{jU}

Wright's Decomposition Rule

Wright developed a graphical rule for relating (observed) correlations with path coefficients (i.e., relating probability and causality).

> *Fundamental idea: correlation results from causal influence along certain paths between variables.*

Definition (Admissible Path)

Φ_k is an **admissible path** between X_i and X_j iff it is an undirected path connecting X_i and X_j s.t. it does not go against the direction of an arc *after* having gone forward.

Wright's Decomposition Rule

Wright developed a graphical rule for relating (observed) correlations with path coefficients (i.e., relating probability and causality).

> *Fundamental idea: correlation results from causal influence along certain paths between variables.*

Definition (Admissible Path)

Φ_k is an **admissible path** between X_i and X_j iff it is an undirected path connecting X_i and X_j s.t. it does not go against the direction of an arc *after* having gone forward.

Wright's Decomposition Rule

Wright developed a graphical rule for relating (observed) correlations with path coefficients (i.e., relating probability and causality).

> *Fundamental idea: correlation results from causal influence along certain paths between variables.*

Definition (Admissible Path)

Φ_k is an **admissible path** between X_i and X_j iff it is an undirected path connecting X_i and X_j s.t. it does not go against the direction of an arc *after* having gone forward.

Wright's Decomposition Rule

This can be thought of as 3 rules in 1 for defining paths supporting causal influence:

1. Directed chains support causal influence
2. Common ancestors support causal influence between descendants
3. Common descendants don't support causal influence between ancestors

(This prefigures Pearl's d-separation rules.)

Wright's Decomposition Rule

To assess the strength of causal influence along an admissible path:

Definition (Valuation)

The valuation of a path is

$$v(\Phi_k) = \prod_{lm} p_{lm} \text{ for all } X_m \rightarrow X_l \in \Phi_k$$

Wright's Decomposition Rule

To assess the strength of causal influence along an admissible path:

Definition (Valuation)

The valuation of a path is

$$v(\Phi_k) = \prod_{lm} p_{lm} \text{ for all } X_m \rightarrow \ X_l \in \Phi_k$$

Wright's Decomposition Rule

Theorem (Wright's Decomposition Rule)

The correlation r_{ij} between variables X_i and X_j, where X_i is an ancestor of X_j, can be rewritten as:

$$r_{ij} = \sum_k v(\Phi_k)$$

where Φ_k is an admissible path between X_i and X_j and $v(\cdot)$ is a valuation of that path.

Wright's Decomposition Rule

This gives a direct relation between path coefficients and correlations:

$$
\begin{aligned}
r_{12} &= p_{21} \\
r_{13} &= p_{31} + p_{21}p_{32} \\
r_{23} &= p_{32} + p_{21}p_{31}
\end{aligned}
$$

We can solve for the p_{ij}:

$$
\begin{aligned}
p_{21} &= r_{12} \\
p_{31} &= \frac{r_{13} - r_{23}r_{12}}{1 - r_{12}^2} \\
p_{32} &= \frac{r_{23} - r_{13}r_{12}}{1 - r_{12}^2}
\end{aligned}
$$

Hence, we can parameterize (identify) any (recursive) path model, given a correlation table.

Wright's Decomposition Rule

This gives a direct relation between path coefficients and correlations:

$$
\begin{aligned}
r_{12} &= p_{21} \\
r_{13} &= p_{31} + p_{21}p_{32} \\
r_{23} &= p_{32} + p_{21}p_{31}
\end{aligned}
$$

We can solve for the p_{ij}:

$$
\begin{aligned}
p_{21} &= r_{12} \\
p_{31} &= \frac{r_{13} - r_{23}r_{12}}{1 - r_{12}^2} \\
p_{32} &= \frac{r_{23} - r_{13}r_{12}}{1 - r_{12}^2}
\end{aligned}
$$

Hence, we can parameterize (identify) any (recursive) path model, given a correlation table.

Wright's Decomposition Rule

This gives a direct relation between path coefficients and correlations:

$$
\begin{aligned}
r_{12} &= p_{21} \\
r_{13} &= p_{31} + p_{21}p_{32} \\
r_{23} &= p_{32} + p_{21}p_{31}
\end{aligned}
$$

We can solve for the p_{ij}:

$$
\begin{aligned}
p_{21} &= r_{12} \\
p_{31} &= \frac{r_{13} - r_{23}r_{12}}{1 - r_{12}^2} \\
p_{32} &= \frac{r_{23} - r_{13}r_{12}}{1 - r_{12}^2}
\end{aligned}
$$

Hence, we can parameterize (identify) any (recursive) path model, given a correlation table.

Wright's Power Theory

Wright's implicit causal power theory:

The causal power of C for E is:

$$CP(C,E) = \sum_k \prod_{lm} p_{lm} \quad \begin{array}{l} \text{for all } X_m \rightarrow X_l \in \Phi_k \\ \text{for all } \Phi_k = C \rightarrow \ldots \rightarrow E \end{array}$$

NB: This is *implicit* in Wright's treatment; Wright had no explicit causal power theory.

Wright's Power Theory

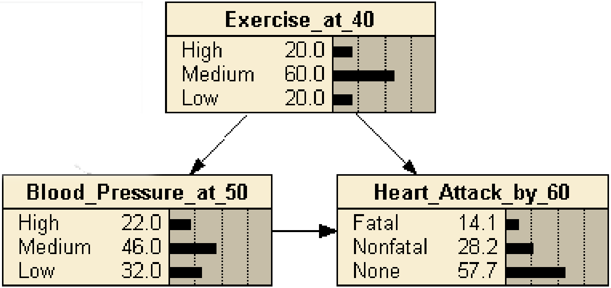
Wright's implicit causal power theory:

The causal power of C for E is:

$$CP(C,E) = \sum_k \prod_{lm} p_{lm} \quad \begin{array}{l} \text{for all } X_m \rightarrow \; X_l \in \Phi_k \\ \text{for all } \Phi_k = C \; \rightarrow \; \ldots \rightarrow \; E \end{array}$$

NB: This is *implicit* in Wright's treatment; Wright had no explicit causal power theory.

Heart Attack Example



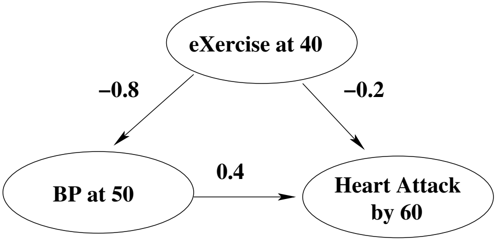
So: What is the causal power of BP for HA?

Note:

- Backpath BP $\leftarrow$ X $\rightarrow$ HA
- Messy interaction btw BP and X upon HA

Heart Attack Example

Consider the linear approximation (dropping the messy interaction):



$$r_{BP,HA} = p_{BP,X}p_{HA,X} + p_{HA,BP} = 0.56$$

The Wright causal power of BP for HA

- Discounts the backpath BP $\leftarrow$ X $\rightarrow$ HA
- Equals 0.4

Wright's Power Theory

- Relates *variables* C and E, not their *values*
  - To relate values, we should need to discretize variable ranges in some way
- Wright's theory has been very successful
- Wright's theory is compatible with current Bayesian network theory

Desideratum 2

Causal power theory should generalize Wright's power theory.

Wright's Power Theory

- Relates *variables* C and E, not their *values*
  - To relate values, we should need to discretize variable ranges in some way
- Wright's theory has been very successful
- Wright's theory is compatible with current Bayesian network theory

Desideratum 2

Causal power theory should generalize Wright's power theory.

Wright's Power Theory

- Relates *variables* C and E, not their *values*
  - To relate values, we should need to discretize variable ranges in some way
- Wright's theory has been very successful
- Wright's theory is compatible with current Bayesian network theory

Desideratum 2

Causal power theory should generalize Wright's power theory.

Wright's Power Theory

- Relates *variables* C and E, not their *values*
  - To relate values, we should need to discretize variable ranges in some way
- Wright's theory has been very successful
- Wright's theory is compatible with current Bayesian network theory

Desideratum 2

Causal power theory should generalize Wright's power theory.

Wright's Power Theory

- Relates *variables* C and E, not their *values*
  - To relate values, we should need to discretize variable ranges in some way
- Wright's theory has been very successful
- Wright's theory is compatible with current Bayesian network theory

Desideratum 2

Causal power theory should generalize Wright's power theory.

Modern Causal Power Theory

Cheng & Glymour

The Cheng (1997) and Glymour & Cheng (1998) PC Theory applies to binary variables taking particular values, $C = c$ and $E = e$, given assumptions:

- $\exists$ a direct causal connection $C \rightarrow E$
- C is independent of any other cause of E
- C does not interact with any other cause of E
- Probabilistic relevance:

  $$\Delta P = p(e|c) - p(e|\neg c) \neq 0$$
- Spurious causes must be eliminated
  - e.g., replaced by common causes

(Echoing Salmon on SR explanation and Suppes on probabilistic causation)

Modern Causal Power Theory

Cheng & Glymour

The Cheng (1997) and Glymour & Cheng (1998) PC Theory applies to binary variables taking particular values, $C = c$ and $E = e$, given assumptions:

- $\exists$ a direct causal connection $C \rightarrow E$
- C is independent of any other cause of E
- C does not interact with any other cause of E
- Probabilistic relevance:

  $$\Delta P = p(e|c) - p(e|\neg c) \neq 0$$
- Spurious causes must be eliminated
  - e.g., replaced by common causes

(Echoing Salmon on SR explanation and Suppes on probabilistic causation)

Modern Causal Power Theory

Cheng & Glymour

The Cheng (1997) and Glymour & Cheng (1998) PC Theory applies to binary variables taking particular values, $C = c$ and $E = e$, given assumptions:

- $\exists$ a direct causal connection $C \rightarrow E$
- C is independent of any other cause of E
- C does not interact with any other cause of E
- Probabilistic relevance:

  $$\Delta P = p(e|c) - p(e|\neg c) \neq 0$$

- Spurious causes must be eliminated
  - e.g., replaced by common causes

(Echoing Salmon on SR explanation and Suppes on probabilistic causation)

Modern Causal Power Theory

Cheng & Glymour

The Cheng (1997) and Glymour & Cheng (1998) PC Theory applies to binary variables taking particular values, $C = c$ and $E = e$, given assumptions:

- $\exists$ a direct causal connection $C \rightarrow E$
- C is independent of any other cause of E
- C does not interact with any other cause of E
- Probabilistic relevance:

$$\Delta P = p(e|c) - p(e|\neg c) \neq 0$$

- Spurious causes must be eliminated
  - e.g., replaced by common causes

(Echoing Salmon on SR explanation and Suppes on probabilistic causation)

Modern Causal Power Theory

Cheng & Glymour

The Cheng (1997) and Glymour & Cheng (1998) PC Theory applies to binary variables taking particular values, $C = c$ and $E = e$, given assumptions:

- $\exists$ a direct causal connection $C \rightarrow E$
- C is independent of any other cause of E
- C does not interact with any other cause of E
- Probabilistic relevance:

$$\Delta P = p(e|c) - p(e|\neg c) \neq 0$$

- Spurious causes must be eliminated
  - e.g., replaced by common causes

(Echoing Salmon on SR explanation and Suppes on probabilistic causation)

Modern Causal Power Theory

Cheng & Glymour

The Cheng (1997) and Glymour & Cheng (1998) PC Theory applies to binary variables taking particular values, $C = c$ and $E = e$, given assumptions:

- $\exists$ a direct causal connection $C \rightarrow E$
- C is independent of any other cause of E
- C does not interact with any other cause of E
- Probabilistic relevance:

  $$\Delta P = p(e|c) - p(e|\neg c) \neq 0$$
- Spurious causes must be eliminated
  - e.g., replaced by common causes

(Echoing Salmon on SR explanation and Suppes on probabilistic causation)

Modern Causal Power Theory

Cheng & Glymour

The Cheng (1997) and Glymour & Cheng (1998) PC Theory applies to binary variables taking particular values, $C = c$ and $E = e$, given assumptions:

- $\exists$ a direct causal connection $C \rightarrow E$
- C is independent of any other cause of E
- C does not interact with any other cause of E
- Probabilistic relevance:

  $$\Delta P = p(e|c) - p(e|\neg c) \neq 0$$
- Spurious causes must be eliminated
  - e.g., replaced by common causes

(Echoing Salmon on SR explanation and Suppes on probabilistic causation)

Modern Causal Power Theory

Cheng & Glymour

The Cheng (1997) and Glymour & Cheng (1998) PC Theory applies to binary variables taking particular values, $C = c$ and $E = e$, given assumptions:

- $\exists$ a direct causal connection $C \rightarrow E$
- C is independent of any other cause of E
- C does not interact with any other cause of E
- Probabilistic relevance:
  $$\Delta P = p(e|c) - p(e|\neg c) \neq 0$$
- Spurious causes must be eliminated
  - e.g., replaced by common causes

(Echoing Salmon on SR explanation and Suppes on probabilistic causation)

"Power PC" Theory

Definition (Causal Power)

For positive ΔP (generative cause), the power of c to bring about e:

$$p_c = \frac{\Delta P}{1 - P(e|\neg c)}$$

Idea: ΔP directly is not a fair measure of p_c

- since there is a background rate $P(e|\neg c)$
- ΔP should be relativized to the remainder
  - those cases that would have been $\neg e$ but for c

"Power PC" Theory

Definition (Causal Power)

For positive ΔP (generative cause), the power of c to bring about e:

$$p_c = \frac{\Delta P}{1 - P(e|\neg c)}$$

Idea: ΔP directly is not a fair measure of p_c

- since there is a background rate $P(e|\neg c)$
- ΔP should be relativized to the remainder
  - those cases that would have been $\neg e$ but for c

"Power PC" Theory

Definition (Causal Power)

For positive ΔP (generative cause), the power of c to bring about e:

$$p_c = \frac{\Delta P}{1 - P(e|\neg c)}$$

Idea: ΔP directly is not a fair measure of p_c

- since there is a background rate $P(e|\neg c)$
- ΔP should be relativized to the remainder
  - those cases that would have been $\neg e$ but for c

"Power PC" Theory

Definition (Causal Power)

For positive ΔP (generative cause), the power of c to bring about e:

$$p_c = \frac{\Delta P}{1 - P(e|\neg c)}$$

Idea: ΔP directly is not a fair measure of p_c

- since there is a background rate $P(e|\neg c)$
- ΔP should be relativized to the remainder
  - those cases that would have been $\neg e$ but for c

PC Theory

Definition (Preventive Causal Power)

For negative ΔP (preventive cause), the power of c to stop e:

$$\overline{p_c} = \frac{-\Delta P}{P(e|\neg c)}$$

Symmetrically

- there is a background rate of failure to reach e, $P(\neg e|\neg c) = 1 - P(e|\neg c)$
- so $-\Delta P$ should be measured relative to the remainder

PC Theory

Definition (Preventive Causal Power)

For negative ΔP (preventive cause), the power of c to stop e:

$$\overline{p_c} = \frac{-\Delta P}{P(e|\neg c)}$$

Symmetrically

- there is a background rate of failure to reach e, $P(\neg e|\neg c) = 1 - P(e|\neg c)$
- so $-\Delta P$ should be measured relative to the remainder

PC Theory

Definition (Preventive Causal Power)

For negative ΔP (preventive cause), the power of c to stop e:

$$\overline{p_c} = \frac{-\Delta P}{P(e|\neg c)}$$

Symmetrically

- there is a background rate of failure to reach e, $P(\neg e|\neg c) = 1 - P(e|\neg c)$
- so $-\Delta P$ should be measured relative to the remainder

PC Theory

Definition (Preventive Causal Power)

For negative ΔP (preventive cause), the power of c to stop e:

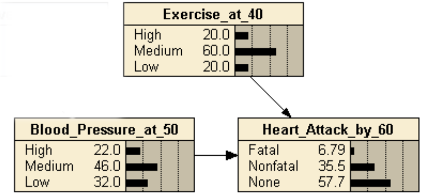
$$\overline{p_c} = \frac{-\Delta P}{P(e|\neg c)}$$

Symmetrically

- there is a background rate of failure to reach e, $P(\neg e|\neg c) = 1 - P(e|\neg c)$
- so $-\Delta P$ should be measured relative to the remainder

PC Theory

Consider the noisy-OR approximation:



Reconstruct variables as binary; delete arc between X and BP; eliminate messy interaction btw X and BP

Then:

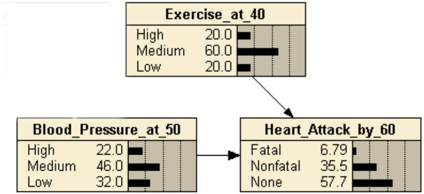
$$\Delta P = P(HA|BP) - P(HA|\neg BP) = 0.195$$

$$p_c = \frac{\Delta P}{1 - P(HA|\neg BP)} = 0.20$$

The prob that high BP will kill someone, given survival o/w relative to the model

PC Theory

Consider the noisy-OR approximation:



Reconstruct variables as binary; delete arc between X and BP; eliminate messy interaction btw X and BP

Then:

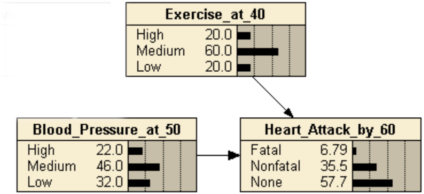
$$\Delta P = P(HA|BP) - P(HA|\neg BP) = 0.195$$

$$p_c = \frac{\Delta P}{1 - P(HA|\neg BP)} = 0.20$$

The prob that high BP will kill someone, given survival o/w relative to the model

PC Theory

Consider the noisy-OR approximation:



Reconstruct variables as binary; delete arc between X and BP; eliminate messy interaction btw X and BP

Then:

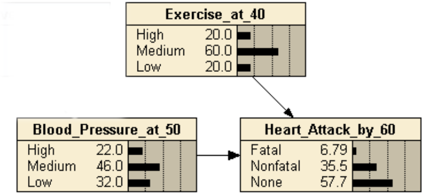
$$\Delta P = P(HA|BP) - P(HA|\neg BP) = 0.195$$

$$p_C = \frac{\Delta P}{1 - P(HA|\neg BP)} = 0.20$$

The prob that high BP will kill someone, given survival o/w relative to the model

PC Theory

Consider the noisy-OR approximation:



Reconstruct variables as binary; delete arc between X and BP; eliminate messy interaction btw X and BP

Then:

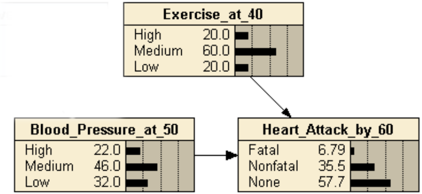
$$\Delta P = P(HA|BP) - P(HA|\neg BP) = 0.195$$

$$p_C = \frac{\Delta P}{1 - P(HA|\neg BP)} = 0.20$$

The prob that high BP will kill someone, given survival o/w relative to the model

PC Theory

Consider the noisy-OR approximation:



Reconstruct variables as binary; delete arc between X and BP; eliminate messy interaction btw X and BP

Then:

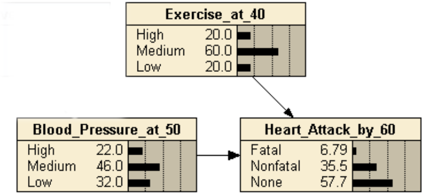
$$\Delta P = P(HA|BP) - P(HA|\neg BP) = 0.195$$

$$p_c = \frac{\Delta P}{1 - P(HA|\neg BP)} = 0.20$$

The prob that high BP will kill someone, given survival o/w relative to the model

PC Theory

Consider the noisy-OR approximation:



Reconstruct variables as binary; delete arc between X and BP; eliminate messy interaction btw X and BP

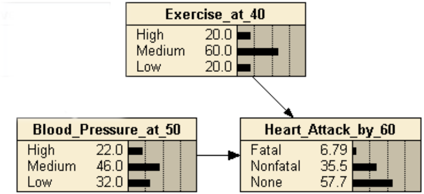
Then:

$$\Delta P = P(HA|BP) - P(HA|\neg BP) = 0.195$$
$$p_c = \frac{\Delta P}{1 - P(HA|\neg BP)} = 0.20$$

The prob that high BP will kill someone, given survival o/w relative to the model

PC Theory

Consider the noisy-OR approximation:



Reconstruct variables as binary; delete arc between X and BP; eliminate messy interaction btw X and BP

Then:

$$\Delta P = P(HA|BP) - P(HA|\neg BP) = 0.195$$

$$p_c = \frac{\Delta P}{1 - P(HA|\neg BP)} = 0.20$$

The prob that high BP will kill someone, given survival o/w relative to the model

PC Theory Limitations

- Binary variables only
- Power measured only between values of variables, not btw variables themselves (as with Wright) – we should like both
- Glymour (2001) shows that PC Theory is limited to "noisy-OR" relations in Bayesian networks
  $\Rightarrow$ Non-interactive and transitive causes only

  But we know, for example, causality is not transitive! (Finesteride, Hesslow's example)

PC Theory Limitations

- Binary variables only
- Power measured only between values of variables, not btw variables themselves (as with Wright) – we should like both
- Glymour (2001) shows that PC Theory is limited to "noisy-OR" relations in Bayesian networks
  $\Rightarrow$ Non-interactive and transitive causes only

But we know, for example, causality is not transitive! (Finesteride, Hesslow's example)

PC Theory Limitations

- Binary variables only
- Power measured only between values of variables, not btw variables themselves (as with Wright) – we should like both
- Glymour (2001) shows that PC Theory is limited to "noisy-OR" relations in Bayesian networks
  ⇒ Non-interactive and transitive causes only

  But we know, for example, causality is not transitive! (Finesteride, Hesslow's example)

PC Theory Limitations

- Binary variables only
- Power measured only between values of variables, not btw variables themselves (as with Wright) – we should like both
- Glymour (2001) shows that PC Theory is limited to "noisy-OR" relations in Bayesian networks

  $\Rightarrow$ Non-interactive and transitive causes only

  But we know, for example, causality is not transitive! (Finesteride, Hesslow's example)

PC Theory Limitations

- Binary variables only
- Power measured only between values of variables, not btw variables themselves (as with Wright) – we should like both
- Glymour (2001) shows that PC Theory is limited to "noisy-OR" relations in Bayesian networks
  $\Rightarrow$ Non-interactive and transitive causes only

But we know, for example, causality is not transitive! (Finesteride, Hesslow's example)

PC Theory Limitations

- Binary variables only
- Power measured only between values of variables, not btw variables themselves (as with Wright) – we should like both
- Glymour (2001) shows that PC Theory is limited to "noisy-OR" relations in Bayesian networks
  $\Rightarrow$ Non-interactive and transitive causes only

But we know, for example, causality is not transitive! (Finesteride, Hesslow's example)

Desiderata

Desideratum 3

Causal power theory should apply both to variables and their values.

Desideratum 4

Causal power theory should allow for non-transitive and interactive relations.

Desiderata

Desideratum 3

Causal power theory should apply both to variables and their values.

Desideratum 4

Causal power theory should allow for non-transitive and interactive relations.

Mutual Information

Consider

$$\frac{P(c,e)}{P(c)P(e)}$$

The deviation of the joint distribution from independence

$$\log \frac{P(c,e)}{P(c)P(e)}$$

Generalize:

$$MI(C,E) =_{df} \sum_{c \in C, e \in E} P(c,e) \log \frac{P(c,e)}{P(c)P(e)}$$

- Mutual Information
- Expected info about C given E, E given C

  $\Rightarrow$ *symmetric*
- The standard measure of prob dependence

Mutual Information

Consider

$$\frac{P(c,e)}{P(c)P(e)}$$

The deviation of the joint distribution from independence

$$\log \frac{P(c,e)}{P(c)P(e)}$$

Generalize:

$$MI(C,E) =_{df} \sum_{c \in C, e \in E} P(c,e) \log \frac{P(c,e)}{P(c)P(e)}$$

- Mutual Information
- Expected info about C given E, E given C

  $\Rightarrow$ *symmetric*

- The standard measure of prob dependence

Mutual Information

Consider

$$\frac{P(c, e)}{P(c)P(e)}$$

The deviation of the joint distribution from independence

$$\log \frac{P(c, e)}{P(c)P(e)}$$

Generalize:

$$MI(C, E) =_{df} \sum_{c \in C, e \in E} P(c, e) \log \frac{P(c, e)}{P(c)P(e)}$$

- Mutual Information
- Expected info about C given E, E given C

  $\Rightarrow$ *symmetric*

- The standard measure of prob dependence

Mutual Information

Consider

$$\frac{P(c,e)}{P(c)P(e)}$$

The deviation of the joint distribution from independence

$$\log \frac{P(c,e)}{P(c)P(e)}$$

Generalize:

$$MI(C,E) =_{df} \sum_{c \in C, e \in E} P(c,e) \log \frac{P(c,e)}{P(c)P(e)}$$

- Mutual Information
- Expected info about C given E, E given C

  $\Rightarrow$ *symmetric*

- The standard measure of prob dependence

Mutual Information

Consider

$$\frac{P(c,e)}{P(c)P(e)}$$

The deviation of the joint distribution from independence

$$\log \frac{P(c,e)}{P(c)P(e)}$$

Generalize:

$$MI(C,E) =_{df} \sum_{c \in C, e \in E} P(c,e) \log \frac{P(c,e)}{P(c)P(e)}$$

- Mutual Information
- Expected info about C given E, E given C

  $\Rightarrow$ *symmetric*

- The standard measure of prob dependence

Mutual Information

Consider

$$\frac{P(c,e)}{P(c)P(e)}$$

The deviation of the joint distribution from independence

$$\log \frac{P(c,e)}{P(c)P(e)}$$

Generalize:

$$MI(C,E) =_{df} \sum_{c \in C, e \in E} P(c,e) \log \frac{P(c,e)}{P(c)P(e)}$$

- Mutual Information
- Expected info about C given E, E given C

  $\Rightarrow$ *symmetric*

- The standard measure of prob dependence

Mutual Information

Consider

$$\frac{P(c,e)}{P(c)P(e)}$$

The deviation of the joint distribution from independence

$$\log \frac{P(c,e)}{P(c)P(e)}$$

Generalize:

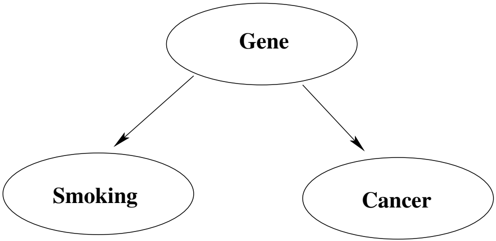
$$MI(C,E) =_{df} \sum_{c \in C, e \in E} P(c,e) \log \frac{P(c,e)}{P(c)P(e)}$$

- Mutual Information
- Expected info about C given E, E given C

  $\Rightarrow$ *symmetric*

- The standard measure of prob dependence

Fisher Model



MI travels up and down any Wrightian path, including back paths; causal influences clearly don't (outside of EPR problems).

Intervention

An intervention upon $V \in g$

- Alters the distribution over V in g
- From outside the system, outside g

Causal Bayesian networks are ideal for representing interventions, augmenting g by adding an intervention variable I, yielding the augmented g^*.

Intervention

An intervention upon $V \in g$

- Alters the distribution over V in g
- From outside the system, outside g

Causal Bayesian networks are ideal for representing interventions, augmenting g by adding an intervention variable I, yielding the augmented g^*.

Intervention

An intervention upon $V \in g$

- Alters the distribution over V in g
- From outside the system, outside g

Causal Bayesian networks are ideal for representing interventions, augmenting g by adding an intervention variable I, yielding the augmented g^*.

Intervention

An intervention upon $V \in g$

- Alters the distribution over V in g
- From outside the system, outside g

Causal Bayesian networks are ideal for representing interventions, augmenting g by adding an intervention variable I, yielding the augmented g^*.

Intervention

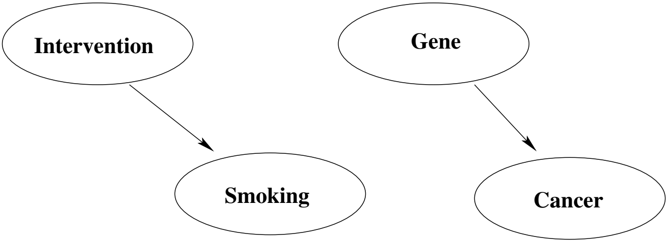
An intervention upon $V \in g$

- Alters the distribution over V in g
- From outside the system, outside g

Causal Bayesian networks are ideal for representing interventions, augmenting g by adding an intervention variable I, yielding the augmented g^*.

Fisher Model II

We shall use perfect (overwhelming) interventions to measure causal power



which was, of course, Fisher's idea!

Causal Information
Idea

Use MI, but *asymmetrically*

- By first intervening perfectly upon C and only then measuring MI

We get the asymmetrical dependence of E upon C, when C is set to a fixed distribution.

This automates Wright's power theory,

- via Bayesian net tools
- extending it automatically to *all* BNs

Remaining problem: which of the $\aleph_1$ distributions should we choose for C?

Causal Information

Idea

Use MI, but *asymmetrically*

- By first intervening perfectly upon C and only then measuring MI

We get the asymmetrical dependence of E upon C, when C is set to a fixed distribution.

This automates Wright's power theory,

- via Bayesian net tools
- extending it automatically to *all* BNs

Remaining problem: which of the $\aleph_1$ distributions should we choose for C?

Causal Information

Idea

Use MI, but *asymmetrically*

- By first intervening perfectly upon C and only then measuring MI

We get the asymmetrical dependence of E upon C, when C is set to a fixed distribution.

This automates Wright's power theory,

- via Bayesian net tools
- extending it automatically to *all* BNs

Remaining problem: which of the $\aleph_1$ distributions should we choose for C?

Causal Information

Idea

Use MI, but *asymmetrically*

- By first intervening perfectly upon C and only then measuring MI

We get the asymmetrical dependence of E upon C, when C is set to a fixed distribution.

This automates Wright's power theory,

- via Bayesian net tools
- extending it automatically to *all* BNs

Remaining problem: which of the $\aleph_1$ distributions should we choose for C?

Causal Information
Idea

Use MI, but *asymmetrically*

- By first intervening perfectly upon C and only then measuring MI

We get the asymmetrical dependence of E upon C, when C is set to a fixed distribution.

This automates Wright's power theory,

- via Bayesian net tools
- extending it automatically to *all* BNs

Remaining problem: which of the $\aleph_1$ distributions should we choose for C?

Causal Information

Idea

Use MI, but *asymmetrically*

- By first intervening perfectly upon C and only then measuring MI

We get the asymmetrical dependence of E upon C, when C is set to a fixed distribution.

This automates Wright's power theory,

- via Bayesian net tools
- extending it automatically to *all* BNs

Remaining problem: which of the $\aleph_1$ distributions should we choose for C?

Causal Information

Idea

Use MI, but *asymmetrically*

- By first intervening perfectly upon *C* and only then measuring MI

We get the asymmetrical dependence of *E* upon *C*, when *C* is set to a fixed distribution.

This automates Wright's power theory,

- via Bayesian net tools
- extending it automatically to *all* BNs

Remaining problem: which of the $\aleph_1$ *distributions should we choose for C?*

Causal Information

Definition (Causal information (CI))

Causal information between a cause C and an effect E in the causal model g is

$$CI(C,E) = \sum_{c \in C, e \in E} p(c)p(e|c) \log \frac{p(e|c)}{p(e)}$$

between the two variables in the augmented model g^*.

- This is precisely MI between C and E in the augmented model g^*.

Causal Information

Definition (Causal information (CI))

Causal information between a cause C and an effect E in the causal model g is

$$CI(C, E) = \sum_{c \in C, e \in E} p(c)p(e|c) \log \frac{p(e|c)}{p(e)}$$

between the two variables in the augmented model g^*.

- This is precisely MI between C and E in the augmented model g^*.

Many-Flavored Causal Information

Definition (CI various)

1. $CI(C, e) = \sum_{c \in C} p(c) p(e|c) \log \frac{p(e|c)}{p(e)}$
2. $CI(c, E) = \sum_{e \in E} p(e|c) \log \frac{p(e|c)}{p(e)}$
3. **Causal Power:** $CI(c, e) = p(e|c) \log \frac{p(e|c)}{p(e)}$

always measured in the augmented g^*.

1. Causal influence of C on $E = e$
2. Causal influence of $C = c$ on E
3. Causal power of c to bring about e

Many-Flavored Causal Information

Definition (CI various)

1. $CI(C, e) = \sum_{c \in C} p(c) p(e|c) \log \frac{p(e|c)}{p(e)}$
2. $CI(c, E) = \sum_{e \in E} p(e|c) \log \frac{p(e|c)}{p(e)}$
3. **Causal Power:** $CI(c, e) = p(e|c) \log \frac{p(e|c)}{p(e)}$

always measured in the augmented g^*.

1. Causal influence of C on $E = e$
2. Causal influence of $C = c$ on E
3. Causal power of c to bring about e

Many-Flavored Causal Information

Definition (CI various)

1. $CI(C, e) = \sum_{c \in C} p(c) p(e|c) \log \frac{p(e|c)}{p(e)}$
2. $CI(c, E) = \sum_{e \in E} p(e|c) \log \frac{p(e|c)}{p(e)}$
3. **Causal Power:** $CI(c, e) = p(e|c) \log \frac{p(e|c)}{p(e)}$

always measured in the augmented g^*.

1. Causal influence of C on $E = e$
2. Causal influence of $C = c$ on E
3. Causal power of c to bring about e

Many-Flavored Causal Information

Definition (CI various)

1. $CI(C, e) = \sum_{c \in C} p(c)p(e|c) \log \frac{p(e|c)}{p(e)}$
2. $CI(c, E) = \sum_{e \in E} p(e|c) \log \frac{p(e|c)}{p(e)}$
3. **Causal Power:** $CI(c, e) = p(e|c) \log \frac{p(e|c)}{p(e)}$

always measured in the augmented g^*.

1. Causal influence of C on $E = e$
2. Causal influence of $C = c$ on E
3. Causal power of c to bring about e

Many-Flavored Causal Information

Definition (CI various)

1. $CI(C, e) = \sum_{c \in C} p(c)p(e|c) \log \frac{p(e|c)}{p(e)}$
2. $CI(c, E) = \sum_{e \in E} p(e|c) \log \frac{p(e|c)}{p(e)}$
3. **Causal Power:** $CI(c, e) = p(e|c) \log \frac{p(e|c)}{p(e)}$

always measured in the augmented g^*.

1. Causal influence of C on $E = e$
2. Causal influence of $C = c$ on E
3. Causal power of c to bring about e

Many-Flavored Causal Information

Definition (CI various)

1. $CI(C, e) = \sum_{c \in C} p(c)p(e|c) \log \frac{p(e|c)}{p(e)}$
2. $CI(c, E) = \sum_{e \in E} p(e|c) \log \frac{p(e|c)}{p(e)}$
3. **Causal Power:** $CI(c, e) = p(e|c) \log \frac{p(e|c)}{p(e)}$

always measured in the augmented g^*.

1. Causal influence of C on $E = e$
2. Causal influence of $C = c$ on E
3. Causal power of c to bring about e

Example Questions

Questions related to the various CI measures:

C for *E*:

How much do heart attack outcomes depend upon BP?

C for *e*:

How many heart attack deaths are due to BP?

c for *E*:

How would heat attack outcomes vary given lowered BP?

c for *e*:

How many lives would be saved by interventions to lower BP?

Example Questions

Questions related to the various CI measures:

C for *E*:
: How much do heart attack outcomes depend upon BP?

C for *e*:
: How many heart attack deaths are due to BP?

c for *E*:
: How would heat attack outcomes vary given lowered BP?

c for *e*:
: How many lives would be saved by interventions to lower BP?

Example Questions

Questions related to the various CI measures:

C for *E*:

How much do heart attack outcomes depend upon BP?

C for *e*:

How many heart attack deaths are due to BP?

c for *E*:

How would heat attack outcomes vary given lowered BP?

c for *e*:

How many lives would be saved by interventions to lower BP?

Example Questions

Questions related to the various CI measures:

C for *E*:
: How much do heart attack outcomes depend upon BP?

C for *e*:
: How many heart attack deaths are due to BP?

c for *E*:
: How would heat attack outcomes vary given lowered BP?

c for *e*:
: How many lives would be saved by interventions to lower BP?

Example Questions

Questions related to the various CI measures:

C for *E*:
: How much do heart attack outcomes depend upon BP?

C for *e*:
: How many heart attack deaths are due to BP?

c for *E*:
: How would heat attack outcomes vary given lowered BP?

c for *e*:
: How many lives would be saved by interventions to lower BP?

$\aleph_1$ Intervention Distributions

- Original $p(C)$

  Given the actual distribution of BP, how does BP influence heart attack? (E.g., Swedes vs non-Swedes)

- Uniform $p(C)$

  As in randomized experimental designs

- Maximizing $p(C)$

  What is the greatest possible influence of C for E? How strongly could lowering BP impact on heart attack outcomes?

The latter two provide a kind of standard baseline for comparing causal powers.

$\aleph_1$ Intervention Distributions

- Original $p(C)$

  Given the actual distribution of BP, how does BP influence heart attack? (E.g., Swedes vs non-Swedes)

- Uniform $p(C)$

  As in randomized experimental designs

- Maximizing $p(C)$

  What is the greatest possible influence of C for E? How strongly could lowering BP impact on heart attack outcomes?

The latter two provide a kind of standard baseline for comparing causal powers.

$\aleph_1$ Intervention Distributions

- Original $p(C)$

  Given the actual distribution of BP, how does BP influence heart attack? (E.g., Swedes vs non-Swedes)

- Uniform $p(C)$

  As in randomized experimental designs

- Maximizing $p(C)$

  What is the greatest possible influence of C for E? How strongly could lowering BP impact on heart attack outcomes?

The latter two provide a kind of standard baseline for comparing causal powers.

$\aleph_1$ Intervention Distributions

- Original $p(C)$

  Given the actual distribution of BP, how does BP influence heart attack? (E.g., Swedes vs non-Swedes)

- Uniform $p(C)$

  As in randomized experimental designs

- Maximizing $p(C)$

  What is the greatest possible influence of C for E? How strongly could lowering BP impact on heart attack outcomes?

The latter two provide a kind of standard baseline for comparing causal powers.

$\aleph_1$ Intervention Distributions

- Original $p(C)$

  Given the actual distribution of BP, how does BP influence heart attack? (E.g., Swedes vs non-Swedes)

- Uniform $p(C)$

  As in randomized experimental designs

- Maximizing $p(C)$

  What is the greatest possible influence of C for E? How strongly could lowering BP impact on heart attack outcomes?

The latter two provide a kind of standard baseline for comparing causal powers.

$\aleph_1$ Intervention Distributions

- Original $p(C)$

  Given the actual distribution of BP, how does BP influence heart attack? (E.g., Swedes vs non-Swedes)

- Uniform $p(C)$

  As in randomized experimental designs

- Maximizing $p(C)$

  What is the greatest possible influence of C for E? How strongly could lowering BP impact on heart attack outcomes?

The latter two provide a kind of standard baseline for comparing causal powers.

$\aleph_1$ Intervention Distributions

- Original $p(C)$

  Given the actual distribution of BP, how does BP influence heart attack? (E.g., Swedes vs non-Swedes)

- Uniform $p(C)$

  As in randomized experimental designs

- Maximizing $p(C)$

  What is the greatest possible influence of C for E? How strongly could lowering BP impact on heart attack outcomes?

The latter two provide a kind of standard baseline for comparing causal powers.

$\aleph_1$ Intervention Distributions

- Original $p(C)$

  Given the actual distribution of BP, how does BP influence heart attack? (E.g., Swedes vs non-Swedes)

- Uniform $p(C)$

  As in randomized experimental designs

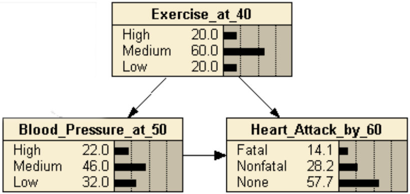
- Maximizing $p(C)$

  What is the greatest possible influence of C for E? How strongly could lowering BP impact on heart attack outcomes?

The latter two provide a kind of standard baseline for comparing causal powers.

Heart Attack Example

MI vs CI



$$
\begin{aligned}
MI(BP, HA) &= \sum_{c \in C, e \in E} P(c, e) \log \frac{P(c, e)}{P(c)P(e)} \\
&= 0.28 \\
CI(BP, HA) &= 0.13
\end{aligned}
$$

- The difference is due to the interventional elimination of the backpath through X

Heart Attack Example

CI causal power

Two CI causal powers for fatal heart attack:

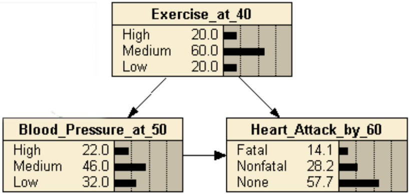
$$CI(c, e) = p(e|c) \log \frac{p(e|c)}{p(e)}$$

- CI(high BP, fatal HA) $= 0.23 \log \frac{0.23}{0.0679} = 0.405$
- CI(low BP, fatal HA) $= 0.052 \log \frac{0.052}{0.0679} = -0.02$

Heart Attack Example

Cheng

What happens to Cheng's PC Theory when we apply it to the original model?



The reintroduction of backpath and interaction

- $p_c = \Delta P/[1 - P(HA|\neg BP)] = 0.16$
  – a decline of 20%

This shows significant errors in attempting to apply PC Theory.

PC Theories

	Variables	Structures	Causality
Wright	Linear	Open	Transitive
Cheng/Glymour	Binary	Noisy-OR Isolated Causes	Transitive
CI	Various	Open	Various (Interactions, thresholds)

Desiderata

1. Causal power theory should apply to **any** kind of causal Bayesian network – linear, binomial, multinomial.
2. Causal power theory should generalize Wright's power theory.
3. Causal power theory should apply both to variables and their values.
4. Causal power theory should allow for non-transitive and interactive relations.

CI Summary

- CI reports the expected code length needed to report the value of E given the value of C in g^*
  - This can be converted back into the language of probabilities
- CI satisfies all of our desiderata, unlike any known alternative
- CI can summarize the explanatory import of hypothetical causes, making causal BNs intelligible
- CI can be applied to test theories of causal attribution
- CI can be applied to test theories of token causation

CI Summary

- CI reports the expected code length needed to report the value of E given the value of C in g^*
  - This can be converted back into the language of probabilities
- CI satisfies all of our desiderata, unlike any known alternative
- CI can summarize the explanatory import of hypothetical causes, making causal BNs intelligible
- CI can be applied to test theories of causal attribution
- CI can be applied to test theories of token causation

CI Summary

- CI reports the expected code length needed to report the value of E given the value of C in g^*
  - This can be converted back into the language of probabilities
- CI satisfies all of our desiderata, unlike any known alternative
- CI can summarize the explanatory import of hypothetical causes, making causal BNs intelligible
- CI can be applied to test theories of causal attribution
- CI can be applied to test theories of token causation

CI Summary

- CI reports the expected code length needed to report the value of E given the value of C in g^*
  - This can be converted back into the language of probabilities
- CI satisfies all of our desiderata, unlike any known alternative
- CI can summarize the explanatory import of hypothetical causes, making causal BNs intelligible
- CI can be applied to test theories of causal attribution
- CI can be applied to test theories of token causation

CI Summary

- CI reports the expected code length needed to report the value of E given the value of C in g^*
  - This can be converted back into the language of probabilities
- CI satisfies all of our desiderata, unlike any known alternative
- CI can summarize the explanatory import of hypothetical causes, making causal BNs intelligible
- CI can be applied to test theories of causal attribution
- CI can be applied to test theories of token causation

CI Summary

- CI reports the expected code length needed to report the value of E given the value of C in g^*
  - This can be converted back into the language of probabilities
- CI satisfies all of our desiderata, unlike any known alternative
- CI can summarize the explanatory import of hypothetical causes, making causal BNs intelligible
- CI can be applied to test theories of causal attribution
- CI can be applied to test theories of token causation

CI Summary

- CI reports the expected code length needed to report the value of E given the value of C in g^*
  - This can be converted back into the language of probabilities
- CI satisfies all of our desiderata, unlike any known alternative
- CI can summarize the explanatory import of hypothetical causes, making causal BNs intelligible
- CI can be applied to test theories of causal attribution
- CI can be applied to test theories of token causation