CAUSALITY: COUNTERFACTUAL AND HYPOTHETICAL

Philip Dawid
University College London

Some Distinctions

• Token	• ***Type***
• Retrospective	• ***Prospective***
• Cause of effect	• ***Effect of cause***
• Counterfactual	• ***Hypothetical***
• Deterministic	• ***Stochastic***
• Value	• ***Distribution***
• Observation	• ***Intervention***

Some quotes

- "Today, the counterfactual, *or potential outcome*, model of causality has become more or less standard in epidemiology" (Höfler 2005a)
- "In the past two decades, statisticians and econometricians have adopted a common conceptual framework for thinking about the estimation of causal effects—the counterfactual account of causality" (Winship & Morgan 1999)

Some quotes

- "Counterfactuals are a hot topic in economics today… I shall argue that on the whole this is a mistake" (Cartwright 2006)
- For making inference about the likely effects of applied causes, counterfactual arguments are unnecessary and potentially misleading (Dawid 2000)

Counterfactual or Hypothetical?

- If I had taken aspirin half an hour ago, would my headache would have gone by now?
- If I take aspirin now, will my headache be gone within half an hour?

A SIMPLE PROBLEM

- Randomised experiment
- Binary (0/1) treatment decision variable T
- Response variable Y

Define/measure/estimate *"the effect of treatment"*

Statistical Model *(Fisher)*

- Specify/estimate *conditional distributions*
  P_t for Y *given* $T = t$ $(t = 0, 1)$

$$\left(\text{e.g., } \begin{array}{l} Y \mid T = 0 \sim N(\mu_0, 1) \\ Y \mid T = 1 \sim N(\mu_1, 1) \end{array}\right)$$

- Measure effect of treatment by appropriate comparison of distributions P_0 and P_1
  – *e.g.* difference of expected responses

$$\delta = \mu_1 - \mu_0$$

Counterfactual use??

- I took aspirin ($T = 1$) and my headache lasted 30 minutes ($Y = 30$)
  – "indicative" conditioning on $T = 1$
- How long would it have lasted if I had not taken aspirin ($T = 0$) ?
  – "subjunctive" conditioning on $T = 0$
- Can not do both conditionings at once!

- ***NEED NEW FRAMEWORK??***

Potential Response Model

- Split Y in two:
  Y_0: response to $T = 0$ • *Hypothetical?*
  Y_1: response to $T = 1$ • *Counterfactual?* • *Complementary!*
- Consider (for any unit) the *pair* $\mathbf{Y} = (Y_0, Y_1)$
  – *with simultaneous existence and joint distribution*
- Treatment "uncovers" pre-existing response:
  $Y = Y_T$ (determined)
- Unit-level (individual) [random] causal effect
  $Y_1 - Y_0$
  – *necessarily unobservable!*

CONNEXIONS

- A PR model determines a statistical model
  –just extract marginal distributions:

$$Y \mid (T = t) \approx Y_t \quad (t = 0, 1)$$

- But distinct PR models can determine the same statistical model
  – the dependence between Y_1 and Y_0 in PR makes no difference to SM
  – but does affect counterfactual analyses

Potential Responses: Problems?

- *PR* model:

$$\begin{cases} Y_t \sim N(\mu_t, 1) \quad (t = 0, 1) \\ \text{corr}(Y_0, Y_1) = \rho \end{cases}$$

- Corresponding *statistical* model:

$$Y \mid (T = t) \sim N(\mu_t, 1)$$

NB: ρ does not enter! – *can not estimate* ρ
– does this matter??

Potential Responses: Problems?

- Under PR model:

$$\text{var}(Y_1 - Y_0) = 2(1 - \rho)$$

➢ *We can not identify population variation in individual causal effect*

$$E(Y_1 - Y_0 \mid Y_1 = y_1) = (1 - \rho)\, y_1 + (\rho\mu_1 - \mu_0)$$

➢ *We can not identify the (counterfactual) individual causal effect, having observed the response to actual treatment*

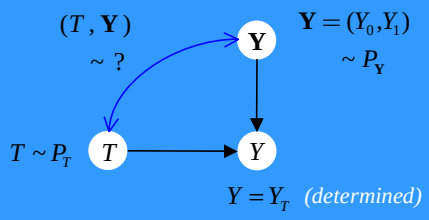
OBSERVATIONAL STUDY

- Treatment decision taken may be associated with patient's state of health
- What assumptions are required to make causal inferences?
- When/how can such assumptions be justified?

Some more quotes

- "How is it possible to draw a distinction between causal relations and non-causal associations? In order to meet this concern a further element must be added to the definition—a counterfactual" (Parascandola & Weed 2001)
- "Probabilistic causal inference (of which Dawid is an advocate) in observational studies would inevitably require counterfactuals" (Höfler 2005b)

Potential Response Model



"Ignorable treatment assignment" $T \perp\!\!\!\perp \mathbf{Y}$

– *treatment independent of potential responses*

PR Model: Some Comments

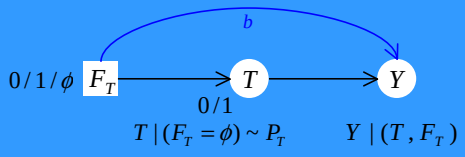
- Value of $\mathbf{Y} = (Y_0, Y_1)$ on any unit supposed the same for both observational and experimental cases (as well as for both choices of T)
- How are we to judge independence of T from $\mathbf{Y}$?

Decision Theoretic Model

- Introduce explicit "treatment regime indicator" variable F_T
- Values:
  - $F_T = 0$: Assign treatment 0 $(\Rightarrow T = 0)$
  - $F_T = 1$: Assign treatment 1 $(\Rightarrow T = 1)$
  - $F_T = \phi$: Just observe
- "Ignorable treatment assignment":
  - identity of observational and experimental *distributions* for $Y \mid T$:

$$Y \perp\!\!\!\perp F_T \mid T$$

Influence Diagram



Absence of arrow b expresses $Y \perp\!\!\!\perp F_T \mid T$

$e.g.\ \ Y \mid (T = t, F_T = \phi) \sim N(\mu_t, 1)$

Causal Model

- Simply a *more ambitious* non-causal model, expressing the invariance of certain modular structures across different regimes (e.g. interventional / observational)
- For PR model:
  - invariant *values* of potential responses
    - implicit, inescapable
- For DT model
  - invariant *conditional distributions* of response
    - explicit, inessential

Advantages

- No potential responses
- Stochastic, not deterministic, relationships
- Simple, explicit, in principle *testable* assumptions
  - –what constitutes a valid basis for such assumptions when experimentation is not pragmatically possible?

DYNAMIC TREATMENT REGIMES

observe L_0 *act* A_0 *observe* L_1 *act* A_1 *observe* Y

time

Consider (deterministic) *treatment regime* g:

$$\begin{array}{rcrcl} l_0 & \stackrel{g}{\rightarrow} & a_0 & = & g(l_0) \\ (l_0, l_1) & \stackrel{g}{\rightarrow} & a_1 & = & g(l_0, l_1) \end{array}$$

Distribution of Y under g?

PR Approach

For each regime g have potential intermediate and response variables:

$$(L_{0g},\ L_{1g},\ Y_g)$$

"Consistency":

$$\begin{array}{lcl} & & L_0 = L_{0g} \\ A_0 = g(L_0) & \Rightarrow & L_1 = L_{1g} \\ A_0 = g(L_0),\ A_1 = g(L_1) & \Rightarrow & Y = Y_g \end{array}$$

– if history to date consistent with g, so is next observable

"Sequential Ignorability"

For each regime g and all possible (l_0, l_1):

$$\begin{array}{lcll} A_0 & \perp\!\!\!\perp & (L_{1g}, Y_g) & | \quad L_0 = l_0 \\ A_1 & \perp\!\!\!\perp & Y_g & | \quad L_0 = l_0, A_0 = g(l_0), L_1 = l_1 \end{array}$$

– actions independent of future potential observables, *given current (consistent) history*

– when valid, allows estimation of distribution of any Y_g from observational data

– by "*G-computation*" formula

DT Approach

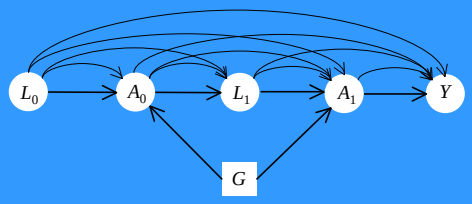
- Introduce explicit "dynamic regime indicator" variable G
- Values include various (now possibly randomised) experimental regimes g *and observational regime* ϕ

Sequential Ignorability

Conditional distribution of each observable, given history, is the same under all regimes:

$$
\begin{aligned}
L_0 &\perp\!\!\!\perp G \\
L_1 &\perp\!\!\!\perp G \mid L_0, A_0 \\
Y &\perp\!\!\!\perp G \mid L_0, A_0, L_1, A_1
\end{aligned}
$$

Influence Diagram



Advantages

- *No unobservable potential responses*
- *Stochastic, not deterministic, relationships*
- *Simple, explicit,* testable *assumptions*
- No consistency requirement
- Permits randomised treatment regimes
- Yields simple argument for *G*-computation
- Approach readily extends to *e.g.* "structural nested models"

Conclusion: Philosophical Comparisons

- Ontology
  - "tablets of stone" / open
- Expressiveness
  - counterfactuals / contexts
- Epistemology
  - unlearnable / testable
- Pragmatics
  - cumbersome / fit for purpose
- Agency
  - implicit / avoidable

Further Reading

- A. P. Dawid (2000). Causal inference without counterfactuals (with Discussion). *J. Amer. Statist. Ass.* **95**, 407– 448.
- A. P. Dawid (2002). Influence diagrams for causal modelling and inference. *Intern. Statist. Rev.* **70**, 161– 189. Corrigenda, ibid., 437.
- A. P. Dawid (2003). Causal inference using influence diagrams: The problem of partial compliance (with Discussion). In *Highly Structured Stochastic Systems*, edited by Peter J. Green, Nils L. Hjort and Sylvia Richardson. Oxford University Press, 45 – 81.
- A. P. Dawid (2004). Probability, causality and the empirical world: A Bayes–de Finetti–Popper–Borel synthesis. *Statistical Science* **19**, 44 – 57.
- A. P. Dawid and V. Didelez (2005). Identifying the consequences of dynamic treatment strategies. Research Report 262, Department of Statistical Science, University College London. http://www.ucl.ac.uk/Stats/research/Resrprts/abs05.html#262
- V. Didelez, A. P. Dawid and S. Geneletti . Direct and indirect effects of sequential treatments. Research Report 265, Department of Statistical Science, University College London. http://www.ucl.ac.uk/Stats/research/Resrprts/abs06.html#265

Additional References

- N. Cartwright. (2006). Counterfactuals in economics: A commentary. In *Explanation and Causation: Topics in Contemporary Philosophy*, O'Rourke *et al.* (eds.) Vol. 4. MIT Press (to appear).
- M. Höfler (2005a). The Bradford Hill considerations on causality: A counterfactual perspective. *Emerging Themes in .Epidemiology* 2:11. doi:10.1186/1742-7622-2-11. http://www.ete-online.com/content/2/1/11.
- M Höfler (2005b). Causal inference based on counterfactuals. *BMC Medical Research Methodology* 5:28. doi:10.1186/1471-2288-5-28. http://www.biomedcentral.com/1471-2288/5/28.
- M Parascandola and D L Weed (2001). Causation in epidemiology. *J. Epidemiol. Community Health* **55**, 905-912. doi:10.1136/jech.55.12.905
- B. Pascal (1669). Pensées sur la réligion, et sur quelques autres sujets. Paris: Guillaume Desprez (Edition Garnier Frères, 1964).
- G. Shafer (1996). *The Art of Causal Conjecture.* Cambridge, Mass.: MIT Press.
- C. Winship and S. L. Morgan (1999). The estimation of causal effects from observational data. *Annual Reviews of Sociology* **25**, 659–706.