# Causal Inference with Graphical Models

Jose M. Peña
STIMA, IDA, LiU

Lecture 3: Actions, Plans and Direct Effects

# Contents

# Literature

- Main sources
  - Pearl, J. *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press, 2009. Chapter 4.
  - Pearl, J. Direct and Indirect Effects. *UAI 17*, 411-420, 2001.
- Additional sources
  - Pearl, J. and Robins, J. Probabilistic Evaluation of Sequential Plans from Causal Models with Hidden Variables. *UAI 11*, 444-453, 1995.
  - Pearl, J., Glymour, M. and Jewell, N. P. *Causal Inference in Statistics: A Primer*. Wiley, 2016. Chapters 3.7 and 4.5.2.
  - Pearl, J. *Causality: Models, Reasoning, and Inference* (1st ed.). Cambridge University Press, 2000. Chapter 4.

# Conditional and Stochastic Actions

- Plan evaluation is relevant in fields such as health management, economic policy making or robot motion planning. Say $Z_k$ represents the process state at time $t_k$ (e.g., temperature), $X_k$ stands for some control variables (e.g., chemicals), and $Y$ is the process outcome (e.g., product quality).

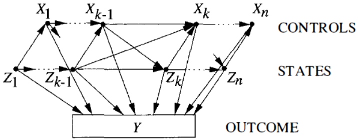



**Figure 3.3** Dynamic causal diagram illustrating typical dependencies among the control variables $X_1, \ldots, X_n$, the state variables $Z_1, \ldots, Z_n$, and the outcome variable $Y$ of a sequential process.

- Factorization:

$$p(y, z_1, \ldots, z_n, x_1, \ldots, x_n)$$
$$= p(y|z_1, \ldots, z_n, x_1, \ldots, x_n) \prod_k p(x_k|x_{k-1}, z_k, z_{k-1}) \prod_k p(z_k|z_{k-1}, x_{k-1}).$$

- Plan evaluation, i.e. the plan is a set of actions $\hat{x}_k$ ($\equiv do(x_k)$):

$$p^*(y) = p(y|\hat{x}_1, \ldots, \hat{x}_n) = \sum_{z_1, \ldots, z_n} p(y|z_1, \ldots, z_n, x_1, \ldots, x_n) \prod_k p(z_k|z_{k-1}, x_{k-1}).$$

- Plan evaluation is a key ingredient of decision theory, which instructs a rational agent to perform the action $x$ with maximum expected utility:

$$EU(x) = \sum_y p(y|\hat{x}) u(y).$$

## Conditional and Stochastic Actions

- A plan that consists of actions $do(x_k)$ is called **unconditional**. A plan that consists of actions $do(X_k = g(x_{k-1}, z_k, z_{k-1}))$ is called **conditional**. A plan that consists of actions chosen according to a probability distribution $p^*(x_k|x_{k-1}, z_k, z_{k-1})$ is called **stochastic**.
- Note that unconditional $\subseteq$ conditional $\subseteq$ stochastic plans.
- Factorization:

$$p(y, z_1, \ldots, z_n, x_1, \ldots, x_n)$$
$$= p(y|z_1, \ldots, z_n, x_1, \ldots, x_n) \prod_k p(x_k|x_{k-1}, z_k, z_{k-1}) \prod_k p(z_k|z_{k-1}, x_{k-1}).$$

- Stochastic plan evaluation:

$$p*(y) = \sum_{z_1,\ldots,z_n} p^*(y, z_1, \ldots, z_n, x_1, \ldots, x_n)$$
$$= \sum_{z_1,\ldots,z_n} p(y|z_1, \ldots, z_n, x_1, \ldots, x_n) \prod_k p^*(x_k|x_{k-1}, z_k, z_{k-1}) \prod_k p(z_k|z_{k-1}, x_{k-1}).$$

- Conditional plan evaluation:

$$\begin{aligned} p*(y) &= \sum_{z_1,\ldots,z_n} p^*(y, z_1, \ldots, z_n, x_1, \ldots, x_n) \\ &= \sum_{z_1,\ldots,z_n} p(y|z_1, \ldots, z_n, g_1, \ldots, g_n) \prod_k p(z_k|z_{k-1}, g_{k-1}) \end{aligned}$$

  where $g_1 = g(z_1)$ and $g_k = g(g_{k-1}, z_k, z_{k-1})$.

# Conditional and Stochastic Actions

- More generally, we have now three types of actions or interventions at our disposal: Unconditional, conditional and stochastic.
- Causal effect of a conditional action $do(X = g(z))$:

$$\begin{aligned} p^*(y) = p(y|do(X = g(z))) &= \sum_z p(y|do(X = g(z)), z) p(z|do(X = g(z))) \\ &= \sum_z p(y|\hat{x}, z)|_{x=g(z)} p(z) \end{aligned}$$

  because $Z$ is assumed to be measured prior to taking the action $X$ and, thus, $Z$ cannot be a descendant of $X$.
- Causal effect of a stochastic action $p^*(x|z)$:

$$p^*(y) = \sum_x \sum_z p(y|\hat{x}, z) p^*(x|z) p(z).$$

- Therefore, identifiability for conditional and stochastic actions is stricter than identifiability for unconditional actions, because conditioning on $z$ may create dependencies that prevent identifying $p(y|\hat{x}, z)$.

# Sequential Back-Door Criterion

- Consider a causal structure $G$ over $\{X, Z, U, Y\}$ where
  - $X = \{X_1, \ldots, X_n\}$ is a set of control variables, and
  - $Z$ is a set of observed variables, and
  - $U$ is a set of latent variables, and
  - $Y$ is a **single** outcome variable.
- Assume that $X_k$ is a non-descendant of $X_{k+1}$, and $Y$ is a descendant of $X_n$ (maybe of others $X_k$ too). Let $N_k$ be the non-descendants of $\{X_k, \ldots, X_n\}$.
- The plan $p(y|\hat{x}_1, \ldots, \hat{x}_n)$ is identifiable if
  1. there is a set $Z_k \subseteq N_k$ for all $1 \leq k \leq n$ st
  2. $Y \perp_{G_{\underline{X}_k, \overline{X}_{k+1}, \ldots, \overline{X}_n}} X_k | X_1, \ldots, X_{k-1}, Z_1, \ldots, Z_k$.

  Moreover, the plan evaluation is given by

$$p(y|\hat{x}_1, \ldots, \hat{x}_n) = \sum_{z_1, \ldots, z_n} p(y|z_1, \ldots, z_n, x_1, \ldots, x_n) \prod_k p(z_k|z_1, \ldots, z_{k-1}, x_1, \ldots, x_{k-1}).$$

- Note that condition 2 is equivalent to the back-door criterion on $G$ after deleting the (bi)directed edges into future actions.

# Sequential Back-Door Criterion

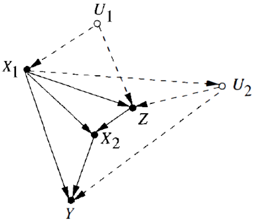



**Figure 4.4** The problem of evaluating the effect of the plan $(do(x_1), do(x_2))$ on $Y$, from nonexperimental data taken on $X_1$, $Z$, $X_2$, and $Y$.

To motivate the discussion, consider an example discussed in Robins (1993, apx. 2), as depicted in Figure 4.4. The variables $X_1$ and $X_2$ stand for treatments that physicians prescribe to a patient at two different times, $Z$ represents observations that the second physician consults to determine $X_2$, and $Y$ represents the patient's survival. The hidden variables $U_1$ and $U_2$ represent, respectively, part of the patient's history and the patient's disposition to recover. A simple realization of such structure could be found among AIDS patients, where $Z$ represents episodes of PCP. This is a common opportunistic infection of AIDS patients that (as the diagram shows) does not have a direct effect on survival $Y$ because it can be treated effectively, but it is an indicator of the patient's underlying immune status ($U_2$), which can cause death. The terms $X_1$ and $X_2$ stand for bactrim, a drug that prevents PCP ($Z$) and may also prevent death by other mechanisms. Doctors used the patient's earlier PCP history ($U_1$) to prescribe $X_1$, but its value was not recorded for data analysis.

- The plan is identifiable with $Z_1 = \varnothing$ and $Z_2 = \{Z\}$ as

$$p(y|\hat{x}_1, \hat{x}_2) = \sum_z p(y|x_1, x_2, z)p(z|x_1).$$

# Sequential Back-Door Criterion

- Proof: Condition 1 implies that no node in $\{Z_1, \ldots, Z_k, X_1, \ldots, X_{k-1}\}$ is a descendant of $\{X_k, \ldots, X_n\}$ and, thus, rule 3 implies that

$$p(z_k | z_1, \ldots, z_{k-1}, x_1, \ldots, x_{k-1}, \hat{x}_k, \ldots, \hat{x}_n) = p(z_k | z_1, \ldots, z_{k-1}, x_1, \ldots, x_{k-1}).$$

Moreover, condition 2 and rule 2 imply that

$$p(y | z_1, \ldots, z_k, x_1, \ldots, x_{k-1}, \hat{x}_k, \ldots, \hat{x}_n) = p(y | z_1, \ldots, z_k, x_1, \ldots, x_k, \hat{x}_{k+1}, \ldots, \hat{x}_n).$$

Putting all together, we have that

$$\begin{aligned}
p(y | \hat{x}_1, \ldots, \hat{x}_n) &= \sum_{z_1} p(y | z_1, \hat{x}_1, \ldots, \hat{x}_n) p(z_1 | \hat{x}_1, \ldots, \hat{x}_n) \\
&= \sum_{z_1} p(y | z_1, x_1, \hat{x}_2, \ldots, \hat{x}_n) p(z_1) \\
&= \sum_{z_1, z_2} p(y | z_1, z_2, x_1, \hat{x}_2, \ldots, \hat{x}_n) p(z_1) p(z_2 | z_1, x_1, \hat{x}_2, \ldots, \hat{x}_n) \\
&= \sum_{z_1, z_2} p(y | z_1, z_2, x_1, x_2, \hat{x}_3, \ldots, \hat{x}_n) p(z_1) p(z_2 | z_1, x_1) \\
&= \ldots
\end{aligned}$$

# Sequential Back-Door Criterion

- Choosing $Z_k$: Exhaustive search. Alternatively, there exist sets $Z_k$ satisfying conditions 1 and 2 iff

$$Y \perp_{G_{\underline{X}_k, \overline{X}_{k+1}, \ldots, \overline{X}_n}} X_k | X_1, \ldots, X_{k-1}, W_1, \ldots, W_k$$

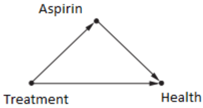
  where $W_k$ is the set of nodes that are non-descendants of $\{X_k, \ldots, X_n\}$ in $G$ and have either $Y$ or $X_k$ as descendant in $G_{\underline{X}_k, \overline{X}_{k+1}, \ldots, \overline{X}_n}$. Moreover,

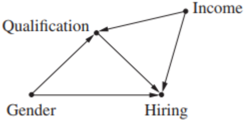
$$p(y|\hat{x}_1, \ldots, \hat{x}_n) = \sum_{w_1, \ldots, w_n} p(y|w_1, \ldots, w_n, x_1, \ldots, x_n) \prod_k p(w_k|w_1, \ldots, w_{k-1}, x_1, \ldots, x_{k-1}).$$

- Choosing the ordering of the actions $X$: Exhaustive search.
- The above can be extended to conditional plans.
- Note that *do*-calculus is an alternative to the above, albeit less intuitive.

## Direct and Indirect Effects

- The direct effect is the effect of $X$ on $Y$ that is not mediated by other variables in $G$, i.e. all the other variables are held **fixed**. Therefore, the direct effect is of interest only if $X \in Pa_Y$.
- **Total effect**: $p(y|do(x))$.
- **Direct effect**: $p(y|do(x), do(v \setminus \{x,y\}))$ where $V$ are the observed variables in $G$. Alternatively, $p(y|do(x), do(pa_Y \setminus \{x\}))$.




- The direct effect is important when evaluating the effectiveness of a treatment, when investigating possible race or sex discrimination, etc.
- In general, it is wrong to condition on $Pa_Y \setminus \{X\}$. For instance, $p(h|do(t), do(a)) = p(h|t,a)$ but $p(h|do(g), do(q)) \neq p(h|g,q)$.
- Since the direct effect is of interest only if $X \in Pa_Y$, the direct effect corresponds to a **plan** where some $Pa_Y$ are the control variables. E.g.,

$$p(h|do(g), do(q)) = \sum_i p(h|i,g,q)p(i|g)$$

  with $Z_1 = \varnothing$ and $Z_2 = \{I\}$. Moreover, $p(i|g) = p(i)$.
- The direct effect is not identifiable if $X \leftrightarrow Y$ is in $G$.

# Direct and Indirect Effects

- More exactly, the direct effect of $X$ on $Y$ is defined as the change in $Y$ that is induced by changing $X$ from the reference value $x^*$ to the value $x$, while $Pa_Y \setminus \{X\}$ are held fixed, i.e.

$$p(y|do(x), do(pa_Y \setminus \{x\})) - p(y|do(x^*), do(pa_Y \setminus \{x\})).$$

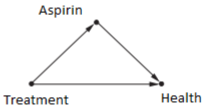
- Note that the direct effect may depend on the value $pa_Y \setminus \{x\}$. Therefore, we may report it for
  - a prescribed value, a.k.a. average **controlled** direct effect, or
  - the value under $do(x^*)$, a.k.a. average **natural** direct effect, i.e. the average improvement in health if we start the treatment but the patients are given as much aspirin as with no treatment, or the average increase in female hiring if females are trained to have the same qualifications as males.
- Formally,
  - $CDE(x, x^*, Y) = E[Y_{xz}] - E[Y_{x^*z}]$
  - $NDE(x, x^*, Y) = E[Y_{xZ_{x^*}}] - E[Y_{x^*}]$

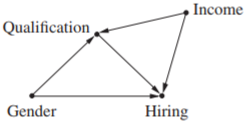
  where $Z = Pa_Y \setminus \{X\}$, and $Y_x$ denotes the value of $Y$ under regime $do(x)$.
- CDE is identifiable if the plan $Y_{xz}$ is identifiable.
- For causal structures **without latent variables**, NDE is identifiable as

$$NDE(x, x^*, Y) = \sum_s \sum_z (E[Y|x, z] - E[Y|x^*, z])p(z|x^*, s)p(s)$$

where $S$ are any variables satisfying the back-door criterion wrt $(X, Z)$.

# Direct and Indirect Effects




- The average **natural indirect** effect is the expected change in $Y$ when $X$ is held fixed at $x^*$ but $Z$ changes to whatever value it would have attained had $X$ be set to $x$, i.e.

$$NIE(x, x^*, Y) = E[Y_{x^* Z_x}] - E[Y_{x^*}].$$

- In the examples above, it is the average improvement in health if we stop the treatment but the patients are given as much aspirin as under the treatment, or the average increase in male hiring if males are trained to have the same qualifications as females.
- NIE is identifiable if the total effect and NDE are identifiable, because

$$\begin{aligned} TE(x, x^*, Y) = E[Y_x] - E[Y_{x^*}] &= NIE(x, x^*, Y) - NDE(x^*, x, Y) \\ &= NDE(x, x^*, Y) - NIE(x^*, x, Y). \end{aligned}$$

- However, in general, $TE(x, x^*, Y) \neq NDE(x, x^*, Y) + NIE(x, x^*, Y)$ because the change in $Y$ may depend on the interaction between $X$ and $Z$.
- CIE does not make sense because we cannot nullify the direct effect.

## Direct and Indirect Effects

- Assume binary random variables and, thus, $E[H_{t,a}] = E[H|t,a] = p(h|t,a)$.

| $T$ | $A$ | $p(h\|t,a)$ |
|---|---|---|
| 1 | 1 | 0.8 |
| 1 | 0 | 0.4 |
| 0 | 1 | 0.3 |
| 0 | 0 | 0.2 |

| $T$ | $p(a\|t)$ |
|---|---|
| 1 | 0.75 |
| 0 | 0.4 |

$$
\begin{aligned}
TE(t,\bar{t},H) &= E[H_t] - E[H_{\bar{t}}] = p(h|t) - p(h|\bar{t}) \\
&= p(h|t,a)p(a|t) + p(h|t,\bar{a})p(\bar{a}|t) - p(h|\bar{t},a)p(a|\bar{t}) - p(h|\bar{t},\bar{a})p(\bar{a}|\bar{t}) \\
&= 0.8 \cdot 0.75 + 0.4 \cdot 0.25 - 0.3 \cdot 0.4 - 0.2 \cdot 0.6 = 0.46 \\
NDE(t,\bar{t},H) &= (E[H|t,a] - E[H|\bar{t},a])p(a|\bar{t}) + (E[H|t,\bar{a}] - E[H|\bar{t},\bar{a}])p(\bar{a}|\bar{t}) \\
&= (0.8 - 0.3)0.4 + (0.4 - 0.2)0.6 = 0.32 \\
NIE(t,\bar{t},H) &= TE(t,\bar{t},H) + NDE(\bar{t},t,H) \\
&= 0.46 + (0.3 - 0.8)0.75 + (0.2 - 0.4)0.25 = 0.035
\end{aligned}
$$

- The average health improvement due to the treatment is 46 %.
- The treatment alone (i.e., keeping the pre-treatment aspirin dose) is responsible for 70 % of this improvement (i.e., $NDE/TE$).
- Therefore, a significant portion (30 %) is due to the treatment being able to stimulate the intake of aspirin.
- However, stimulating the intake of aspiring by other means than the treatment explains just 7 % of the improvement (i.e., $NIE/TE$).
- Therefore, the treatment is crucial and should not be replaced by a (less expensive) program for aspiring intake encouragement.

# Summary

- Conditional and Stochastic Actions
- Sequential Back-Door Criterion
- Direct and Indirect Effects

Thank you