G–Formula, Inverse Probability of Treatment Weighting and Optimal Sequential Treatments

Vanessa Didelez
Department of Mathematics
University of Bristol

Durham, July 2008

Overview

- Motivating examples
- Identifiability: stability and optimal strategies
- Illustrating g–computation
- Inverse probability weighting (IPTW)
- Comparison and conclusions

Example 1: Hospital Acquired Infection

Population: Patients in intensive care.

Response variable: Duration of stay in intensive care.
$\Rightarrow$ 'Survival' type of response.

Fixed covariates: Age, sex etc.

Time dependent covariates: Time of hospital acquired infection, organ function (score).

Question: Effect of 'preventing infection' on duration in intensive care.

Example 1: Hospital Acquired Infection

Causal question: What if we can prevent infection (higher standards of hygene etc.) for future patients, so that *no one* develops this infection?

Problem: In the data, time of infection (if at all) may depend on patient's recovery process and general health.
$\Rightarrow$ Cannot compare patients who *happen* not to have an infection with situation where infection is prevented for *everyone.*

Causal inference: Assume that certain aspects of the data are stable so that they are still valid under an intervention that prevents this infection.

Example 2: Chemotherapy for Cancer Patients

Population: Patients with operable breast cancer.

Response variable: Absence of residual malignant cells at surgery.
$\Rightarrow$ One binary outcome measured at end of study period.

Fixed covariates: Age, tumour size at start etc.

Time dependent covariates: Tumour size, number of chemotherapy cycles.

Question: Effect of the number of cycles on presence of tumour.

Example 2: Chemotherapy for Cancer Patients

Causal question: If we *prescribe* a certain number of chemotherapy cycles to future patients will this have a different effect than prescribing a lower (higher) number?

Problem: In the data, number (and timing) of chemotherapy cycles vary and may depend on patient history, patient's wish to terminate, toxic response, death, negative palpation result...
$\Rightarrow$ Patients who *happen* to have had e.g. four cycles cannot be compared with situation where *everyone* is prescribed four cycles.

Causal inference: Assume that certain aspects of the data are stable so that they are still valid under an intervention of prescribing no. of cycles.

Data Situation

$A_1, \ldots, A_N$ "action" variables $\longrightarrow$ can be 'manipulated'

$L_1, \ldots, L_N$ covariates $\longrightarrow$ (available) background information

$Y = L_{N+1}$ response variable

all measured over time, L_i before A_i

$\mathbf{A}^{<i} = (A_1, \ldots, A_{i-1})$ past up to before i; $\mathbf{A}^{\leq i}$, $\mathbf{A}^{>i}$ etc. similarly

Strategies

Strategy $\mathbf{s} = (s_1, \ldots, s_N)$ set of functions assigning an action

$$a_i = s_i(\mathbf{a}^{<i}, \mathbf{l}^{\leq i}) \text{ to each history } (\mathbf{a}^{<i}, \mathbf{l}^{\leq i})$$

If s_i constant for $(\mathbf{a}^{<i}, \mathbf{l}^{\leq i})$, $i = 1, \ldots, N$, then **unconditional strategy**

Otherwise **conditional/dynamic strategy**.

Causal Inference

Three issues:

1) What is the **causal target** of inference?
— effect of unconditional strategy;
— effect of *conditional* strategy;
— finding *optimal* strategy?

2) Under what **assumption** can we learn anything at all from observed data about our causal target of inference?
$\Rightarrow$ they are slighlty different depending on taget.

3) What **methods** should we use? For sequential treatments:
— G–computation
— Inverse probability of treatment weighting (IPTW)
— (G–estimation)

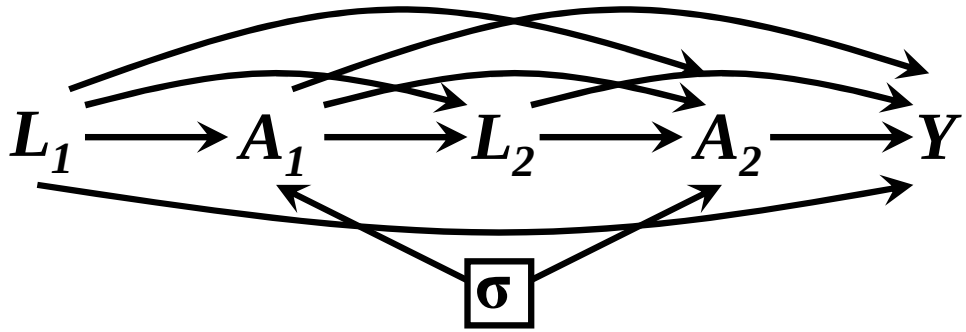
Identifiability

Wanted: $p(y; \mathrm{s})$ (or as before $E(k(Y); \mathrm{s})$) from observables.

Have seen that **simple stability** is sufficient to identify this.

$$L_i \perp\!\!\!\perp \sigma | (\mathbf{A}^{<i}, \mathbf{L}^{<i}) \quad \text{for all } i = 1, \ldots, N+1$$

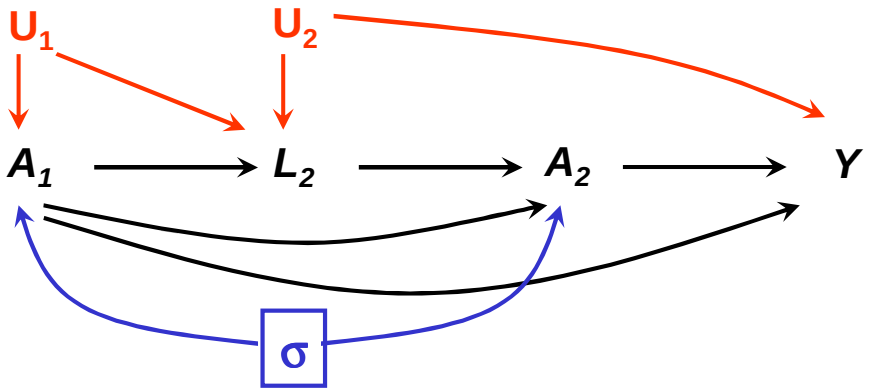
Or graphically:



(References: Robins (1986), Dawid & Didelez (2005))

Examples

Example 3: (from earlier)



Simple stability violated: $L_2 \not\perp\!\!\!\perp \sigma \mid A_1$ and $Y \not\perp\!\!\!\perp \sigma \mid (A_1, A_2, L_2)$

Relax Simple Stability?

For given strategy $\mathbf{s}$ can relax conditions for identifiability.

Assume extended stability holds wrt. $\mathbf{A}, \mathbf{L}, \mathbf{U}, Y$, and define 'new' joint distributions $p_i(\mathbf{a}, \mathbf{l}, \mathbf{u}, y) =$

$$p(\mathbf{a}^{\leq i}, \mathbf{l}^{\leq i}, \mathbf{u}^{\leq i}; o) \times p(\mathbf{a}^{>i}, \mathbf{l}^{>i}, \mathbf{u}^{>i}, y | \mathbf{a}^{\leq i}, \mathbf{l}^{\leq i}, \mathbf{u}^{\leq i}; \mathbf{s})$$

Relax Simple Stability?

For given strategy $\mathbf{s}$ can relax conditions for identifiability.

Assume extended stability holds wrt. $\mathbf{A}, \mathbf{L}, \mathbf{U}, Y$, and define 'new' joint distributions $p_i(\mathbf{a}, \mathbf{l}, \mathbf{u}, y) =$

$$\underbrace{p(\mathbf{a}^{\leq i}, \mathbf{l}^{\leq i}, \mathbf{u}^{\leq i}; o)}_{\text{obs. for } \leq i} \times p(\mathbf{a}^{>i}, \mathbf{l}^{>i}, \mathbf{u}^{>i}, y | \mathbf{a}^{\leq i}, \mathbf{l}^{\leq i}, \mathbf{u}^{\leq i}; \mathbf{s})$$

Relax Simple Stability?

For given strategy $\mathbf{s}$ can relax conditions for identifiability.

Assume extended stability holds wrt. $\mathbf{A}, \mathbf{L}, \mathbf{U}, Y$, and define 'new' joint distributions $p_i(\mathbf{a}, \mathbf{l}, \mathbf{u}, y) =$

$$p(\mathbf{a}^{\leq i}, \mathbf{l}^{\leq i}, \mathbf{u}^{\leq i}; o) \times \underbrace{p(\mathbf{a}^{>i}, \mathbf{l}^{>i}, \mathbf{u}^{>i}, y | \mathbf{a}^{\leq i}, \mathbf{l}^{\leq i}, \mathbf{u}^{\leq i}; \mathbf{s})}_{\text{strategy for } > i}$$

Relax Simple Stability?

For given strategy $\mathbf{s}$ can relax conditions for identifiability.

Assume extended stability holds wrt. $\mathbf{A}, \mathbf{L}, \mathbf{U}, Y$, and define 'new' joint distributions $p_i(\mathbf{a}, \mathbf{l}, \mathbf{u}, y) =$

$$p(\mathbf{a}^{\leq i}, \mathbf{l}^{\leq i}, \mathbf{u}^{\leq i}; o) \times p(\mathbf{a}^{>i}, \mathbf{l}^{>i}, \mathbf{u}^{>i}, y | \mathbf{a}^{\leq i}, \mathbf{l}^{\leq i}, \mathbf{u}^{\leq i}; \mathbf{s})$$

Theorem 1: sufficient condition for identifiability of $\mathbf{s}$ is

$$p_{i-1}(y|\mathbf{a}^{\leq i}, \mathbf{l}^{\leq i}) = p_i(y|\mathbf{a}^{\leq i}, \mathbf{l}^{\leq i}), \qquad i = 1, \ldots, N.$$

(Dawid & Didelez, 2005)

Simple stability implies the above.

Comments

Theorem 1, in words:
once we know a_i and the observable past variables, the distribution of Y does not depend on *how* a_i was generated, when $\mathbf{a}^{<i}$ is observational and $\mathbf{a}^{>i}$ follows the strategy.

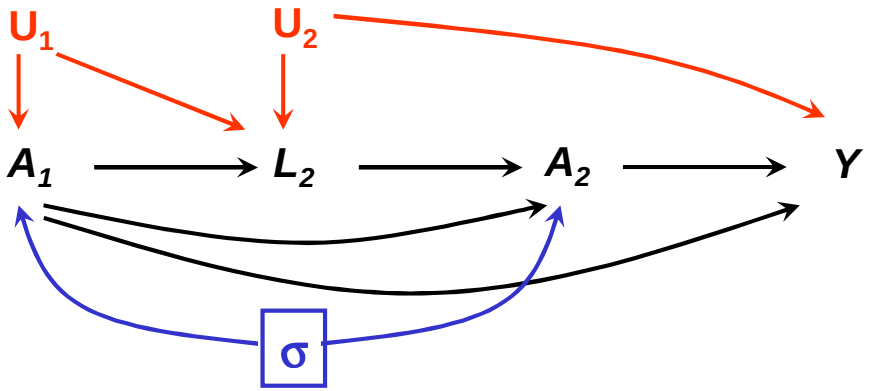
Note:
Essentially same as Pearl & Robins (1995) for *unconditional* strategies.

Graphical check: draw graphs D_i with
— $\text{pa}_{D_i}(A_k) = \text{pa}_o(A_k)$ for $k < i$
— $\text{pa}_{D_i}(A_k) = \text{pa}_s(A_k)$ as under strategy for $k > i$
— $\text{pa}_{D_i}(A_i) = \text{pa}_s(A_i) \cup \text{pa}_o(A_i) \cup \sigma$.

$\Rightarrow$ check separation $Y \perp\!\!\!\perp \sigma | (\mathbf{A}^{\leq i}, \mathbf{L}^{\leq i})$ in D_i, $i = 1, \ldots, N$

Example 3 ctd.

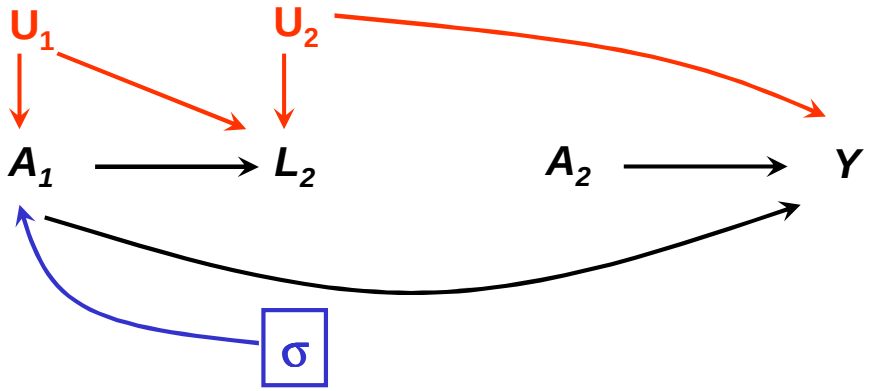
Assumed underlying structure (note: $L_1 = \emptyset$ here)

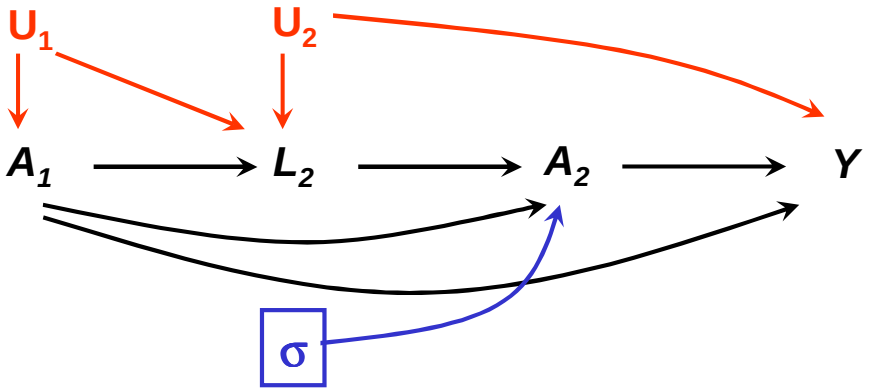


Now: also assume that s_2 is unconditional, i.e. choice of action A_2 in our strategy does not depend on past observations.

Example 3 ctd.

Then D_1 and D_2 are given by

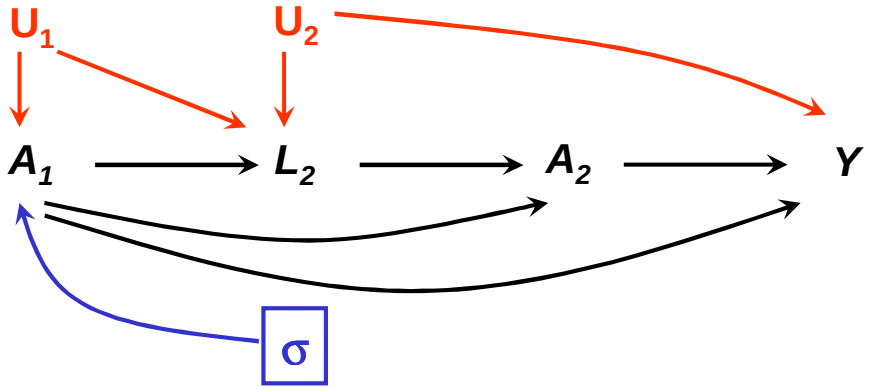


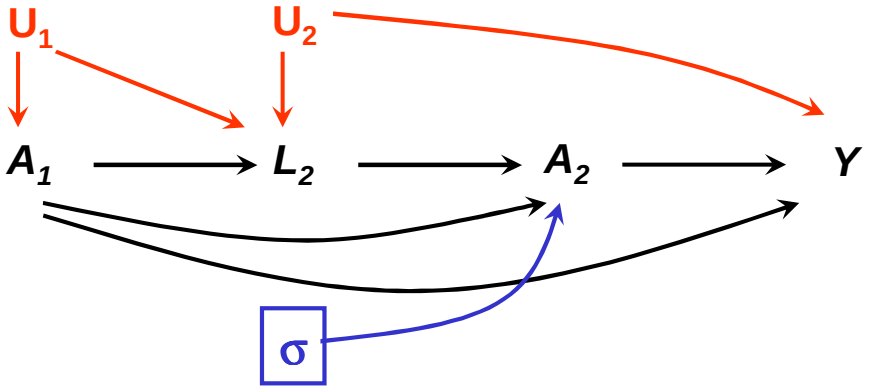


Can see that $Y \perp\!\!\!\perp \sigma | A_1$ in D_1 and $Y \perp\!\!\!\perp \sigma | (A_1, A_2, L_2)$ in D_2.

Example 3 ctd.

However, if s_2 is conditional, i.e. A_2 depends on past observations in our strategy, then D_1 and D_2 are given by





Now $Y \not\perp\!\!\!\perp \sigma | A_1$ in D_1.

This suggests that the 'relaxed' conditions are not so 'relaxed' for conditional interventions.

Optimal Treatment Strategy

Really, what we want to find is the *optimal* treatment strategy.

$\Rightarrow$ will typically be *conditional* strategy.

$\Rightarrow$ have to investigate whether all relevant conditional strategies are identified from data.

$\Rightarrow$ have to allow a_i to depend on whole observed past $(\mathbf{a}^{<i}, \mathbf{l}^{\leq i})$.

Result

Assumption 1: $\text{pa}_{\mathbf{s}}(A_i) \subset \text{pa}_o(A_i)$ for all $i = 1, \ldots, N$.

Assumption 2: each $L_1, \ldots, L_N$ is an ancestor of Y in D_0 (under strategy $\mathbf{s}$), $i = 1, \ldots, N$.

Theorem 2: With these assumptions, if the graphical check of Theorem 1 succeeds then we also have simple stability. (Dawid & Didelez, 2008)

Optimal strategies: Assumption 2 satisfied because
— actions A_i must be allowed to depend on past $\mathbf{L}^{\leq i}$
— and A_i ancestors of Y.

G–Computation

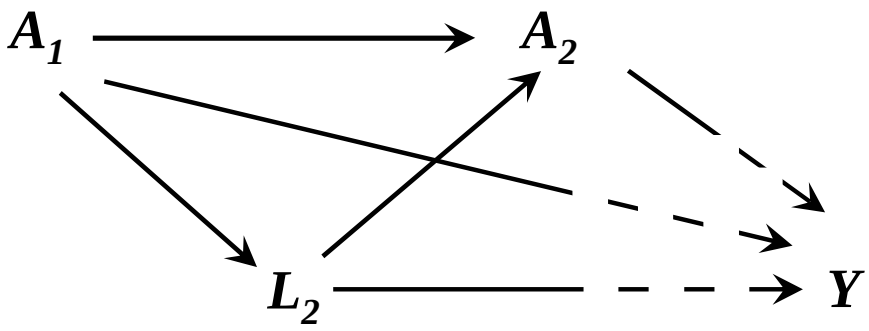
Assuming simple stability (or Theorem 1).

If $\mathbf{s}$ is (non–random) strategy fixing actions at $\mathbf{a}^{\leq N}$ then

$$p(y;\mathbf{s}) = \sum_{\mathbf{l}^{\leq N}} p(y|\mathbf{l}^{\leq N}, \mathbf{a}^{\leq N}; o) \prod_{i=1}^{N} p(l_i|\mathbf{l}^{<i}, \mathbf{a}^{<i}; o)$$

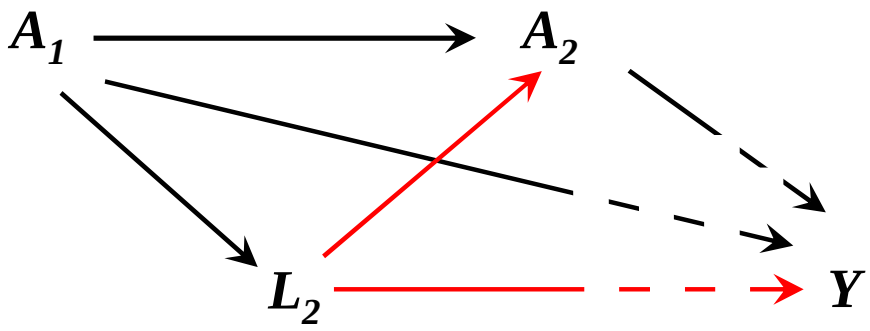
Why? A bit of intuition (assuming $\mathbf{s}$ is unconditional)...

Illustration



Simple scenario with $L_1 = \emptyset$ and only two actions.

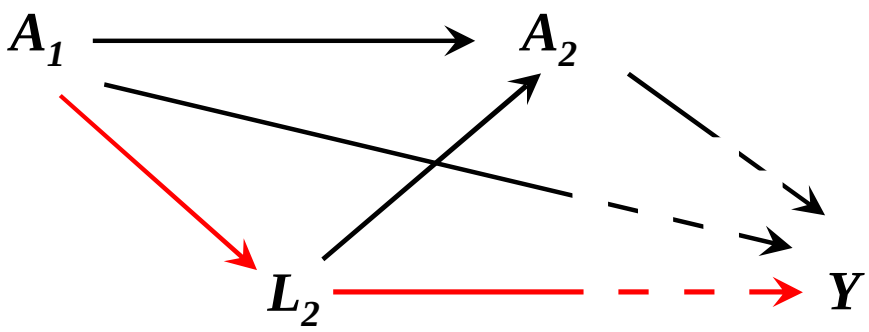
Illustration



L_2 'confounder' for A_2 and Y.

$\Rightarrow$ have to condition on L_2.

Illustration

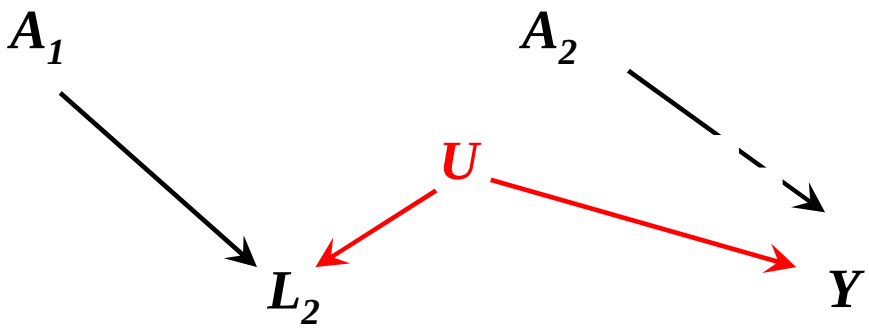


L_2 'mediator' for A_1 on Y.

$\Rightarrow$ *must not* condition on L_2!

$\Rightarrow$ coefficients in regression $p(y|a_1, a_2, l_2; o)$ not suitable for causal effect.

Illustration

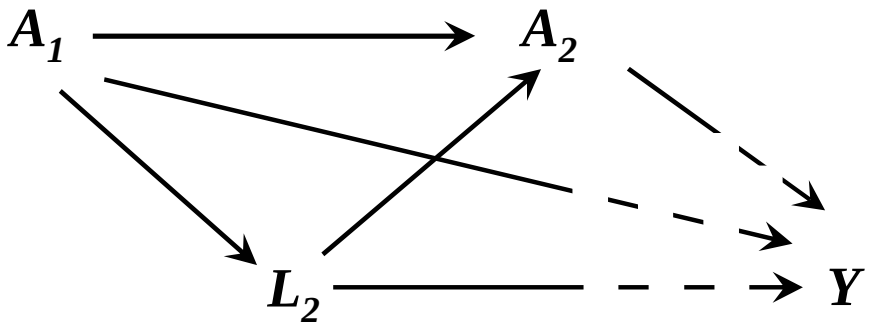


Also, conditioning on L_2 could induce selection effect.

$\Rightarrow$ *must not* condition on L_2!

$\Rightarrow$ coefficients in regression $p(y|a_1, a_2, l_2; o)$ not suitable for causal effect.

Illustration

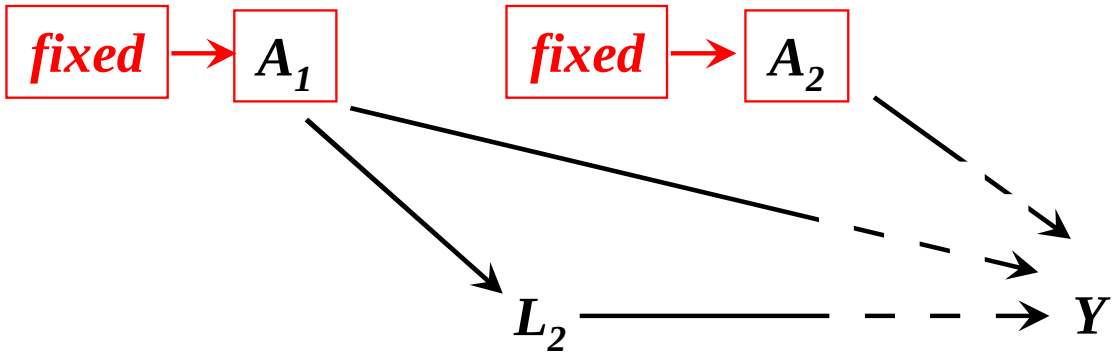


Under observational regime:

$$p(y, a_1, a_2, l_2; o) =$$

$$p(y|a_1, a_2, l_2; o) \underbrace{p(a_2|a_1, l_2; o)}_{\text{chosen by 'nature'}} p(l_2|a_1; o) \underbrace{p(a_1; o)}_{\text{chosen by 'nature'}}$$

Illustration



Under **intervention**:

Actions A_1, A_2 not random anymore, but fixed.

$\Rightarrow$ not generated by $p(a_1; o)$ and $p(a_2|a_1, l_2; o)$ anymore.

$\Rightarrow p(y, l_2; \mathbf{s}) = p(y|a_1, a_2, l_2; o)p(l_2|a_1; o)$

Stability (or Theo. 1) ensures remaining factors stay the same.

G–Computation

Integrate / sum out covariates $\mathbf{l}^{\leq N}$ to obtain $p(y; \mathbf{s})$

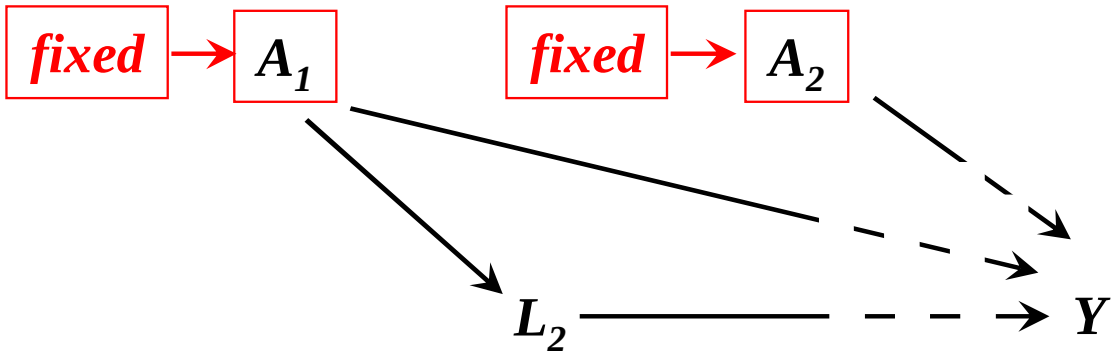
$$p(y; \mathbf{s}) = \sum_{\mathbf{l}^{\leq N}} p(y|\mathbf{l}^{\leq N}, \mathbf{a}^{\leq N}; o) \prod_{i=1}^{N} p(l_i|\mathbf{l}^{<i}, \mathbf{a}^{<i}; o)$$

So, estimate $p(y|\mathbf{l}^{\leq N}, \mathbf{a}^{\leq N}; o)$ and $p(l_i|\mathbf{l}^{<i}, \mathbf{a}^{<i}; o)$ from data.
$\Rightarrow$ plug in and done!

Note: for realistic settings will need to *model* these conditional distributions; misspecification may "multiply".

Alternative: inverse probability of treatment weighting.

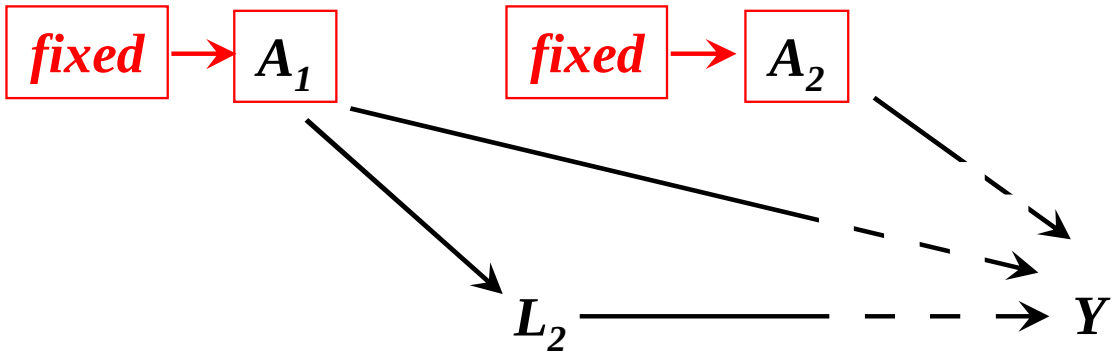
Reminder



Under **intervention**:

$$p(y, l_2; \mathbf{s}) = p(y|a_1, a_2, l_2; o)p(l_2|a_1; o)$$

Inverse Probability of Treatment Weighting



Under **intervention**:

$$p(y, l_2; \mathbf{s}) = p(y|a_1, a_2, l_2; o)p(l_2|a_1; o)$$

Same as

$$\ldots = \frac{p(y, a_1, a_2, l_2; o)}{p(a_1; o)p(a_2|a_1, l_2; o)} \longleftarrow \text{IPTW}$$

IPTW — More Generally

Assuming stability (or Theo. 1), and unconditional strategy $\mathbf{s}$

$$p(Y;\mathbf{s}) \longleftarrow \text{ intervention distribution}$$

$$= \sum_{\mathbf{l}^{\leq N}} p(y|\mathbf{l}^{\leq N}, \mathbf{a}^{\leq N}; o) \prod_{i=1}^{N} p(l_i|\mathbf{l}^{<i}, \mathbf{a}^{<i}; o) \longleftarrow \text{ G–comp.}$$

$$= \sum_{\mathbf{l}^{\leq N}} \frac{p(y, \mathbf{l}^{\leq N}, \mathbf{a}^{\leq N}; o)}{\prod_{i=1}^{N} p(a_i|\mathbf{l}^{\leq i}, \mathbf{a}^{<i}; o)} \longleftarrow \text{ IPTW}$$

where $a_1, \ldots, a_K$ are fixed according to intervention strategy $\sigma = \mathbf{s}$.

Not obvious how to *use* IPTW yet.

IPTW — Even More Generally

Define a probability measure P^* as

$$p^*(y, \mathbf{l}^{\leq N}, \mathbf{a}^{\leq N}) = p(y, \mathbf{l}^{\leq N}, \mathbf{a}^{\leq N}; o) \overbrace{\frac{\prod_i \tilde{p}(a_i|\mathbf{a}^{<i})}{\prod_i p(a_i|\mathbf{l}^{\leq i}, \mathbf{a}^{<i}; o)}}^{\text{weights } W(\mathbf{l}^{\leq N}, \mathbf{a}^{\leq N})^{-1}}$$

where $\tilde{P}$ is an arbitrary joint distribution for $\mathbf{A}^{\leq N}$.

Then we can show

– under P^*: $A_i \perp\!\!\!\perp \mathbf{L}^{\leq i} \mid \mathbf{A}^{<i}$ for all i

– and $p^*(\mathbf{a}^{\leq i}) = \tilde{p}(\mathbf{a}^{\leq i})$ for all i

IPTW: Marginal Structural Models (MSMs)

A MSM (Robins, 1999) is a model for $E(Y;\mathbf{s}) = \mu(\mathbf{a}^{\leq N};\beta)$ e.g. logistic regression etc.

Note: parameters β have immediate causal interpretation.

With P^* as before, we have

$$E(Y;\mathbf{s}) = E^*(Y|\mathbf{A}^{\leq N} = \mathbf{a}^{\leq N})$$

if $\sigma = \mathbf{s}$ is strategy fixing $\mathbf{A}^{\leq N}$ at $\mathbf{a}^{\leq N}$.

Estimating Equations

Hence, we obtain

$$E^*\{\phi(\mathbf{A}^{\leq N})(Y - \mu(\mathbf{A}^{\leq N}; \beta))\} = 0$$

i.e. unbiased estimating equations under P^*.

But this is the same as

$$E^*\{\phi(\mathbf{A}^{\leq N})(Y - \mu(\mathbf{A}^{\leq N}; \beta))\} = E\left\{\frac{\phi(\mathbf{A}^{\leq N})(Y - \mu(\mathbf{A}^{\leq N}; \beta))}{W(\mathbf{L}^{\leq N}, \mathbf{A}^{\leq N})}; o\right\}$$

i.e. *weighted* estimating equations are unbiased under observational regime.

‘Survival’ Outcome

Example: Duration of stay in intensive care.

If response is such a survival / duration, don’t want to model $E(Y; \mathbf{s})$ but instead hazard rate $\lambda(t; \mathbf{s})$ under strategy $\sigma = \mathbf{s}$.

$\Rightarrow$ Same principle, but with time varying weights.

$\Rightarrow$ Weights must be conditional on ‘survived so far’.

Comparison

G–computation:

- Need models for $p(y|\mathbf{l}^{\leq N}, \mathbf{a}^{\leq N}; o)$ and $p(l_i|\mathbf{l}^{<i}, \mathbf{a}^{<i}; o)\ \forall\ i$.
- Under misspecification might not even include the null hypothesis of no causal effect.
- No simple relation between parameters of above conditional distributions and causal effect.
- Apart from that, very general method.

Comparison

IPTW:

- Need models for $E(Y; \mathbf{s})$ and $p(a_i | \mathbf{l}^{\leq i}, \mathbf{a}^{<i}; o)$ for all i.
- By parameterising $E(Y; \mathbf{s}) = \mu(\mathbf{a}^{\leq N}; \beta)$ we obtain a causally interpretable parameter.
- Can easily be carried out by slight modification of popular regression methods.
- Cannot (easily) deal with interactions between L_i and A_i on Y and hence has problems with conditional strategies.

Conclusion

- Conditions for identifiability of a strategy may depend on type of strategy.
  If we want to find an optimal strategy, just check simple stability.

- Note: have not talked about *how* to find an optimal strategy; not easy!

- *Unconditional* strategies can be estimated with IPTW as an alternative to g–computation.

- Further method, g–estimation: based on counterfactuals — can we reformulate that?

- In applications need to scrutinise carefully assumption allowing to link observational and interventional regimes.