RANDOMIZATION AND CAUSATION (MATH-336)

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Midterm exam - solutions

Date: 14th of April, 2025
Time: 10:15–11:45

Name: _____

SCIPER: _____

INSTRUCTIONS TO CANDIDATES

- This midterm exam will contribute 20% to your final grade. To obtain the maximum number of points you should be clear about your reasoning and present your arguments explicitly. You have **90 minutes** to complete the exam.
- All that can be used for this exam is a pen. No notes, books, summaries, formula collections or calculators are allowed. All questions should be answered.
- The finest enumerated item in each question will be marked on a scale of $0-2$ points, indicating an incorrect, partially correct and completely correct answer respectively (half-points are not given). **The exam has 3 questions with a total of 48 points**.
- **Write the answer to every question in the other booklet (midterm exam - answers)**. Scrap paper will be provided for rough work, but only answers written in the booklet will be marked.
- At the end of the exam, you will have to return everything : the booklet with the questions, the booklet with your answers, and the scrap paper.

Mark question 1 (TOT: 12 points):

Mark question 2 (TOT: 14 points):

Mark question 3 (TOT: 22 points):

Question 1.

Let X, V, L, A, Y be a set of discrete variables topologically ordered as (X, V, L, A, Y). You are given the following NPSEM-IE (non-parametric structural equation model with independent errors) model:

$$\begin{aligned} X &= f_X(U_X) \\ V &= f_V(U_V) \\ L &= f_L(X, V, U_L) \\ A &= f_A(X, U_A) \\ Y &= f_Y(A, V, L, U_Y) \end{aligned}$$

where the U_X, U_V, U_L, U_A, U_Y are mutually independent error terms. We assume consistency and positivity throughout the question.

(1) Draw the causal DAG G for the above causal model.

(2) Give the Markov factorization for the joint density of (X,V,L,A,Y) under the DAG G.

(3) List all independencies (both unconditional and conditional) implied by the DAG on Y. That is: write all independencies of the form $Y \perp\!\!\!\perp N|T$ where N is a single random variable, and T is a set of random variables (T could be empty).

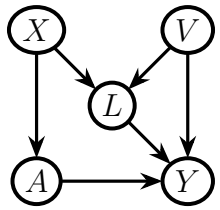
(4) Draw the SWIG $\mathcal{G}_a$ corresponding to the intervention $A = a$.

(5) Re-write all structural equations above under an intervention that sets $A = a$.

(6) Is the expected potential outcome $E[Y^a]$ identifiable with the g-formula? If yes, give an identification formula. Your expression should be minimal, in the sense that it does not include any unnecessary variables. If not identified, explain why.

Solutions:

(1) The causal DAG for the above causal model is:



G

(2) The Markov factorization of the joint density is

$$p(x, v, l, a, y) = p(x)p(v)p(l|x, v)p(a|x)p(y|a, l, v)$$

(3) There is only one single independence condition implied on Y:

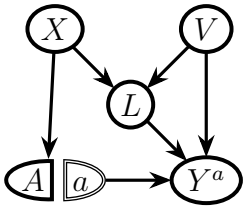
$$Y \perp\!\!\!\perp X|A, L, V$$

To see this, notice first that because there are direct arrows from A, L and V into Y, Y is not independent of any of A, L or V, neither unconditionally, nor conditionally. Next, from the Markov factorization, we see that $Y \perp\!\!\!\perp X|A, L, V$. Because of the path $X \to A \to Y$, A could not be removed from the conditioning set. Similarly, because of the path $X \to L \to Y$, L could not be removed from the conditioning set. Finally, Y is not independent of X given A and L only because conditioning on the collider L opens the path between X and Y through V. Consequently, V cannot be removed from the conditioning.

Thus, the single independence condition that holds is

$$\boxed{Y \perp\!\!\!\perp X|A, V, L}$$

(4) The requested SWIG is:



$\mathcal{G}_a$

(5) Under the intervention $A = a$, the NPSEM-IE models become:

$$\begin{aligned} X &= f_X(U_X) \\ V &= f_V(U_V) \\ L &= f_L(X, V, U_L) \\ A^{a+} &= a \qquad (A = a \text{ is also accepted.}) \\ Y^a &= f_Y(a, V, L, U_Y) \end{aligned}$$

(6) Yes, $E[Y^a]$ is identifiable as:

$$\begin{aligned} E[Y^a] &= E[E[Y^a|X]] && \text{(LOTE)} \\ &= E[E[Y^a|X, A=a]] && (Y^a \perp\!\!\!\perp A|X) \\ &= E[E[Y|X, A=a]] && \text{(Consistency)} \end{aligned}$$

Question 2.

In this exercise, consider a randomized controlled trial investigating the effect of a vaccine on the probability of contracting an infectious disease (for example, influenza or COVID-19). Also, consider that there is a genetic predisposition that would protect patients from getting the disease, regardless of whether they are vaccinated.

Specifically, let $A \in \{0, 1\}$ be the binary treatment assignment ($A = 1$ is to get the vaccine; A is randomized), Y be the binary outcome ($Y = 1$ is to have the disease in the year following the vaccine), and T be a binary indicator of having the genetic predisposition ($T = 1$ is to have the predisposition *i.e* to be immune to the disease, T is measured prior to treatment assignment).

The assumptions for T can be written as:

$$(A1) : Pr(Y^{a=1} = 1|T = 1) = Pr(Y^{a=0} = 1|T = 1) = 0$$
$$(A2) : T^{a=1} = T^{a=0}$$

We further assume that $Pr(Y^{a=0} = 1) > 0$, and $0 < Pr(T = 0) < 1$, along with consistency and positivity.

(1) Draw a DAG for this problem. Your DAG should include nodes for the variables A, T and Y only, and should only include necessary edges.

(2) Comment briefly on the plausibility of the consistency assumption in this setting (max 5 sentences).

(3) **Causal Risk Ratio (CRR).** In this question, we assume that the researchers are interested in estimating the Causal Risk Ratio (CRR), defined as :

$$CRR = \frac{Pr(Y^{a=1} = 1)}{Pr(Y^{a=0} = 1)}$$

We also define the Causal Risk Ratio in the non-immune subpopulation $T = 0$, denoted as $CRR_{T=0}$, as follows:

$$CRR_{T=0} = \frac{Pr(Y^{a=1} = 1|T = 0)}{Pr(Y^{a=0} = 1|T = 0)}$$

(a) Is the CRR identified? If yes, derive an identification formula. If no, explain why.

(b) Show that $CRR_{T=0} = CRR$.

(c) Now imagine the the genetic predisposition T is not measured. Is $CRR_{T=0}$ identified? Explain in maximum 3 sentences.

(4) Now imagine that the researchers are instead interested in estimating the Average Treatment Effect (ATE), defined as :

$$ATE = Pr(Y^{a=1} = 1) - Pr(Y^{a=0} = 1)$$

We also define the Conditional Average Treatment Effect in the non-immune subpopulation $T = 0$, denoted as $CATE_{T=0}$, as follows:

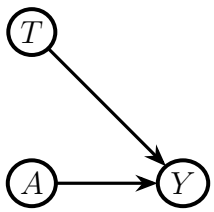
$$CATE_{T=0} = Pr(Y^{a=1} = 1|T = 0) - Pr(Y^{a=0} = 1|T = 0)$$

(a) Is the ATE identified? If yes, derive an identification formula. If no, explain why.

(b) Express $CATE_{T=0}$ as a function of the ATE and (potentially) some observed parameters and discuss the relationship between $CATE_{T=0}$ and the ATE : is $CATE_{T=0} >, =,$ or $< ATE$?

Solutions:

(1) The DAG is :



G'

(2) The consistency assumption requires two sub-assumptions to hold: (1) no multiple versions of the treatment, and (2) no between-unit interference. While it is reasonable to assume no multiple versions of the treatment, as long as there is only one type of vaccine and the administration is similar for all individuals, there may be between-unit interference. In fact, one patient receiving the vaccine may protect other individuals from contracting the disease. In a randomized clinical trial, we generally assume that the number of participants is so small that it is very unlikely that patients will interact with each other, making the consistency assumption plausible.

(3) (a) Yes, the CRR is identified. Indeed, $Y^a \perp\!\!\!\perp A = a$ in our setting (this can be read-off from the transformation of the DAG G' into a SWIG). From this, consistency, and positivity, it is straightforward to show that:

$$CRR = \frac{Pr(Y = 1|A = 1)}{Pr(Y = 1|A = 0)}$$

(b) We have that :

$$\begin{aligned}
CRR_{T=0} &= \frac{Pr(Y^{a=1} = 1|T = 0)}{Pr(Y^{a=0} = 1|T = 0)} \\
&= \frac{Pr(Y^{a=1} = 1, T = 0)Pr(T = 0)}{Pr(Y^{a=0} = 1, T = 0)Pr(T = 0)} \\
&= \frac{Pr(Y^{a=1} = 1, T = 0)}{Pr(Y^{a=0} = 1, T = 0)} \\
&= \frac{Pr(Y^{a=1} = 1) - Pr(Y^{a=1} = 1, T = 1)}{Pr(Y^{a=0} = 1) - Pr(Y^{a=0} = 1, T = 1)} \\
&= \frac{Pr(Y^{a=1} = 1) - Pr(Y^{a=1} = 1|T = 1)Pr(T = 1)}{Pr(Y^{a=0} = 1) - Pr(Y^{a=0} = 1|T = 1)Pr(T = 1)} \\
&= \frac{Pr(Y^{a=1} = 1)}{Pr(Y^{a=0} = 1)} && (Pr(Y^a = 1|T = 1) = 0) \\
&= CRR,
\end{aligned}$$

which concludes the proof.

(c) Yes, $CRR_{T=0}$ is identified by $\frac{Pr(Y=1|A=1)}{Pr(Y=1|A=0)}$, since $CRR_{T=0} = CRR$ and the CRR is identified

(4) (a) Yes, the ATE is identified as:

$$ATE = Pr(Y = 1|A = 1) - Pr(Y = 1|A = 0)$$

where we used $Y^a \perp\!\!\!\perp A = a$, positivity, and consistency.

(b) Let $a \in 0, 1$. We have that :

$$\begin{aligned} Pr(Y^a|T = 0) &= \frac{Pr(Y^a = 1, T = 0)}{Pr(T = 0)} \\ &= \frac{Pr(Y^a = 1) - Pr(Y^a = 1, T = 1)}{Pr(T = 0)} \\ &= \frac{Pr(Y^a = 1)}{Pr(T = 0)} \qquad \text{(by assumption (A1))} \end{aligned}$$

From that, we have :

$$\begin{aligned} CATE_{T=0} &= \frac{Pr(Y^{a=1} = 1)}{Pr(T = 0)} - \frac{Pr(Y^{a=0} = 1)}{Pr(T = 0)} \\ &= \frac{ATE}{Pr(T = 0)} \end{aligned}$$

Thus, we proved that :

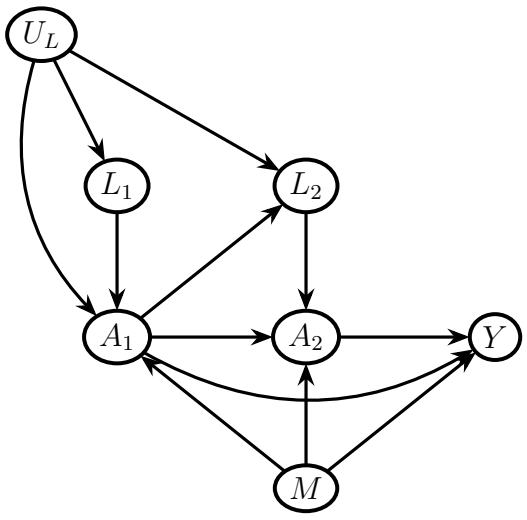
$$\boxed{CATE_{T=0} = \frac{ATE}{Pr(T = 0)}}$$

Since $Pr(T = 0) < 1$, $CATE_{T=0} > ATE$.

Question 3.

Motivation (not strictly needed to answer the question). Many treatments for chronic diseases (*e.g.*, diabetes, heart disease, cancer) must be taken continuously for years. However, these treatments are often associated with many side effects. Patients struggling with side effects may not be able to adhere to treatment. Non-adherence has implications for the interpretation of causal effects. For these reasons, researchers are often interested in the effectiveness of interventions that ensure adherence. However, it may be unrealistic to expect study participants to adhere to a strict treatment strategy that requires continuous use of treatment (perfect adherence). Other, more relaxed, adherence interventions are possible. In this exercise, we explore possible definitions of interventions on adherence. For simplicity, we will place ourselves in a setting with only two time points.

Notation and assumptions. Let A_1, A_2 be indicators of treatment at time points 1 and 2, respectively, Y be the outcome (measured after time 2), M is a set of measured covariates at baseline, L_1, L_2 measured (and discrete) covariates at time point 1 and 2, respectively, and U_L is an **unmeasured** covariate. We assume that A_1, A_2 and Y are binary, and that all other variables are discrete. We assume consistency for both static and dynamic time-varying regimes. We assume the following topological order $(U_L, M, L_1, A_1, L_2, A_2, Y)$ and the following DAG (G_1) throughout the question:



G_1

(1) **Perfect adherence**. First, let's consider a time-varying static strategy that sets perfect adherence :

(a) Draw the SWIG $\mathcal{G}_1(a_1, a_2)$ corresponding to the intervention that assigns A_1 to a_1 and A_2 to a_2 (for perfect adherence, we would set both a_1 and a_2 to 1).

(b) Find sets of variables V and W such that the following exchangeability conditions hold:

$$Y^{a_1,a_2} \perp\!\!\!\perp A_1 | V$$

$$Y^{a_1,a_2} \perp\!\!\!\perp A_2^{a_1} | V, A_1 = a_1, W.$$

The sets V and W should be **"minimal"**, in the sense that they should not contain unnecessary variables. In addition, V and W should only contain variables that are topologically ordered before A_1 for V and before $A_2^{a_1}$ for W.

(c) State the positivity condition(s) needed for identification.

(d) Derive an identification formula for $E[Y^{a_1,a_2}]$. Your expression should be "minimal", in the sense that it should not contain unnecessary variables.

Perfect adherence may not be feasible for certain patients who struggle with side effects. Researchers may be interested in strategies implementing relaxed version of adherence. We will explore two families of such strategies.

(2) **Stochastic strategies**.

We now consider an alternative dynamic stochastic strategy, denoted as g_π and defined as follows. At time point 1, patient is required to take treatment value a_1 : $A_1^{g_\pi +} = a_1$. At time point 2, let $X_2 \sim \mathcal{U}(0,1)$ be an exogenous randomly drawn value from a standard uniform distribution and Let $\pi : L_2^{g_\pi} \to (0,1)$ be an investigator-specified function of current covariate value. We set

$$A_2^{g_\pi +} = \begin{cases} 1 \text{ if } X_2 \leq \pi(L_2^{g_\pi}) \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

(a) Draw the dynamic SWIG $\mathcal{G}_2(g_\pi)$ corresponding to the above intervention g_π. *Hint: you may want to introduce a node for X_2.*

(b) State sets of variables V_2 and W_2 such that the following exchangeability conditions hold:

$$Y^{g_\pi} \perp\!\!\!\perp A_1 | V_2$$

$$Y^{g_\pi} \perp\!\!\!\perp A_2^{g_\pi} | V_2, A_1 = a_1, W_2.$$

The sets V_2 and W_2 should be **"minimal"**, in the sense that it should not contain unnecessary variables. The sets **should only contain variables that are topologically ordered before A_1 for V_2 and before $A_2^{g_\pi}$ for W_2.**

(c) What arrow could be removed from the DAG for $E[Y^{g_\pi}]$ to be identified with the extended g-formula? Give a single arrow (not a set of arrows).

(d) For this question and for the next question (2)(e), we assume that this arrow was deleted from the DAG. Derive an identification formula for $E[Y^{g_\pi}]$. Your expression should be "minimal", in the sense that it should not contain unnecessary variables. You do not have to state the required positivity assumptions and you can assume that they hold.

(e) For this question, we set $a_1 = 1$ and the function π to be constant equal to 1. Prove that the identification formula you found in previous question reduces to the identification formula you found in question (1)(d) with $a_1 = a_2 = 1$.

(3) **Natural grace period strategies**.

Natural grace period strategies were proposed as another family of strategies. It consists in allowing periods without treatment, as long as treatment breaks do not exceed a certain length of time. In this example, we allow one period of time maximum without treatment. Specifically, we want to study a time-varying dynamic strategy where treatment is assigned according to the decision rule g given by:

$$A_1^{g+} = A_1$$
$$A_2^{g+} = \begin{cases} A_2^g \text{ if } A_1^{g+} = 1 \\ 1 \text{ otherwise} \end{cases}$$

In words, at time point 1, patients take treatment as they would naturally do. If a patient did not take treatment at time point 1, then the patient is forced to take treatment at time point 2. If the patient did take treatment at time point 1, the patient takes treatment as he/she would naturally do at time point 2.

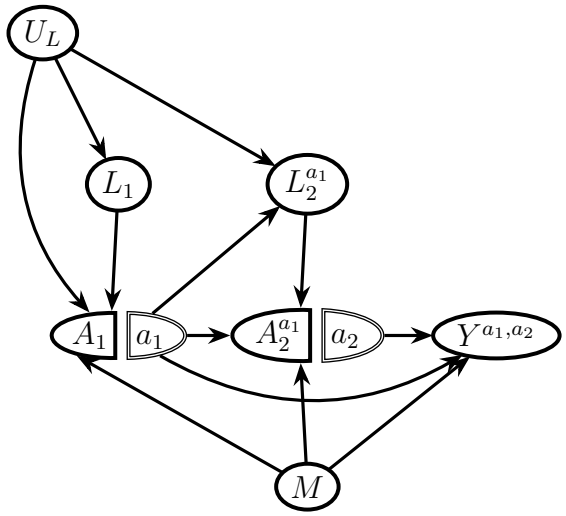
(a) Draw the dynamic SWIG $\mathcal{G}_3(g)$ corresponding to the above intervention.

(b) Why **can't** you use the same exchangeability conditions we discussed in the lectures for the extended g-formula to identify $E[Y^g]$?

Solutions:

(1) **Perfect adherence**

(a) The desired SWIG is shown below :



$\mathcal{G}(a_1, a_2)$

(b) Using d-separation we can see from the graph that :

- $(\boldsymbol{E_1}) : \boldsymbol{Y^{a_1,a_2}} \perp\!\!\!\perp \boldsymbol{A_1}|\boldsymbol{M}$. This is because paths from Y^{a_1,a_2} to A_1 either go from the constant intervened values a_1 or a_2; or through M which is not a collider in any of the paths.
- Similarly, $(\boldsymbol{E_2}) : \boldsymbol{Y^{a_1,a_2}} \perp\!\!\!\perp \boldsymbol{A_2^{a_1}}|\boldsymbol{M}, \boldsymbol{A_1} = \boldsymbol{a_1}$. Indeed, conditioning on A_1 could have open paths on which it is a collider, but all of these paths are d-separated by M.

Thus, $V = \{M\}$ and $W = \{\}$.

(c) Identifying $E[Y^{a_1,a_2}]$ requires the following positivity assumptions:

$$\forall m \text{ such that } Pr(M = m) > 0, Pr(A_1 = a_1|M = m) > 0 \qquad (P_1)$$

$$\forall m \text{ such that } Pr(M = m, A_1 = a_1) > 0, Pr(A_2^{a_1} = a_2|M = m, A_1 = a_1) > 0 \qquad (P_2)$$

(d) $E[Y^{a_1,a_2}]$ could be identified as follows:

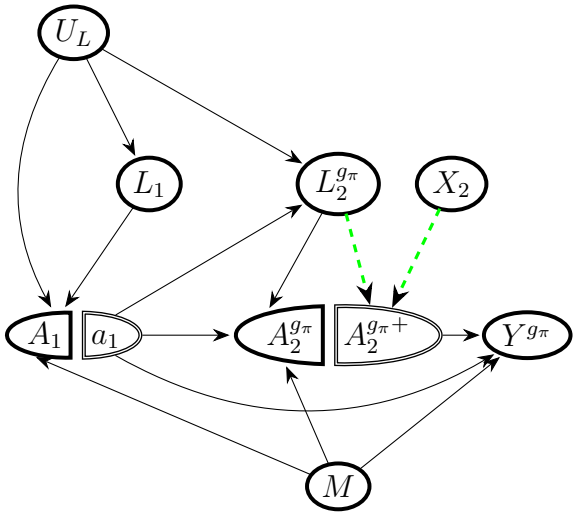
$$\begin{aligned}
E[Y^{a_1,a_2}] &= Pr(Y^{a_1,a_2} = 1) && (Y \text{ is binary}) \\
&= \sum_m Pr(Y^{a_1,a_2} = 1|M = m)Pr(M = m) && (\text{Law of total probability}) \\
&= \sum_m Pr(Y^{a_1,a_2} = 1|M = m, A_1 = a_1)Pr(M = m) && (P_1 \text{ and } E_1) \\
&= \sum_m Pr(Y^{a_1,a_2} = 1|M = m, A_1 = a_1, A_2^{a_1} = a_2)Pr(M = m) && (P_2 \text{ and } E_2) \\
&= \sum_m Pr(Y = 1|M = m, A_1 = a_1, A_2 = a_2)Pr(M = m) && (\text{Consistency})
\end{aligned}$$

Thus, we proved that:

$$\boxed{E[Y^{a_1,a_2}] = \sum_m E[Y|M=m, A_1=a_1, A_2=a_2]Pr(M=m)}$$

(2) **Stochastic strategies**

(a) The desired d-SWIG is shown below :



$\mathcal{G}_{\in}(g_\pi)$

where, with a slight abuse of notations, we directly replaced $A_1^{g_\pi+}$ by a_1.

(b) From the d-SWIG, and using the rules of d-separation, we can read:

- $(\boldsymbol{E'_1}) : \boldsymbol{Y^{g_\pi}} \perp\!\!\!\perp \boldsymbol{A_1}|\boldsymbol{M}, \boldsymbol{U_L}$. This set is minimal because :
  (i) M is required to block the path $A_1 \leftarrow M \rightarrow Y^{g_\pi}$.
  (ii) U_L is required to block the path $A_1 \leftarrow U_L \rightarrow L_2^{g_\pi} \rightarrow A_2^{g_\pi+} \rightarrow Y^{g_\pi}$.
- $(\boldsymbol{E'_2}) : \boldsymbol{Y^{g_\pi}} \perp\!\!\!\perp \boldsymbol{A_2^{g_\pi}}|\boldsymbol{U_L}, \boldsymbol{A_1}, \boldsymbol{M}, \boldsymbol{L_2^{g_\pi}}$
  The set is minimal because :
  (i) $L_2^{g_\pi}$ is required to block the path $A_2^{g_\pi} \leftarrow L_2^{g_\pi} \rightarrow A_2^{g_\pi+} \rightarrow Y^{g_\pi}$.

Finally, we have $V_2 = \{M, U_L\}$ and $W_2 = \{L_2^{g_\pi}\}$.

(c) The arrow from $U_L \rightarrow L_2$.

(d) Without the arrow from from $U_L \rightarrow L_2^{g_\pi}$ in the SWIG, the exchangeability conditions from previous questions become:

- $(\boldsymbol{E'_1}) : \boldsymbol{Y^{g_\pi}} \perp\!\!\!\perp \boldsymbol{A_1}|\boldsymbol{M}$.
- $(\boldsymbol{E'_2}) : \boldsymbol{Y^{g_\pi}} \perp\!\!\!\perp \boldsymbol{A_2^{g_\pi}}|\boldsymbol{A_1}, \boldsymbol{M}, \boldsymbol{L_2^{g_\pi}}$

Then, we can prove that :

$$\begin{aligned}
E[Y^{g_\pi}] &= \sum_m Pr(Y^{g_\pi}=1|M=m)p(m) && \text{(LOTP)}\\
&= \sum_m Pr(Y^{g_\pi}=1|A_1=a_1, M=m)p(m) && \text{(E'1 + positivity)}\\
&= \sum_{m,l_2} Pr(Y^{g_\pi}=1|l_2, A_1=a_1, M=m)p(L_2^{g_\pi}=l_2|A_1=a_1, M=m)p(m) && \text{(LOTP)}
\end{aligned}$$

$$= \sum_{m,l_2,a_2'} Pr(Y^{g_\pi} = 1 | A_2^{g_\pi +} = a_2', L_2^{g_\pi} = l_2, A_1 = a_1, M = m) p^g(l_2, a_2') p(l_2 | a_1, m) p(m)$$

(LOTP)

$$= \sum_{m,l_2,a_2'} Pr(Y^{g_\pi} = 1 | A_2^{g_\pi} = a_2', A_2^{g_\pi +} = a_2', L_2^{g_\pi} = l_2, A_1 = a_1, M = m) p^g(l_2, a_2') p(l_2 | a_1, m) p(m)$$

(E'2)

$$= \sum_{m,l_2,a_2'} Pr(Y = 1 | A_2 = a_2', L_2 = l_2, A_1 = a_1, M = m) p^g(l_2, a_2') p(L_2 = l_2 | A_1 = a_1, M = m) p(m)$$

(consistency)

where

$$p^g(l_2, a_2') = Pr(A_2^{g_\pi +} = a_2' | L_2^{g_\pi} = l_2, A_1 = a_1, M = m) = a_2' \cdot \pi(l_2) + (1 - a_2') \cdot (1 - \pi(l_2)).$$

Thus:

$$\boxed{E[Y^{g_\pi}] = \sum_{m,l_2,a_2'} E[Y | A_2 = a_2', M = m, A_1 = a_1, L_2 = l_2] p^g(l_2, a_2') p(l_2 | a_1, m) p(m)}$$

where $p^g(l_2, a_2') = a_2' \cdot \pi(l_2) + (1 - a_2') \cdot (1 - \pi(l_2))$

(e) In this case $\forall\, l_2 : \pi(l_2) = 1$ so that $p^g(l_2, a_2') = a_2'$. Also, the intervened value of treatment at time point 2 does not rely on the value of $L_2^{g_\pi}$, so that there is no green arrow in the SWIG above anymore. This implies that $Y^{g_\pi} \perp\!\!\!\perp L_2^{g_\pi} | M, A_2 = 1, A_1 = 1$. Knowing that, the identification formula reduces to :

$$E[Y^{g_\pi}] = \sum_{m,l_2} Pr[Y = 1 | A_2 = 1, M = m, A_1 = 1, L_2 = l_2] p(l_2 | A_1 = 1, m) p(m)$$

(The term with $a_2' = 0$ is 0)

$$= \sum_{m,l_2} Pr[Y^{g_\pi} = 1 | A_2 = 1, M = m, A_1 = 1, L_2^{g_\pi} = l_2] p(l_2 | A_1 = 1, m) p(m)$$

(consistency)

$$= \sum_{m,l_2} Pr[Y^{g_\pi} = 1 | A_2 = 1, M = m, A_1 = 1] p(l_2 | A_1 = 1, m) p(m)$$

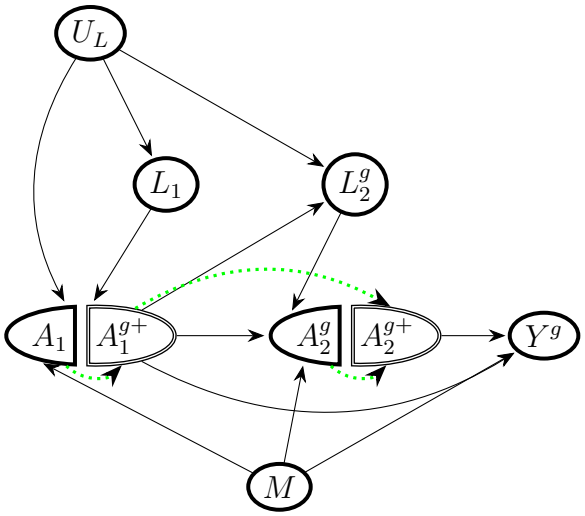
$(Y^{g_\pi} \perp\!\!\!\perp L_2^{g_\pi} | M, A_2 = 1, A_1 = 1)$

$$= \sum_{m} Pr[Y^{g_\pi} = 1 | A_2 = 1, M = m, A_1 = 1] p(m), \qquad (p(l_2 | A_1 = 1, m) \text{ sums to } 1)$$

which is the formula we found in question (1)(c) above.

(3) **Natural grace period strategies**

(a) The desired d-SWIG is shown below :



$\mathcal{G}_3(g)$

(b) Because A_2^{g+} is a function of previous intervened treatment value A_1^{g+} (see slide 179 from the lecture).