Graphical Models for Causal Inference using LaTeX

Eleanor Murray

July 23, 2017

Contents

Chapter 1

Introduction

This document contains LaTeX code for creating graphical models commonly used in causal inference and comparative effectiveness research. There are four types of graphical models included in this document:

- Directed acyclic graphs (DAGs);

- Single-world intervention graphs (SWIGs);

- Probability trees for inverse probability weighting;

- Decision trees for Markov models.

For more information, the most up-to-date code, and additional software for causal inference and data visualization, or to suggest improvements, please visit: github.com/eleanormurray

1.1 Setting up your LaTeX style

For the graphical models presented here, I use the tikzpicture package in LaTeX. To easily standardize the appearance of your graphical models, I recommend creating a custom arrow command. To do this, use the following commands.

```
Install the package tikzpicture, with the following libraries:
\usepackage{tikz}
\usetikzlibrary{positioning, calc, shapes.geometric, shapes.multipart,
        shapes, arrows.meta, arrows,
        decorations.markings, external, trees}

Create custom arrow style:
\tikzstyle{Arrow} = [
        thick,
        decoration={
                markings,
                mark=at position 1 with {
                        \arrow[thick]{latex}
                        }
                },
        shorten >= 3pt, preaction = {decorate}
        ]
```

Your arrows will look like this: $\longrightarrow$

Define the following variables for use throughout this document:

- Treatments, exposures, interventions, or independent variables for which the causal effect is of interest represented with A

- Outcomes or dependent variables of interest with Y

- Known covariates or confounders with L

- Unknown covariates or confounders with U

- Censoring, selection, loss to follow-up, or dropout with C

- When multiple time points are of interest, they are indexed with a subscript

- When counterfactual variables are of interest, they are indexed with a superscript

Chapter 2

Directed acyclic graphs (DAGs)

Directed acyclic graphs (DAGs) are tools for evaluating hypothesized relationships between variables in terms of their potential to cause bias in a given study or analysis. They are useful tools for designing a study or analysis, for assessing potential causes of unexpected results, or for teaching causal inference.

There are two main types of DAGs that are commonly used in causal inference:

- Statistical DAGs contain measured or known variables only;
- Causal DAGs contain all variables of interest plus all common parents of two or more variables of interest even if the identity of these variables is unknown.

DAGs typically also indicate the flow of time through a temporal ordering from either left (past) to right (future or present) or top (past) to bottom (future or present). The DAGs here use a left to right temporal ordering. These DAGs are also drawn under the null, meaning that they represent the null hypothesis that A has no causal effect on Y.

2.1 Statistical DAGs

Statistical DAGs encode beliefs about associations between known variables. In statistical DAGs arrows imply potential associations which may or may not be causal. Any variables not connected by arrows in the statistical DAG are known or assumed not to be associated.

Statistical DAGs can be built from datasets using machine learning or other statistical techniques, but should not be used to determine whether two variables are causally related.

2.1.1 Statistical DAG for point exposures

The two DAGs below represent the simplest possible statistical DAGs, with only two variables, an exposure or treatment variable, A, and an outcome or dependent variable, Y. The first DAG represents the assumption that the null hypothesis of no association between the exposure and the outcome is true (i.e. $E[Y|A = a] = E[Y|A = a']$). The second DAG does not make this assumption – this DAG includes an arrow between A and Y and therefore allows the possibility of an association between exposure and outcome but does not require that this association exists.

```
DAG 1 - null hypothesis:

 A        Y

Code for DAG 1:

\begin{tikzpicture}
 \node (1) {A};
 \node [right =of 1] (2) {Y};
\end{tikzpicture}


DAG 2 - no hypothesis:

 A ——————▶ Y

Code for DAG 2:

\begin{tikzpicture}
 \node (1) {A};
 \node [right =of 1] (2) {Y};

 \draw[Arrow] (1.east) -- (2.west);
\end{tikzpicture}
```

Statistical DAG with confounding

In the next DAG, we have added a known confounder, L – that is, a variable which is associated with the exposure, and with the outcome, and occurs before both in time, and which could therefore be an explanation for any spurious (non-causal) association which is detected between exposure and outcome. L could also represent a set of confounding variables with similar relationships to A and Y. The DAG below is drawn under the assumption that the null hypothesis is true.
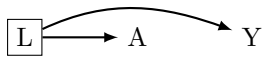
```
DAG:

 L ═══════▶ A      Y

Code:

\begin{tikzpicture}
 \node (1) {L};
 \node [right =of 1] (2) {A};
 \node [right =of 2] (3) {Y};

 \draw[Arrow] (1.east) -- (2.west);
 \draw[Arrow] (1) to [out=25, in=160] (3);
\end{tikzpicture}
```

Conditioning or restricting on a confounder

In this DAG, there is a box around L to indicate that the variable, L, has been restricted to a certain set of values, or has been adjusted for statistically in the analysis. The DAG below is drawn under the assumption that the null hypothesis is true. The only difference in the code is that we have drawn a rectangle around the first node. However, the DAG now encodes the assumption that the exposure and outcome will not be associated in any analysis that conditions or restricts on L.

```
DAG:

L⟶A    Y

Code:

\begin{tikzpicture}
 \node [rectangle, draw] (1) {L};
 \node [right =of 1] (2) {A};
 \node [right =of 2] (3) {Y};

 \draw[Arrow, thick] (1.east) -- (2.west);
 \draw[Arrow, thick] (1) to [out=25, in=160] (3);
\end{tikzpicture}
```

Conditioning or restricting on a collider

In this DAG, there is an additional variable, C, for censoring or loss to follow-up. Since we typically only have data on individuals who remain in our study, C is equal to 0 (uncensored) for all individuals for whom we have data, and is therefore restricted on by definition. Thus, in any DAG with censoring, there should be a box around C to indicate this restriction.

Note, that when we are conditioning on a downstream consequence of exposure we create the possibility of collider bias. Therefore, it is likely that a statistical DAG created from data in which we condition or restrict on C will detect an association between the exposure and outcome. This likelihood is represented in a statistical DAG by an arrow between A and Y. If we omit this arrow from A to Y, we are making an assumption of non-informative censoring. This is equivalent to assuming that any observations with missing outcome information are missing completely at random (MCAR). See the causal DAG section for a better way to represent this relationship.

The DAG below is drawn under the assumption that the null hypothesis is true.
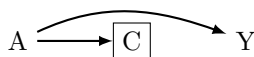
```
DAG:

A⟶C    Y

Code:

\begin{tikzpicture}
 \node (1) {A};
 \node [rectangle, draw, right =of 1] (2) {C};
 \node [right =of 2] (3) {Y};
```

```
  \draw[Arrow, thick] (1) to [out=25, in=160] (3);
  \draw[Arrow, thick] (1.east) -- (2.west);
\end{tikzpicture}
```

### 2.1.2 Statistical DAG for sustained exposures

When the data are longitudinal and available at multiple time points, DAGs must have separate nodes for each variable at each time in order to remain acyclical and avoid loops. The time points can be indicated using subscripts.

Here, there is a time-varying confounder, L, which is also affected by prior treatment. If the analysis does not adjust for L at all time points, there will be bias due to confounding. However, if the analysis adjusts for L at all time points through simple stratification-based regression analyses, there could be bias due to conditioning on a collider at L_k.

The DAG below is a complete DAG, under the assumption that the null hypothesis is true, meaning that all possible arrows are present except those directly from each treatment, A, to the outcome, Y. However, if there are any common causes of the confounder L_k and the outcome, then any attempt to create this statistical DAG from data alone will result in observed associations between A and Y at all time points. See the causal DAGs section for more information.

For simplicity, in the DAG below Y_k is omitted and Y_{k+1} can be considered an indicator for the cumulative probability (incidence) of the outcome by the end of follow-up.

DAG:

$L_{k-1} \longrightarrow A_{k-1} \longrightarrow L_k \longrightarrow A_k \qquad Y_{k+1}$
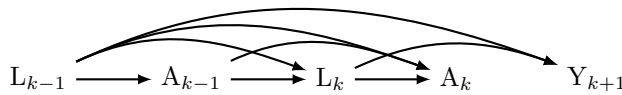
Code:

```
\begin{tikzpicture}
 \node (1) {L$_{k-1}$};
 \node [right =of 1] (2) {A$_{k-1}$};
 \node [right =of 2] (3) {L$_{k}$};
 \node [right =of 3] (4) {A$_{k}$};
 \node [right =of 4] (5) {Y$_{k+1}$};

 \draw[Arrow, thick] (1.east) -- (2.west);
 \draw[Arrow, thick] (1) to [out=25, in=160] (3);
 \draw[Arrow, thick] (1) to [out=25, in=160] (4);
 \draw[Arrow, thick] (1) to [out=25, in=160] (5);

 \draw[Arrow, thick] (2.east) -- (3.west);
 \draw[Arrow, thick] (2) to [out=25, in=160] (4);

 \draw[Arrow, thick] (3.east) -- (4.west);
 \draw[Arrow, thick] (3) to [out=25, in=160] (5);
```

```
\end{tikzpicture}
```

2.2 Causal DAGs

Causal DAGs encode beliefs about causal relationships between variables of interest. In causal DAGs, arrows between variables imply potential causal relationships. Any variables not connected by arrows in the causal DAG are known or assumed to lack of causal relationship. When two variables are known to be associated but not believed to be causally related, this is represented by the inclusion of an unknown variable, U.

2.2.1 Causal DAG for point exposures

The two DAGs below represent the simplest possible causal DAGs, with only two known variables, an exposure or treatment variable, A, and an outcome or dependent variable, Y. Unlike the statistical DAG, the first causal DAG does not represent the assumption that the null hypothesis of no association between the exposure and the outcome is true (i.e. $E[Y|A = a] = E[Y|A = a']$), since an association could be observed due to confounding by U. Instead, the first causal DAG represents the (stronger) assumption that the null hypothesis of no causal relationship between the exposure and the outcome is true (i.e. $E[Y^a|A = a] = E[Y^{a'}|A = a']$).

The second DAG does not make either of these assumptions – this DAG includes an arrow between A and Y, and an unknown common cause U, and therefore allows the possibility of a causal relationship or a non-causal association between exposure and outcome but does not require that either of these exists.

```
DAG 1 - null hypothesis:

 U ───────► A        Y

Code for DAG 1:

\begin{tikzpicture}
 \node (1) {A};
 \node [right =of 1] (2) {Y};
 \node [left =of 1] (3) {U};

 \draw[Arrow] (3.east) -- (1.west);
 \draw[Arrow] (3) to [out=25, in=160] (2);
\end{tikzpicture}


DAG 2 - no hypothesis:

 U ───────► A ──────► Y

Code for DAG 2:

\begin{tikzpicture}
 \node (1) {A};
 \node [right =of 1] (2) {Y};
 \node [left =of 1] (3) {U};

 \draw[Arrow] (1.east) -- (2.west);
 \draw[Arrow] (3.east) -- (1.west);
 \draw[Arrow] (3) to [out=25, in=160] (2);
\end{tikzpicture}
```

Causal DAG with confounding

Any causal DAG must include all possible common causes of any two variables on the DAG. In the simplest case, where all confounders are known, the statistical DAG for confounding will be the same as the causal DAG for confounding. In more complex cases, the causal DAG for confounding can include additional unknown common causes of A and Y, of L and A, and / or of L and Y.

DAG:
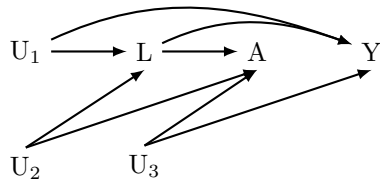

```
Code:

\begin{tikzpicture}
 \node (1) {L};
 \node [right =of 1] (2) {A};
 \node [right =of 2] (3) {Y};
 \node [left =of 2] (5) {U$_1$};
 \node [below =of 5] (6) {U$_2$};
 \node [below =of 2] (7) {U$_3$};

 \draw[Arrow] (2.east) -- (3.west);
 \draw[Arrow] (2) to [out=25, in=160] (4);

 \draw[Arrow] (5.east) -- (2.west);
 \draw[Arrow] (5) to [out=25, in=160] (4);

 \draw[Arrow] (6.north) -- (2.south);
 \draw[Arrow] (6.north) -- (3.south);

 \draw[Arrow] (7.north) -- (3.south);
 \draw[Arrow] (7.north) -- (4.south);
\end{tikzpicture}
```

Causal DAG, conditioning or restricting on a collider

As in the statistical DAG with a collider, the causal DAG has an additional variable, C, for censoring or loss to follow-up. Since we typically only have data on individuals who remain in our study, C is equal to 0 (uncensored) for all individuals for whom we have data, and is therefore restricted on by definition. Thus, in any DAG with censoring, there should be a box around C to indicate this restriction.

Unlike the statistical DAG, the causal DAG also includes an unknown common cause of C and Y. It is likely that most studies with censoring or loss to follow-up have some known or unknown common causes of censoring and the outcome. Omitting U (or some other, known, common cause) from the causal DAG makes the assumption of non-informative censoring or that the outcome information is missing completely at random (MCAR). By including U, we relax this (strong) assumption, and instead allow censoring to be informative. If we can identify, and measure, U then we can adjust for the censoring process and are now only assuming that the outcome is missing at random (MAR), or that censoring is

conditionally non-informative.

Note that since C is a common effect of A and of U, C is a collider. As such, a data analysis that does not account for U will result in an artificial (non-causal) association between A and Y through the path A→C←U→L which is opened by restriction on the collider, C. Using a causal DAG allows us to easily see the potential for collider bias when conditioning or restricting on a post-exposure variable.
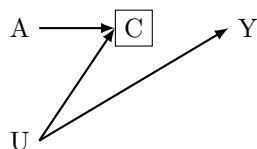
```
DAG:

  A ——————▶ [C]      ▶ Y

  U ╱

Code:

\begin{tikzpicture}
 \node (1) {A};
 \node [rectangle, draw, right =of 1] (2) {C};
 \node [right =of 2] (3) {Y};
 \node [below =of 1] (4) {U};

 \draw[Arrow, thick] (1.east) -- (2.west);
 \draw[Arrow, thick] (4.east) -- (2.west);
 \draw[Arrow, thick] (4.east) -- (3.west);
\end{tikzpicture}
```

2.2.2 Causal DAG for sustained exposures

When the data are longitudinal and available at multiple time points, DAGs must have separate nodes for each variable at each time in order to remain acyclical and avoid loops. The time points can be indicated using subscripts.

Here, there is a time-varying confounder, L, which is also affected by prior treatment. If the analysis does not adjust for L at all time points, there will be bias due to confounding. However, if the analysis adjusts for L at all time points through simple stratification-based regression analyses, there could be bias due to conditioning on a collider at L_k.

In the causal DAG, we represent this potential for bias by including one or more common causes of L and Y. Here, we make the simplest assumption that a single, baseline, cause (or set of causes) affects L and Y at all time points.

The DAG below is a complete DAG, under the assumption that the null hypothesis is true, meaning that all possible arrows are present except those directly from each treatment, A, to the outcome, Y.

For simplicity, in the DAG below Y_k is omitted and Y_{k+1} can be considered an indicator for the cumulative probability (incidence) of the outcome by the end of follow-up.

DAG:

$$L_{k-1} \longrightarrow A_{k-1} \longrightarrow L_k \longrightarrow A_k \qquad Y_{k+1}$$

$$U$$
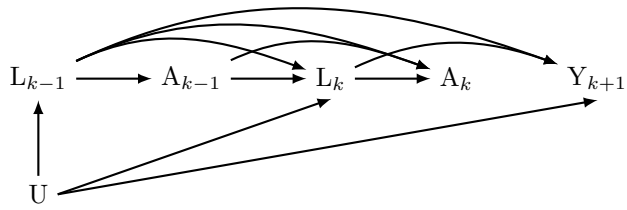
Code:

```
\begin{tikzpicture}
 \node (1) {L$_{k-1}$};
 \node [right =of 1] (2) {A$_{k-1}$};
 \node [right =of 2] (3) {L$_{k}$};
 \node [right =of 3] (4) {A$_{k}$};
 \node [right =of 4] (5) {Y$_{k+1}$};
 \node [below =of 1] (6) {U};

 \draw[Arrow, thick] (1.east) -- (2.west);
 \draw[Arrow, thick] (1) to [out=25, in=160] (3);
 \draw[Arrow, thick] (1) to [out=25, in=160] (4);
 \draw[Arrow, thick] (1) to [out=25, in=160] (5);

 \draw[Arrow, thick] (2.east) -- (3.west);
 \draw[Arrow, thick] (2) to [out=25, in=160] (4);

 \draw[Arrow, thick] (3.east) -- (4.west);
 \draw[Arrow, thick] (3) to [out=25, in=160] (5);

 \draw[Arrow, thick] (6.north) -- (1.south);
 \draw[Arrow, thick] (6.east) -- (3.south);
 \draw[Arrow, thick] (6.east) -- (5.south);
\end{tikzpicture}
```

Chapter 3

Single-world intervention graphs (SWIGs)

Single-world intervention graphs (SWIGs) are similar to DAGs in that they have the same key properties of being directed and acyclic. In addition, SWIGs explicitly include the intervention or experimental condition of interest and the counterfactual outcome of interest, allowing direct visual assessment of the causal inference assumption of exchangeability or conditional exchangeability.

The intervention or experimental condition of interest is indicated by the use of split nodes. These nodes are split in two by a | symbol. The left-hand side of the split node (typically denoted with an upper case letter) indicates the observed value of intervention or experimental variable. The right-hand side of the split node (typically denoted with a lower case letter) indicates the investigator-determined intervention or experimental variable.

Counterfactual nodes are denoted by a superscript. For example, counterfactual outcomes or dependent variables of interest may be denoted Y^a, and known covariates or confounders which are affected by prior interventions may be denoted L^a and are now also counterfactuals.

Note that to demonstrate the change in arrows related to the intervention in a SWIG, all SWIGs below are not drawn under the assumption that the null hypothesis is true – that is, the SWIGs include arrows from the intervention nodes directly to the outcome. This is not a required feature.

We can draw two general types of SWIGs:

- Intervention-specific SWIGs, and
- Template SWIGs

For teaching purposes, we will first work through intervention-specific SWIGs but then we will describe the general template. SWIG templates are more often used, since they allow assessment of a range of interventions.

3.1 Intervention-specific SWIGs

Intervention-specific SWIGs specify the exact value which we are interested in for a specific research question. This could be the intervention that we will assign to individuals in one arm of a randomized trial, or it could be one of the hypothetical interventions we are interested in emulating using observational data to emulate a hypothetical target trial.
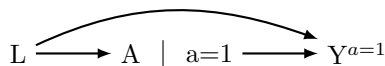
3.1.1 Intervention-specific SWIG for interventions on point exposures

The intervention-specific SWIG for an intervention on a point exposure looks similar to the DAG for point exposures, except that the intervention variable, A, is now split into two nodes - an observed node, and an intervention node. In the intervention-specific SWIG below, we are interested in the point intervention 'all individuals to take treatment (i.e. a=1)'.

In this SWIG, the intervention node is therefore set to 'a=1'. All downstream consequences of the intervention node are now counterfactual (unless we make assumptions), and are therefore labeled with superscripts which indicate the counterfactual intervention world in which they occur.

As noted above, the SWIG below is not drawn under the assumption that the null hypothesis is true.

SWIG:

L $\longrightarrow$ A $\mid$ a=1 $\longrightarrow$ Y$^{a=1}$

Code:

```
\begin{tikzpicture}
[
array/.style={rectangle split,
        rectangle split parts = 3,
        rectangle split horizontal,
    minimum height = 2em
    }
]
  \node (1) {L};
  \node [array, right =of 1] (2) {%
        {A}
    \nodepart{two}{$|$}
    \nodepart{three}{a=1}
    };
  \node [right =of 2] (3) {Y$^{a=1}$};

  \draw[Arrow, thick] (1.east) -- (2.west);
  \draw[Arrow, thick] (2.two east) -- (3.west);
  \draw[Arrow, thick] (1) to [out=25, in=160] (3);

\end{tikzpicture}
```
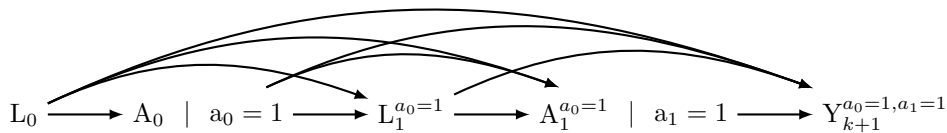
3.1.2 Intervention-specific SWIG for static interventions on sustained exposures

The intervention-specific SWIG for a sustained intervention looks similar to the DAG for sustained exposures, except for the SWIG-specific features. However, in the SWIG we can now represent two different types of interventions on sustained exposures: static interventions, and dynamic interventions. Static interventions are interventions which are determined from baseline and are either identical for every individual or dependent only on baseline or pre-baseline characteristics. Dynamic interventions are interventions which are not fully known until the end of follow-up; these interventions depend on characteristics which change over time.

In the intervention-specific SWIG below, we are interested in the point intervention 'all individuals to take treatment at all time points (i.e. $a_k=1$)'. In this SWIG the intervention node is therefore set to 'a=1' at every time point.

All downstream consequences of the intervention node are now counterfactual (unless we make assumptions), and are therefore labeled with superscripts which indicate the counterfactual intervention world in which they occur. Now, these nodes are labeled with all prior interventions, so that variables which are downstream of multiple intervention nodes will have multiple counterfactual superscripts. Note that the L_1 node is not split because this is a time-varying covariate and not an intervention variable.

SWIG:

$$L_0 \longrightarrow A_0 \mid a_0 = 1 \longrightarrow L_1^{a_0=1} \longrightarrow A_1^{a_0=1} \mid a_1 = 1 \longrightarrow Y_{k+1}^{a_0=1,a_1=1}$$

Code:

```
\begin{tikzpicture}[
array/.style={rectangle split,
        rectangle split parts = 3,
        rectangle split horizontal,
    minimum height = 2em
    }
]
  \node (1) {L$_0$};
  \node [array, right =of 1] (2) {%
        {A$_0$}
    \nodepart{two}{$|$}
    \nodepart{three}{a$_0 =1$}
    };
  \node [right =of 2] (3) {L$_1^{a_{0}=1}$};
    \node [array, right =of 3] (4) {%
        {A$_1^{a_{0}=1}$}
    \nodepart{two}{$|$}
    \nodepart{three}{a$_1 =1$}
    };
  \node [right =of 4] (5) {Y$_{k+1}^{a_0 =1, a_1 =1}$};

  \draw[Arrow, thick] (1.east) -- (2.west);
  \draw[Arrow, thick] (1) to [out=25, in=160] (3);
  \draw[Arrow, thick] (1) to [out=25, in=160] (4);
```

```
  \draw[Arrow, thick] (1) to [out=25, in=160] (5);

  \draw[Arrow, thick] (2.east) -- (3.west);
  \draw[Arrow, thick] (2) to [out=25, in=160] (4);
  \draw[Arrow, thick] (2) to [out=25, in=160] (5);

  \draw[Arrow, thick] (3.east) -- (4.west);
  \draw[Arrow, thick] (3) to [out=25, in=160] (5);

  \draw[Arrow, thick] (4.east) -- (5.west);

\end{tikzpicture}
```

3.1.3 Intervention-specific SWIG for dynamic interventions on sustained exposures

Intervention-specific SWIGs for dynamic interventions allow the intervention value to depend on time-varying characteristics of the individual. Here, we are interested in the dynamic intervention 'all individuals to take treatment at baseline, and continue treatment only while the time-varying covariate $L=1$'. This means that we treat everyone at baseline, and then continue to treat only those individuals whose the time-varying covariate indicates that treatment remains appropriate. In the example below, this is indicated by $L=1$.

We indicate that the specific treatment received by an individual during follow-up depends on L by assigning the intervention value $a_1 = L$. This shorthand allows us to see that treatment is given at time $t = 1$ (i.e. $a_1=1$) only when the time-varying covariate indicates this is appropriate (i.e. $L_1 =1$), and that treatment is withheld (i.e. $a_1=0$) when the time-varying covariate indicates treatment is no longer appropriate (i.e. $L_1 =0$).

We indicate the association between the variable which guides the intervention value and the intervention node with a dotted line since this arrow may or may not exist in the real world (outside our interventions on A), but must exist in the world of our intervention-specific SWIG.

```
SWIG:
```
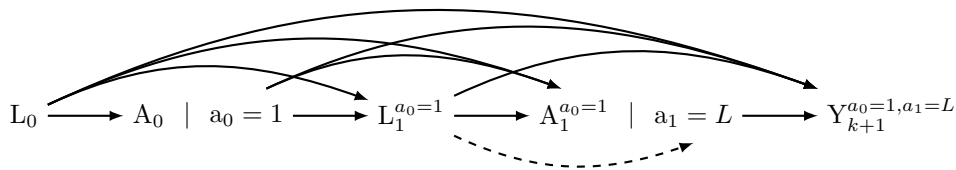


```
Code:

\begin{tikzpicture}[
array/.style={rectangle split,
        rectangle split parts = 3,
        rectangle split horizontal,
    minimum height = 2em
    }
]
```

```
  \node (1) {L$_0$};
   \node [array, right =of 1] (2) {%
          {A$_0$}
     \nodepart{two}{$|$}
     \nodepart{three}{a$_0 =1$}
     };
  \node [right =of 2] (3) {L$_1^{a_{0}=1}$};
    \node [array, right =of 3] (4) {%
          {A$_1^{a_{0}=1}$}
     \nodepart{two}{$|$}
     \nodepart{three}{a$_1 = L$}
     };
  \node [right =of 4] (5) {Y$_{k+1}^{a_0 =1, a_1 = L}$};

\draw[Arrow, thick] (1.east) -- (2.west);
 \draw[Arrow, thick] (1) to [out=25, in=160] (3);
 \draw[Arrow, thick] (1) to [out=25, in=160] (4);
 \draw[Arrow, thick] (1) to [out=25, in=160] (5);

 \draw[Arrow, thick] (2.east) -- (3.west);
 \draw[Arrow, thick] (2) to [out=25, in=160] (4);
 \draw[Arrow, thick] (2) to [out=25, in=160] (5);

 \draw[Arrow, thick] (3.east) -- (4.west);
 \draw[Arrow, dashed] (3) to [out=-25, in =-160] (4.three south);
 \draw[Arrow, thick] (3) to [out=25, in=160] (5);

 \draw[Arrow, thick] (4.east) -- (5.west);

\end{tikzpicture}
```

3.2 SWIG templates

Template SWIGs denote all possible intervention values through the use of a random intervention variable. To create a SWIG template, we construct the same SWIGs as above but no longer label the intervention node nor the counterfactual with a specific intervention value.

Note that, in general, SWIG templates should not be used for assessing potential collider bias or whether a lack identifiability resulting from cross-world problems, since SWIG templates no longer represent only one single world.

3.2.1 SWIG template for interventions on point exposures

The SWIG template for interventions on point exposures represents all possible interventions on the variable of interest. This means that the intervention node is no longer labeled with a specific value as above. Instead, we label with a generic intervention a.

As above, this and all SWIG templates below do not make the assumption that the null hypothesis is true, but SWIG templates can be drawn under this assumption if preferred.

```
SWIG:

L ———→ A  |  a ———→ Y^a

Code:

\begin{tikzpicture}
[
array/.style={rectangle split,
        rectangle split parts = 3,
        rectangle split horizontal,
    minimum height = 2em
    }
]
 \node (1) {L};
 \node [array, right =of 1] (2) {%
        {A}
    \nodepart{two}{$|$}
    \nodepart{three}{a}
    };
 \node [right =of 2] (3) {Y$^{a}$};

 \draw[Arrow, thick] (1.east) -- (2.west);
 \draw[Arrow, thick] (2.two east) -- (3.west);
 \draw[Arrow, thick] (1) to [out=25, in=160] (3);

\end{tikzpicture}
```

3.2.2 SWIG template for static interventions on sustained exposures

The SWIG template for all possible static interventions on sustained exposures operates the same way as the point intervention template. Static interventions are any intervention that does not depend on characteristics of the individual which can change over time. Therefore, the sequence of intervention values is fixed at baseline for every individual.
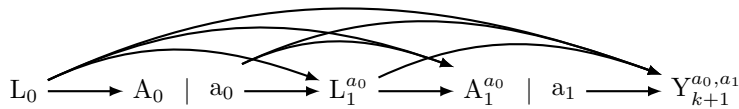
```
SWIG:

L_0 ———→ A_0  |  a_0 ———→ L_1^{a_0} ———→ A_1^{a_0}  |  a_1 ———→ Y_{k+1}^{a_0,a_1}

Code:

\begin{tikzpicture}[
array/.style={rectangle split,
        rectangle split parts = 3,
        rectangle split horizontal,
    minimum height = 2em
```

```
        }
]
  \node (1) {L$_0$};
   \node [array, right =of 1] (2) {%
          {A$_0$}
      \nodepart{two}{$|$}
      \nodepart{three}{a$_0$}
      };
  \node [right =of 2] (3) {L$_1^{a_{0}}$};
    \node [array, right =of 3] (4) {%
          {A$_1^{a_{0}}$}
      \nodepart{two}{$|$}
      \nodepart{three}{a$_1$}
      };
  \node [right =of 4] (5) {Y$_{k+1}^{a_0, a_1}$};

  \draw[Arrow, thick] (1.east) -- (2.west);
  \draw[Arrow, thick] (1) to [out=25, in=160] (3);
  \draw[Arrow, thick] (1) to [out=25, in=160] (4);
  \draw[Arrow, thick] (1) to [out=25, in=160] (5);

  \draw[Arrow, thick] (2.east) -- (3.west);
  \draw[Arrow, thick] (2) to [out=25, in=160] (4);
  \draw[Arrow, thick] (2) to [out=25, in=160] (5);

  \draw[Arrow, thick] (3.east) -- (4.west);
  \draw[Arrow, thick] (3) to [out=25, in=160] (5);

  \draw[Arrow, thick] (4.east) -- (5.west);

\end{tikzpicture}
```

3.2.3   SWIG template for dynamic interventions on sustained exposures

The template SWIG for dynamic interventions on sustained exposures allows the intervention value to depend on time-varying characteristics of the individual, but does not specify that value. We indicate the association between the variable which guides the intervention value and the intervention node with a dotted line.

```
SWIG:
```
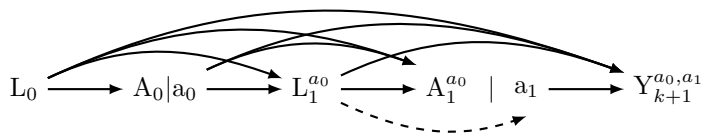


```
Code:

\begin{tikzpicture}[
```

```
array/.style={rectangle split,
        rectangle split parts = 3,
        rectangle split horizontal,
    minimum height = 2em
    }
]
 \node (1) {L$_0$};
 \node [right =of 1] (2) {A$_0 | $a$_0$};
 \node [right =of 2] (3) {L$_1^{a_{0}}$};
 \node [array, right =of 3] (4) {%
        {A$_1^{a_{0}}$}
    \nodepart{two}{$|$}
    \nodepart{three}{a$_1$}
    };
 \node [right =of 4] (5) {Y$_{k+1}^{a_0, a_1}$};

 \draw[Arrow, thick] (1.east) -- (2.west);
 \draw[Arrow, thick] (1) to [out=25, in=160] (3);
 \draw[Arrow, thick] (1) to [out=25, in=160] (4);
 \draw[Arrow, thick] (1) to [out=25, in=160] (5);

 \draw[Arrow, thick] (2.east) -- (3.west);
 \draw[Arrow, thick] (2) to [out=25, in=160] (4);
 \draw[Arrow, thick] (2) to [out=25, in=160] (5);

 \draw[Arrow, thick] (3.east) -- (4.west);
 \draw[Arrow, dashed] (3) to [out=-25, in =-160] (4.three south);
 \draw[Arrow, thick] (3) to [out=25, in=160] (5);

 \draw[Arrow, thick] (4.east) -- (5.west);

\end{tikzpicture}
```

Chapter 4

Probability trees for inverse probability weighting

Probability trees can be useful teaching tools for explaining the use of inverse probability weighting for estimating causal effects. When drawing probability trees for inverse probability weighting we use the following conventions:

- Temporal ordering such that nodes to the left-most side of the tree represent the earliest occurring variables,

- Circular nodes for covariates and outcome variables,

- Tree branches within the circular covariate and outcome nodes to indicate the possible values of these variables, with labels to specify the conditional probability of each value given all prior tree branches,

- Tree branches outside circular nodes to indicate intervention variables (treatment or exposure), labeled with the conditional probability of each intervention variable value given all prior tree branches and the number of individuals expected in each intervention branch,

- Final tally of individuals in each tree leaf at the right-most side of the probability tree.

Using these conventions, the probability tree includes all required information for calculation of inverse probability of treatment weights for simple examples. When inverse probability of censoring weights are desired, branches for censoring can be added in a similar fashion to the intervention nodes.

4.1 Probability tree for point exposures

The example below draws a tree for this scenario for a population of 1,000 individuals, where the intervention probability depends on the baseline confounder only.

Probability tree:

Code:

```
%Step 1: create styles for labeling nodes
\tikzstyle{observed} = [circle, thick, text centered, draw, minimum height=1in,
        minimum width=1in, text width=9em, text height = 16pt]
\tikzstyle{end_text} = [thick, text centered]

%Step 2: draw probability tree
\begin{tikzpicture}
```

```
\node[observed] at (0, 0) (1){};

\draw[black] (1.west) --node[midway, sloped, above]{L=1 (0.4)} (1.north east);
\draw[black] (1.west) --node[midway, sloped, above]{L=0 (0.6)} (1.south east);

\node[observed] at (10, 6) (2){};
\node[observed] at (10, 2) (3){};
\node[observed] at (10,-2) (4){};
\node[observed] at (10,-6) (5){};

\draw[black] (1.north east) -- node[midway, sloped, above]{A=0 (0.1)}
        node[midway, sloped, below]{40} (2.west);
\draw[black] (2.west) --node[midway, sloped, above]{Y=0 (0.75)} (2.north east);
\draw[black] (2.west) --node[midway, sloped, above]{Y=1 (0.25)} (2.south east);
\node[end_text] at (12, 7) {30};
\node[end_text] at (12, 5) {10};

\draw[black] (1.north east) -- node[midway, sloped, above]{A=1 (0.9)}
        node[midway, sloped, below]{360} (3.west);
\draw[black] (3.west) --node[midway, sloped, above]{Y=0 (0.67)} (3.north east);
\draw[black] (3.west) --node[midway, sloped, above]{Y=1 (0.33)} (3.south east);
%Step 2: draw probability tree
\begin{tikzpicture}
\node[observed] at (0, 0) (1){};

\draw[black] (1.west) --node[midway, sloped, above]{L=1 (0.4)} (1.north east);
\draw[black] (1.west) --node[midway, sloped, above]{L=0 (0.6)} (1.south east);

\node[observed] at (10, 6) (2){};
\node[observed] at (10, 2) (3){};
\node[observed] at (10,-2) (4){};
\node[observed] at (10,-6) (5){};

\draw[black] (1.north east) -- node[midway, sloped, above]{A=0 (0.1)}
        node[midway, sloped, below]{40} (2.west);
\draw[black] (2.west) --node[midway, sloped, above]{Y=0 (0.75)} (2.north east);
\draw[black] (2.west) --node[midway, sloped, above]{Y=1 (0.25)} (2.south east);
\node[end_text] at (12, 7) {30};
\node[end_text] at (12, 5) {10};

\draw[black] (1.north east) -- node[midway, sloped, above]{A=1 (0.9)}
        node[midway, sloped, below]{360} (3.west);
\draw[black] (3.west) --node[midway, sloped, above]{Y=0 (0.67)} (3.north east);
\draw[black] (3.west) --node[midway, sloped, above]{Y=1 (0.33)} (3.south east);
\node[end_text] at (12, 3) {240};
\node[end_text] at (12, 1) {120};

\draw[black] (1.south east) -- node[midway, sloped, above]{A=0 (0.5)}
        node[midway, sloped, below]{300} (4.west);
\draw[black] (4.west) --node[midway, sloped, above]{Y=0 (0.9)} (4.north east);
\draw[black] (4.west) --node[midway, sloped, above]{Y=1 (0.1)} (4.south east);
\node[end_text] at (12, -1) {270};
\node[end_text] at (12, -3) {30};
```

```
\draw[black] (1.south east) -- node[midway, sloped, above]{A=0 (0.5)}
        node[midway, sloped, below]{300} (5.west);
\draw[black] (5.west) --node[midway, sloped, above]
        {Y=0 $(\frac{1}{3})$} (5.north east);
\draw[black] (5.west) --node[midway, sloped, above]
        {Y=1 $(\frac{2}{3})$} (5.south east);
\node[end_text] at (12, -5) {100};
\node[end_text] at (12, -7) {200};
\end{tikzpicture}
```
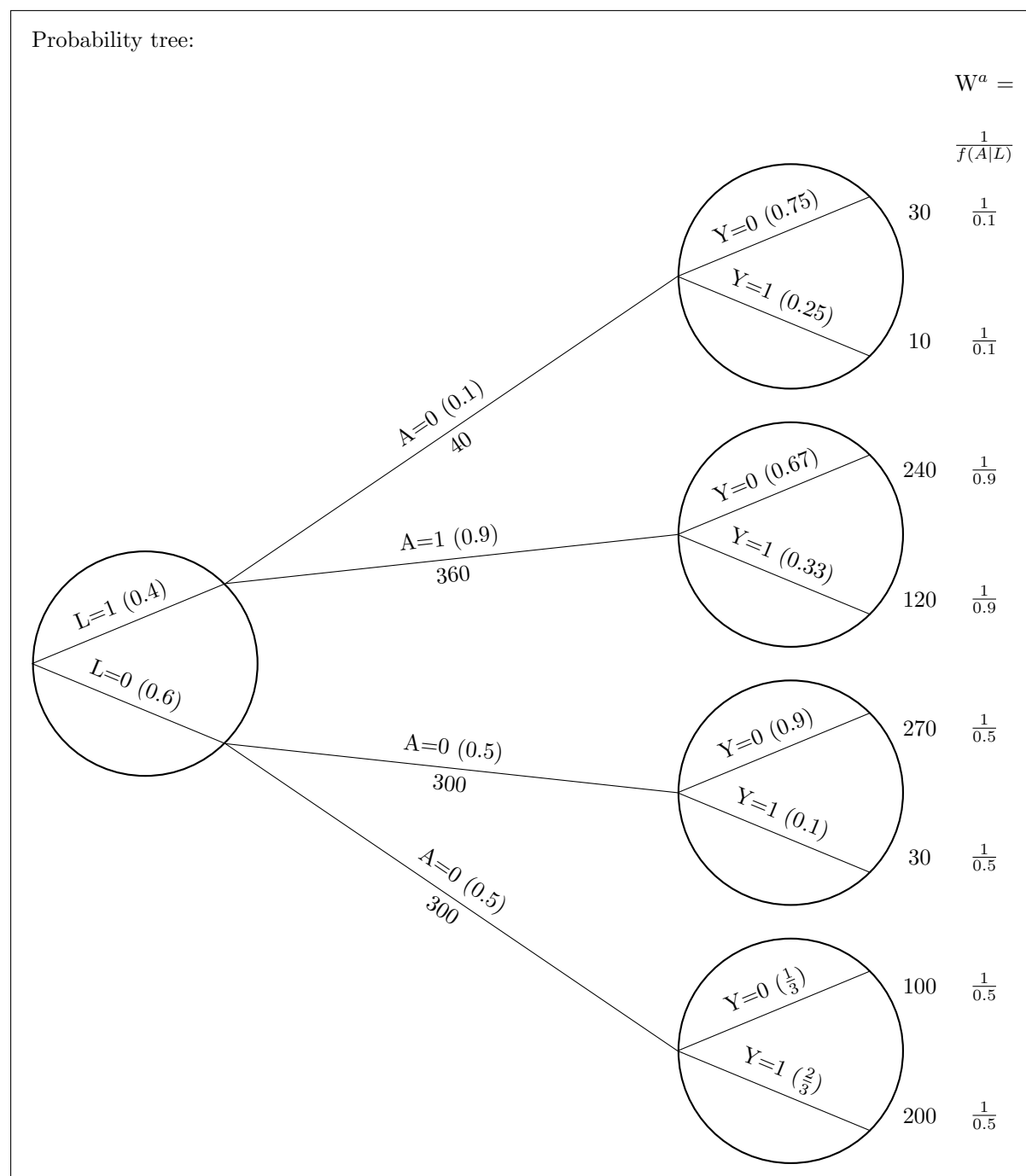
4.1.1 Using probability trees to estimate inverse probability weights for point exposures

The following figure demonstrates the use of probability trees as a didactic tool for estimating inverse probability of treatment weights with a point intervention.

The figure below is identical to the probability tree for a point exposure, except that it now includes a final column displaying the value of the unstabilized inverse probability of treatment weights, calculated as $W^a = \frac{1}{f(A|L)}$. That is, for those individuals who received an intervention value of $A = 1$, their inverse probability of treatment weight is equal to 1 over their conditional probability of having received intervention $A = 1$ given their observed covariate value, L, or $\frac{1}{P(A=1|L)}$.

Probability tree:

$W^a =$

$\frac{1}{f(A|L)}$

Y=0 (0.75)   30 $\frac{1}{0.1}$

Y=1 (0.25)   10 $\frac{1}{0.1}$

A=0 (0.1) 40

A=1 (0.9) 360

Y=0 (0.67)   240 $\frac{1}{0.9}$

Y=1 (0.33)   120 $\frac{1}{0.9}$

L=1 (0.4)

L=0 (0.6)

A=0 (0.5) 300

Y=0 (0.9)   270 $\frac{1}{0.5}$

Y=1 (0.1)   30 $\frac{1}{0.5}$

A=0 (0.5) 300

Y=0 $\left(\frac{1}{3}\right)$ 100 $\frac{1}{0.5}$

Y=1 $\left(\frac{2}{3}\right)$ 200 $\frac{1}{0.5}$

```
Code:

%Step 1: create styles for labeling nodes
\tikzstyle{observed} = [circle, thick, text centered, draw, minimum height=1in,
        minimum width=1in, text width=9em, text height = 16pt]
\tikzstyle{end_text} = [thick, text centered]

%Step 2: draw probability tree
\begin{tikzpicture}
\node[observed] at (0, 0) (1){};
\draw[black] (1.west) --node[midway, sloped, above]{L=1 (0.4)} (1.north east);
\draw[black] (1.west) --node[midway, sloped, above]{L=0 (0.6)} (1.south east);

\node[observed] at (10, 6) (2){};
\node[observed] at (10, 2) (3){};
\node[observed] at (10,-2) (4){};
\node[observed] at (10,-6) (5){};

\draw[black] (1.north east) -- node[midway, sloped, above]{A=0 (0.1)}
        node[midway, sloped, below]{40} (2.west);
\draw[black] (2.west) --node[midway, sloped, above]{Y=0 (0.75)} (2.north east);
\draw[black] (2.west) --node[midway, sloped, above]{Y=1 (0.25)} (2.south east);
\node[end_text] at (12, 7) {30};
\node[end_text] at (12, 5) {10};

\draw[black] (1.north east) -- node[midway, sloped, above]{A=1 (0.9)}
        node[midway, sloped, below]{360} (3.west);
\draw[black] (3.west) --node[midway, sloped, above]{Y=0 (0.67)} (3.north east);
\draw[black] (3.west) --node[midway, sloped, above]{Y=1 (0.33)} (3.south east);
\node[end_text] at (12, 3) {240};
\node[end_text] at (12, 1) {120};

\draw[black] (1.south east) -- node[midway, sloped, above]{A=0 (0.5)}
        node[midway, sloped, below]{300} (4.west);
\draw[black] (4.west) --node[midway, sloped, above]{Y=0 (0.9)} (4.north east);
\draw[black] (4.west) --node[midway, sloped, above]{Y=1 (0.1)} (4.south east);
\node[end_text] at (12, -1) {270};
\node[end_text] at (12, -3) {30};

\draw[black] (1.south east) -- node[midway, sloped, above]{A=0 (0.5)}
        node[midway, sloped, below]{300} (5.west);
\draw[black] (5.west) --node[midway, sloped, above]
        {Y=0 $(\frac{1}{3})$} (5.north east);
\draw[black] (5.west) --node[midway, sloped, above]
        {Y=1 $(\frac{2}{3})$} (5.south east);
\node[end_text] at (12, -5) {100};
\node[end_text] at (12, -7) {200};

\node[end_text] at (13, 9) {W$^a$ = };
\node[end_text] at (13, 8) {$\frac{1}{f(A|L)}$};
\node[end_text] at (13, 7) {$\frac{1}{0.1}$};
\node[end_text] at (13, 5) {$\frac{1}{0.1}$};
\node[end_text] at (13, 3) {$\frac{1}{0.9}$};
```

```
\node[end_text] at (13, 1) {$\frac{1}{0.9}$};
\node[end_text] at (13, -1) {$\frac{1}{0.5}$};
\node[end_text] at (13, -3) {$\frac{1}{0.5}$};
\node[end_text] at (13, -5) {$\frac{1}{0.5}$};
\node[end_text] at (13, -7) {$\frac{1}{0.5}$};

\end{tikzpicture}
```

4.1.2 Probability tree for point exposures with loss to follow-up

When individuals can be lost to follow-up or censored for dropout or non-compliance, we may also be interested in inverse probability of censoring weights. The probability tree below demonstrates the case where censoring occurs after the point intervention but before outcome ascertainment. Censoring is depicted in the tree using an additional set of intervention-style branches.

A similar approach can be used to build probability trees for sustained exposures. However, these trees rapidly become larger than can easily be displayed and their utility as teaching tools is reduced the more complex these trees become.

Probability tree:
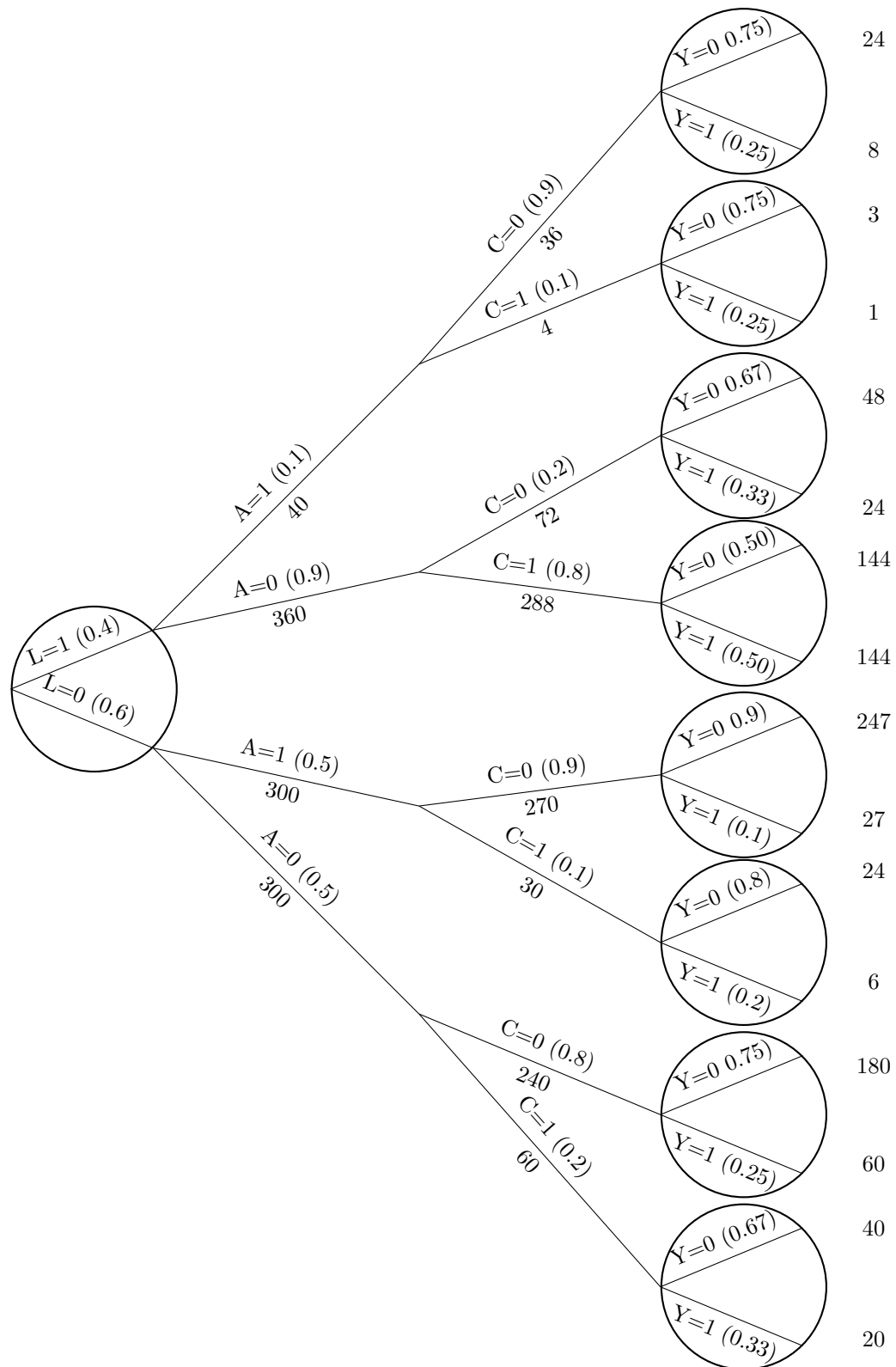
Code:

```
%Step 1: create styles for labeling nodes
\tikzstyle{observed} = [circle, thick, text centered, draw, minimum height=1in,
        minimum width=1in, text height = 8pt]
\tikzstyle{end_text} = [thick, text centered]

%Step 2: draw probability tree
\begin{tikzpicture}
\node[observed] at (0, 0) (1){};

\draw[black] (1.west) --node[midway, sloped, above]{L=1 (0.4)} (1.north east);
\draw[black] (1.west) --node[midway, sloped, above]{L=0 (0.6)} (1.south east);

\draw[black] (1.north east) --node[midway, sloped, above]{A=1 (0.1)}
        node[midway, sloped, below]{40}(5,5);
\draw[black] (1.north east) --node[midway, sloped, above]{A=0 (0.9)}
        node[midway, sloped, below]{360} (5,1.8);
\draw[black] (1.south east) --node[midway, sloped, above]{A=1 (0.5)}
        node[midway, sloped, below]{300} (5,-1.8);
\draw[black] (1.south east) --node[midway, sloped, above]{A=0 (0.5)}
        node[midway, sloped, below]{300} (5,-5);

\node[observed] at (10, 9.20) (2){};
\node[observed] at (10, 6.55) (3){};
\node[observed] at (10, 3.90) (4){};
\node[observed] at (10, 1.32) (5){};

\node[observed] at (10,-1.32) (6){};
\node[observed] at (10,-3.90) (7){};
\node[observed] at (10,-6.55) (8){};
\node[observed] at (10,-9.20) (9){};

\draw[black] (5,5) -- node[midway, sloped, above]{C=0 (0.9)}
        node[midway, sloped, below]{36} (2.west);
\draw[black] (5,5) -- node[midway, sloped, above]{C=1 (0.1)}
        node[midway, sloped, below]{4} (3.west);
\draw[black] (2.west) --node[midway, sloped, above]{Y=0 0.75)} (2.north east);
\draw[black] (2.west) --node[midway, sloped, below]{Y=1 (0.25)} (2.south east);
\node[end_text] at (12, 10) {24};
\node[end_text] at (12, 8.3) {8};
\draw[black] (3.west) --node[midway, sloped, above]{Y=0 (0.75)} (3.north east);
\draw[black] (3.west) --node[midway, sloped, below]{Y=1 (0.25)} (3.south east);
\node[end_text] at (12, 7.3) {3};
\node[end_text] at (12, 5.8) {1};

\draw[black] (5, 1.8) -- node[midway, sloped, above]{C=0 (0.2)}
        node[midway, sloped, below]{72} (4.west);
\draw[black] (5, 1.8) -- node[midway, sloped, above]{C=1 (0.8)}
        node[midway, sloped, below]{288} (5.west);
\draw[black] (4.west) --node[midway, sloped, above]{Y=0 0.67)} (4.north east);
\draw[black] (4.west) --node[midway, sloped, below]{Y=1 (0.33)} (4.south east);
\node[end_text] at (12, 4.5) {48};
\node[end_text] at (12, 2.8) {24};
\draw[black] (5.west) --node[midway, sloped, above]{Y=0 (0.50)} (5.north east);
```

```
\draw[black] (5.west) --node[midway, sloped, below]{Y=1 (0.50)} (5.south east);
\node[end_text] at (12, 2.0) {144};
\node[end_text] at (12, 0.5) {144};

\draw[black] (5,-1.8) -- node[midway, sloped, above]{C=0 (0.9)}
        node[midway, sloped, below]{270} (6.west);
\draw[black] (5,-1.8) -- node[midway, sloped, above]{C=1 (0.1)}
        node[midway, sloped, below]{30} (7.west);
\draw[black] (6.west) --node[midway, sloped, above]{Y=0 0.9)} (6.north east);
\draw[black] (6.west) --node[midway, sloped, below]{Y=1 (0.1)} (6.south east);
\node[end_text] at (12, -0.5) {247};
\node[end_text] at (12, -2.0) {27};
\draw[black] (7.west) --node[midway, sloped, above]{Y=0 (0.8)} (7.north east);
\draw[black] (7.west) --node[midway, sloped, below]{Y=1 (0.2)} (7.south east);
\node[end_text] at (12, -2.8) {24};
\node[end_text] at (12, -4.5) {6};

\draw[black] (5,-5) -- node[midway, sloped, above]{C=0 (0.8)}
        node[midway, sloped, below]{240} (8.west);
\draw[black] (5,-5) -- node[midway, sloped, above]{C=1 (0.2)}
        node[midway, sloped, below]{60} (9.west);
\draw[black] (8.west) --node[midway, sloped, above]{Y=0 0.75)} (8.north east);
\draw[black] (8.west) --node[midway, sloped, below]{Y=1 (0.25)} (8.south east);
\node[end_text] at (12, -5.8) {180};
\node[end_text] at (12, -7.3) {60};
\draw[black] (9.west) --node[midway, sloped, above]{Y=0 (0.67)} (9.north east);
\draw[black] (9.west) --node[midway, sloped, below]{Y=1 (0.33)} (9.south east);
\node[end_text] at (12, -8.3) {40};
\node[end_text] at (12, -10) {20};
\end{tikzpicture}
```

Chapter 5

Decision trees for Markov processes

Decision trees are widely used in health policy and comparative effectiveness research to both visualize the decision making process and to describe Markov models. Markov models can be individual-level simulation models or compartmental models, and are tools for assessing counterfactual outcomes under a set of hypotheses and assumptions about the world.

The trees begin with the potential states into which each individual or type of individual must be categorized at the start of every time-step. Typically, at least one of these states is a sink – i.e. a state from which individuals cannot leave. This state has as a descendant only an end node (typically denoted with a triangle). All other states have as descendants a series of one or more decision nodes. Some of these nodes represent decisions based on the intervention of interest and are indicated by a black box at the decision point. Other nodes represent characteristics that are updated based on a (conditional) probability distribution and are represented by hollow circles at the decision point.
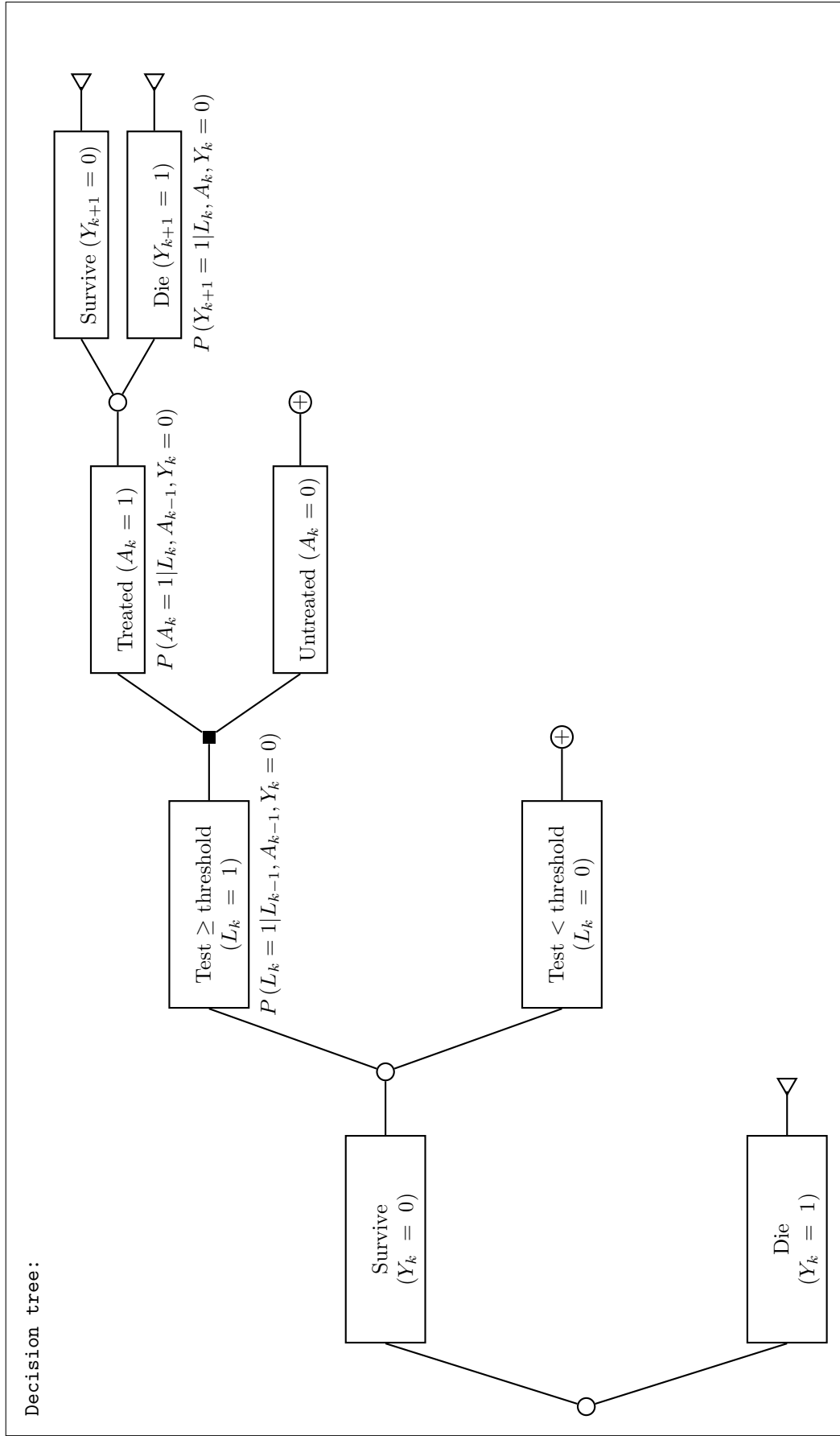
In the decision tree, at any time a series of descendants can be replaced by a 'copy node' (a + symbol inscribed in a circle) to denote that all descendants of this node follow the same series as the descendants of its sibling nodes.

5.1 Decision tree for a Markov model of dynamic interventions on sustained exposures

The decision tree in this section depicts a decision tree for a Markov model when treatment decisions are deterministic and based on some stochastic covariate test, and the potential health states of interest are 'alive' and 'dead'.

In this scenario, individuals who have a covariate score of above a treatment threshold will have a higher probability of receiving the intervention than individuals with a covariate score below the threshold. This represents the same basic scenario depicted by the statistical and causal DAGs for sustained exposures and the intervention-specific SWIG and SWIG template for dynamic interventions on sustained exposures. However, we have omitted the unknown common cause of L and Y because decision trees for Markov models generally include only known variables.

Decision tree:

Survive ($Y_{k+1} = 0$)

Die ($Y_{k+1} = 1$)

$P(Y_{k+1} = 1 | L_k, A_k, Y_k = 0)$

Treated ($A_k = 1$)

$P(A_k = 1 | L_k, A_{k-1}, Y_k = 0)$

Untreated ($A_k = 0$)

Test $\geq$ threshold
($L_k = 1$)

$P(L_k = 1 | L_{k-1}, A_{k-1}, Y_k = 0)$

Test $<$ threshold
($L_k = 0$)

Survive
($Y_k = 0$)

Die
($Y_k = 1$)

```
Code:

%Step 1: create styles for labeling nodes
\tikzstyle{startnode} = [circle, thick, draw = black, text centered,
                minimum height=8pt, minimum width=8pt, fill=white,
                inner sep=0pt, xshift = -1.3cm]
\tikzstyle{choice} = [regular polygon, regular polygon sides = 4, thick,
                minimum width=8pt, fill, inner sep=0pt]
\tikzstyle{textbox} = [rectangle, thick, text width=9em, text height = 16pt,
                text centered, draw]
\tikzstyle{fixedchoice} = [circle, thick,  draw = black, minimum height=8pt,
                minimum width=8pt, fill=white, inner sep=0pt]
\tikzstyle{copynode}=[circle, thick, draw, text centered,
                minimum width=8pt, inner sep=0pt]
\tikzstyle{endnode} = [regular polygon,  draw = black,thick,
                regular polygon sides=3, minimum width=10pt,
                fill=white, inner sep=0pt,shape border rotate=90,
                outer sep=-\pgflinewidth, xshift = -0.2cm]
\tikzstyle{endtext}=[thick, text centered]

%Step 2: draw decision tree -- note tree builds from left to right and bottom to top

\begin{tikzpicture}[
grow=right,child anchor=west,
level 1/.style={sibling distance=6.6cm},level 2/.style={sibling distance=5.8cm},
level 3/.style={sibling distance=5.8cm},level 4/.style={sibling distance=5.8cm},
level 5/.style={sibling distance=3.0cm},level 6/.style={sibling distance=3.0cm},
level 7/.style={sibling distance=1.2cm},level 8/.style={sibling distance=1.2cm},
level 9/.style={sibling distance=1.2cm},
level distance = 2.75cm, thick]

\node[startnode]{}
child{node[textbox]{Die \\($Y_{k}=1$)}
  child{node[endnode]{}}
}
child{node[textbox]{Survive \\(${Y}_{k}=0$)}
 child{node[fixedchoice]{}
  child {node[textbox]{Test $<$ threshold \\ ($L_{k}=0$)}
   child{node[copynode]{+}}
  }
  child{node[textbox](1){Test $\geq$ threshold  \\  ($L_{k}=1$)}
   child{node[choice]{}
    child{node[textbox]{Untreated    ($A_{k}=0$)}
     child{node[copynode]{+}}
    }
    child{node[textbox](2){Treated    ($A_{k}=1$)}
     child{node[fixedchoice]{}
      child{node[textbox](3){Die   ($Y_{k+1}=1$)}
       child{node[endnode]{}}
      }
      child{node[textbox]{Survive ($Y_{k+1}=0$)}
       child{node[endnode]{}}
      }
```

```
        }
      }
    }
  }
}
}
;
%optional code to label conditional probabilities associated with nodes;
\node [below=0.7cm, align=right, thick, xshift = 0.5cm] at (1)
        {$P\left(L_{k}=1|L_{k-1},A_{k-1}, Y_{k}=0\right)$};
\node [below=0.5cm, align=right, thick, xshift = 0.5cm] at (2)
        {$P\left(A_{k}=1|L_{k},A_{k-1}, Y_{k}=0\right)$};
\node [below=0.5cm, align=right, thick, xshift = 0.2cm] at (3)
        {$P\left(Y_{k+1}=1|L_{k},A_{k}, Y_{k}=0\right)$};
\end{tikzpicture}
```