Roadmap for Targeted Learning

1. **Data**: realizations of random variables with a probability distribution.
2. **Statistical Model**: actual knowledge about the shape of the data-generating probability distribution.
3. **Statistical Target Parameter**: a feature/function of the data-generating probability distribution.
4. **Estimator**: an a priori-specified algorithm, benchmarked by a dissimilarity-measure (e.g., MSE) w.r.t. target parameter.

Summary – Data

1. Data are generated as i.i.d. draws from some probability distribution P_0.
2. Statistical inference means drawing some conclusion based on a sample from P_0.
3. To make inference meaningful, we need to understand how the data were sampled.

Statistical Models

A statistical model $\mathcal{M}$ is the set of all probability distributions that we think might have generated the data.

- The choice of model is a reflection of scientific knowledge, not statistical convenience.
- This ensures that $P_0 \in \mathcal{M}$.

Suppose all that is known about P_0 is that the data are n i.i.d. copies of O. This defines our model: the set of all possible probability distributions for O.

- This is called a nonparametric model.

Statistical Models

Suppose $O = (A, Y) \sim P_0$, where $A \in \{0, 1\}$ is a binary treatment and $Y \in \{0, 1\}$ is a binary outcome, $O \in \mathcal{O} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$.

Suppose our model $\mathcal{M}$ is nonparametric. There are four possible values for an observation. Each observation has some (unknown) probability of being observed.

Each distribution in our model, say $P \in \mathcal{M}$, could be parameterized by probabilities, $P(O = (0, 0)), P(O = (0, 1)), P(O = (1, 0))$.

- $\mathcal{M}$ consists of all choices of these three probabilities whose sum is less than or equal to one.

Statistical Models

We could also parameterize the model using probabilities $P(A = 1), P(Y = 1|A = 1), P(Y = 1|A = 0)$, where

$$P(A = 1) = P(A = 1, Y = 0) + P(A = 1, Y = 1)$$

$$P(Y = 1|A = 1) = \frac{P(Y = 1, A = 1)}{P(A = 1)}$$

$$P(Y = 1|A = 0) = \frac{P(Y = 1, A = 0)}{1 - P(A = 1)}$$

Notice that these quantities are variationally independent – the value of one does not restrict what the others can be.

Statistical Models

Suppose A is a new drug we are evaluating and Y is a safety outcome. We have surveillance data that makes us confident that

$$P_0(Y = 1|A = 0) \in (0, u)$$

for some small number u.

Our model still makes no assumptions about $P(A = 1)$ and $P(Y = 1|A = 1)$; however, we have now excluded from the model all values of $P(Y = 1|A = 0)$ that fall outside $(0, u)$.

We refer to this smaller model as a semiparametric model.

- Not the standard definition of semiparametric model that you may be used to hearing.

Statistical Models

Suppose $O = (A, Y) \sim P_0$, where $A \in \{0, 1\}$ is a binary treatment and $Y \in \mathbb{N}$ is an integer-valued outcome.

A nonparametric model could be characterized by $P(O = o)$ as with the binary Y.

We could also parameterize the model using $P(A = 0)$, $P(Y|A = 0)$, and $P(Y|A = 1)$, where

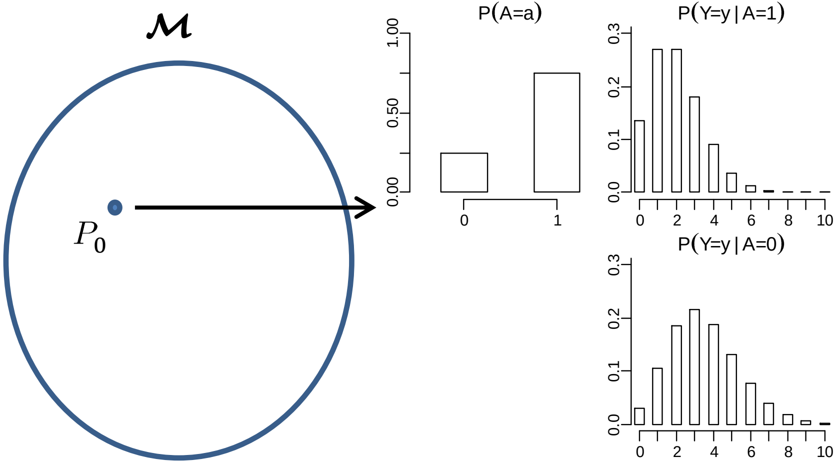
$$P(A = 1) = \sum_{y=0}^{\infty} P(A = 1, Y = y)$$

$$P(Y = y|A = 1) = \frac{P(Y = y, A = 1)}{P(A = 1)}$$

$$P(Y = y|A = 0) = \frac{P(Y = y, A = 0)}{1 - P(A = 1)}$$

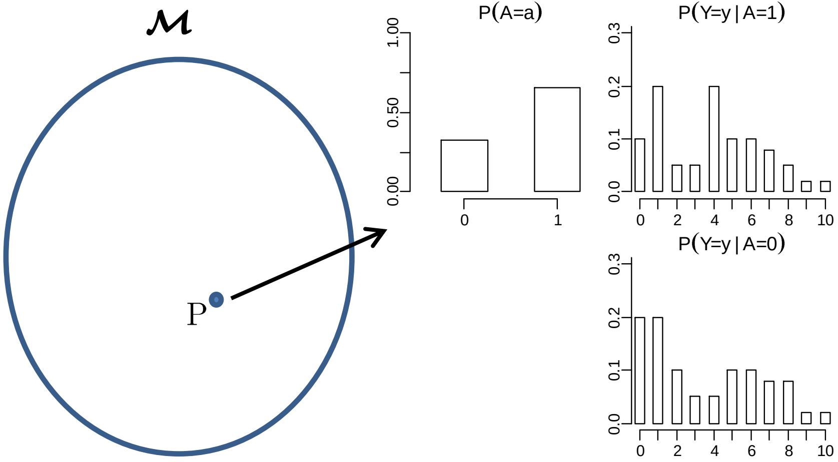
Statistical Model

An illustration of a model for $O = (A, Y)$:



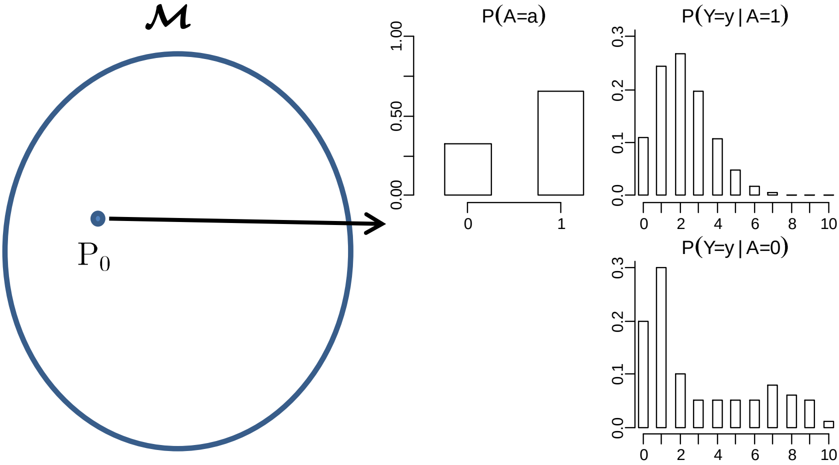
Statistical Model

Each point represents a distribution for O:



Statistical Model

The true distribution is in $\mathcal{M}$:



Statistical Model

Suppose $O = (W, A, Y) \sim P_0$, where $W \in \mathbb{R}$ is a univariate covariate, $A \in \{0, 1\}$ is a binary treatment and $Y \in \{0, 1\}$ is a binary outcome.

The nonparametric model could be characterized by P_W, $P(A = 1|W)$, $P(Y = 1|A = 1, W)$, and $P(Y = 1|A = 0, W)$.

Notation: We often write $P_W(w) = P(W \leq w)$ to denote the distribution of covariates. To average some function f of W over P_W, we will write

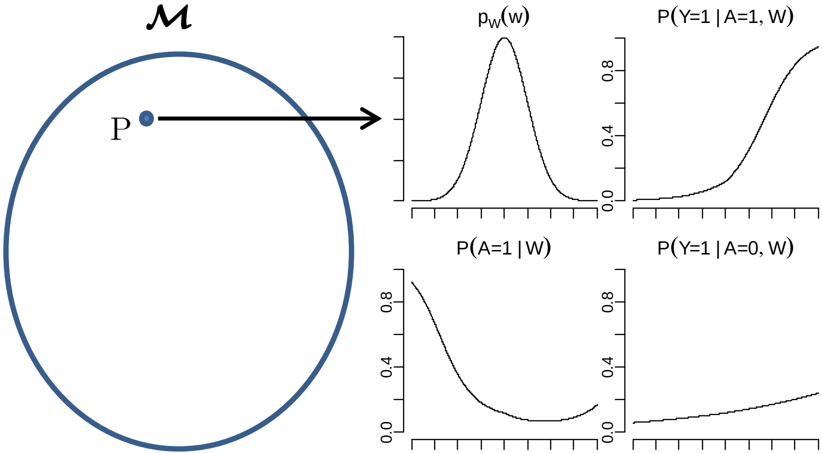
$$E_P\{f(W)\} = \int f(w) dP_W(w) \ .$$

If P_W has a density function p_W with respect to Lebesgue measure, we can write the more familiar Riemann integral form

$$E_P\{f(W)\} = \int f(w) p_W(w) dw \ .$$

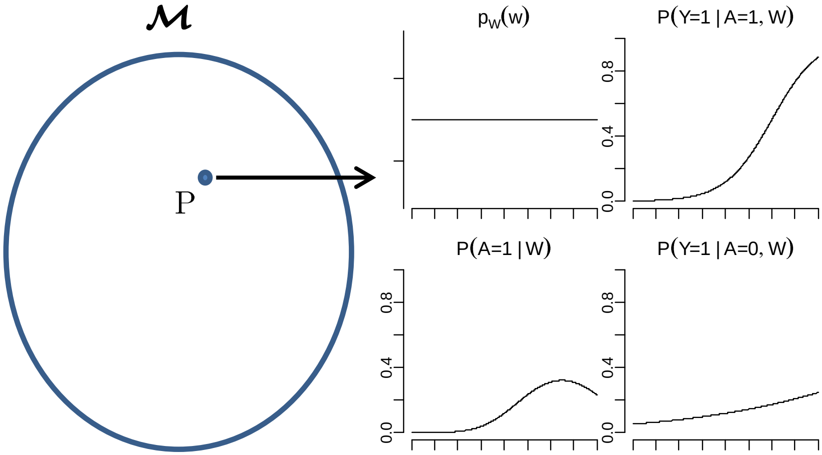
Statistical Model

An illustration of a model for $O = (W, A, Y)$:



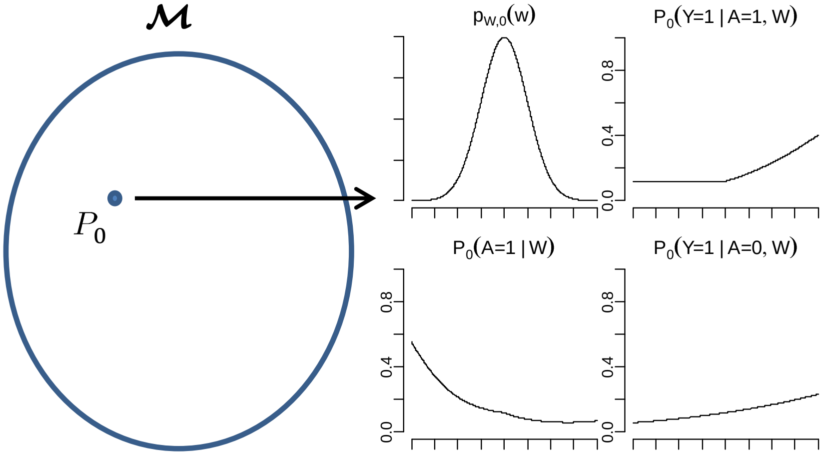
Statistical Model

Each point represents a distribution for O:



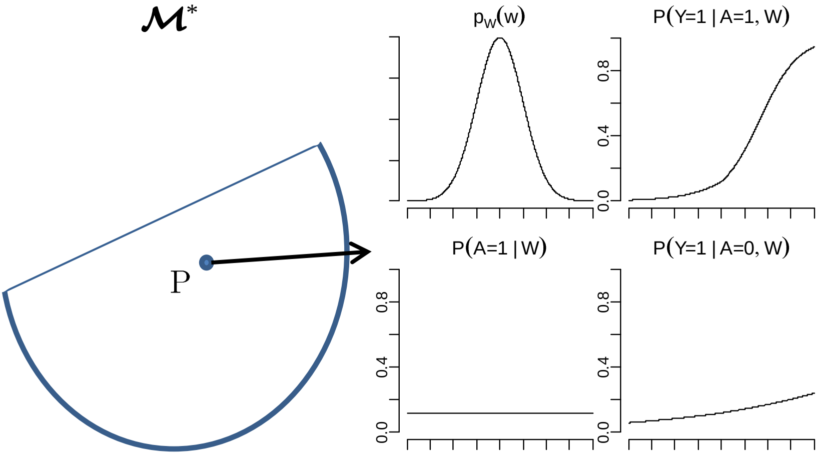
Statistical Model

The true distribution is in $\mathcal{M}$:



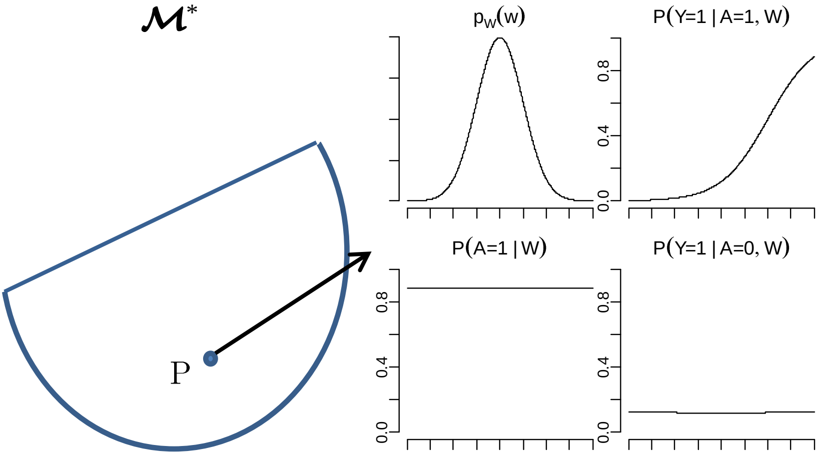
Statistical Model

We can restrict $\mathcal{M}$, e.g., $P(A|W) = P(A)$:



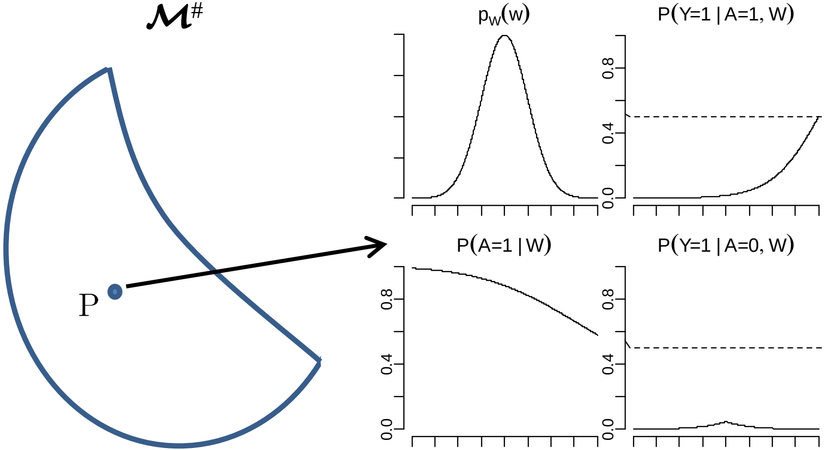
Statistical Model

We still allow $P(Y|A, W)$ and $P(W)$ to vary arbitrarily:



Statistical Model

We can restrict $\mathcal{M}$, e.g., $P(Y|A = a, W = w) < u$ for all (a, w):



Statistical Target Parameter

We refer to the feature/function of the data-generating probability distribution that answers our question of scientific interest the statistical target parameter.

Formally, we define a parameters as a function $\Psi : \mathcal{M} \rightarrow \mathbf{\Psi}$, and we refer to $\mathbf{\Psi}$ as the parameter space.

- You give me a P and I give you back e.g., a real number

The statistical target parameter is not defined in reference to a particular (e.g., parametric) model. It should be well-defined/interpretable for any distribution in our model.

Statistical Target Parameter

Suppose $O = (A, Y) \sim P_0$, where $A \in \{0, 1\}$ is a binary treatment and $Y \in \{0, 1\}$ is a binary outcome.

Example parameters could include univariate summary measures:
Mean of $Y = P(Y = 1|A = 1)P(A = 1) + P(Y = 1|A = 0)P(A = 0)$

Example parameters could include comparisons across levels of A:

$$\text{Risk difference} = P(Y = 1|A = 1) - P(Y = 1|A = 0)$$
$$\text{Risk ratio} = \frac{P(Y = 1|A = 1)}{P(Y = 1|A = 0)}$$
$$\text{Odds ratio} = \frac{P(Y = 1|A = 1)/P(Y = 0|A = 1)}{P(Y = 1|A = 1)/P(Y = 0|A = 1)}$$

Statistical Target Parameter

Suppose $O = (A, Y) \sim P_0$, where $A \in \{0, 1\}$ is a binary treatment and $Y \in \mathbb{N}$ is an integer-valued outcome.

Univariate summary measures: The average value of Y, which we write as

$$\begin{aligned}\Psi_m(P) &= E_P(Y)\\ &= \sum_{a=0}^{1}\sum_{y=1}^{\infty} yP(Y = y, A = a)\ ,\end{aligned}$$

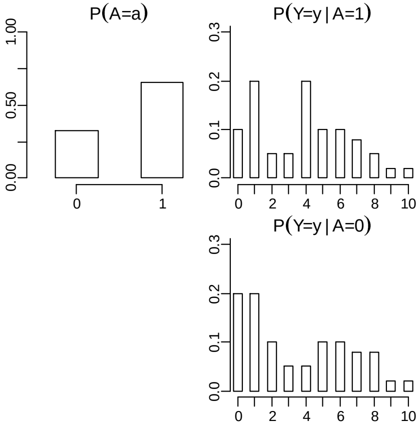
or equivalently as

$$\begin{aligned}\Psi_m(P) &= E_P\{E_P(Y|A)\}\\ &= \sum_{a=0}^{1}\sum_{y=1}^{\infty} yP(Y = y|A = a)P(A = a)\ ,\end{aligned}$$

Statistical Target Parameter

Take a P from $\mathcal{M}$, apply Ψ, get back number.

P $\qquad\qquad$ $\Psi(P)$



$$\rightarrow \sum_a \sum_y y\, P(Y=y|A=a)\, P(A=a)$$

Statistical Target Parameter

Suppose $O = (A, Y) \sim P_0$, where $A \in \{0, 1\}$ is a binary treatment and $Y \in \mathbb{N}$ is an integer-valued outcome.

Example parameters could include comparisons across levels of A:

$$\Psi_{dm}(P) = E_P(Y|A=1) - E_P(Y|A=0)$$

$$\Psi_{rm}(P) = \frac{E_P(Y|A=1)}{E_P(Y|A=0)}$$

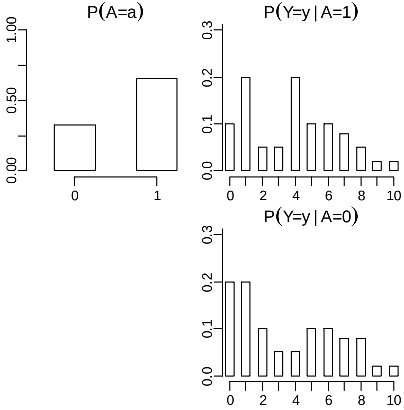
$$\Psi_{rgm}(P) = \frac{\exp[E_P\{\log(Y)|A=1\}]}{\exp[E_P\{\log(Y)|A=0\}]}$$

How do these parameters depend on P compared to Ψ_m?

Statistical Target Parameter

Take a P from $\mathcal{M}$, apply Ψ, get back number.

P $\qquad\qquad\qquad$ $\Psi(P)$



$$\rightarrow \sum_y y\, P(Y|A=1) - \sum_y y\, P(Y|A=0)$$

Statistical Target Parameter

Suppose $O = (W, A, Y) \sim P_0$, where $W \in \mathbb{R}$ is a univariate covariate, $A \in \{0, 1\}$ is a binary treatment and $Y \in \{0, 1\}$ is a binary outcome.

Univariate summary measures:
The average value of Y,

$$\begin{aligned}\Psi_m(P) &= E_P\{E_P(Y|A)\}\\ &= \int \sum_{a=0}^{1} \sum_{y=0}^{1} yP(Y = y|A = a, W = w)P(A = a|W = w)dP_W(w) \ .\end{aligned}$$

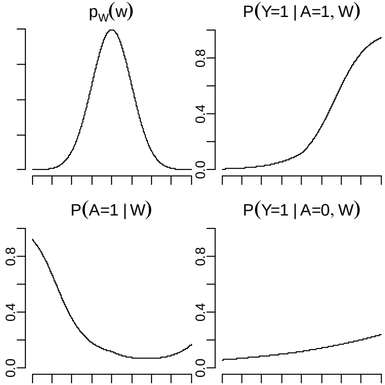
The average value of A,

$$\begin{aligned}\Psi_{ma}(P) &= E_P\{A|W\}\\ &= \int P(A = 1|W = w)dP_W(w)\end{aligned}$$

Statistical Target Parameters

Take a P from $\mathcal{M}$, apply Ψ, get back number.

P $\qquad\qquad$ $\Psi(P)$



$$\rightarrow \int \sum_a \{P(Y=1 | A=a, W=W) \; P(A=a | W=w) \, dP_W(w)\}$$

Statistical Target Parameters

Suppose $O = (W, A, Y) \sim P_0$, where $W \in \mathbb{R}$ is a univariate covariate, $A \in \{0, 1\}$ is a binary treatment and $Y \in \{0, 1\}$ is a binary outcome.

Comparisons across values of A:

$$\Psi_{md,w}(P) = P(Y = 1|A = 1, W = w) - P(Y = 0|A = 0, W = w)$$
$$\Psi_{amd} = E_P\{P(Y = 1|A = 1, W) - P(Y = 0|A = 0, W)\}$$

Comparisons across values of A in standardized population:

$$\begin{aligned}\Psi_{smd} &= E_{\tilde{P}}\{P(Y = 1|A = 1, W) - P(Y = 0|A = 0, W)\} \\ &= \int \{P(Y = 1|A = 1, W = w) - P(Y = 0|A = 0, W = w)\} d\tilde{P}_W(w)\end{aligned}$$

Statistical Target Parameters

Suppose $O = (W, A, Y) \sim P_0$, where $W \in \mathbb{R}$ is a univariate covariate, $A \in \{0, 1\}$ is a binary treatment and $Y \in \{0, 1\}$ is a binary outcome.

Comparisons across values of A and W:

$$\begin{aligned}\Psi_{em,md}(P) = & \, E_P\{P(Y = 1|A = 1, W) - P(Y = 0|A = 0, W)|W > 0\} \\ & - E_P\{P(Y = 1|A = 1, W) - P(Y = 0|A = 0, W)|W \leq 0\}\end{aligned}$$

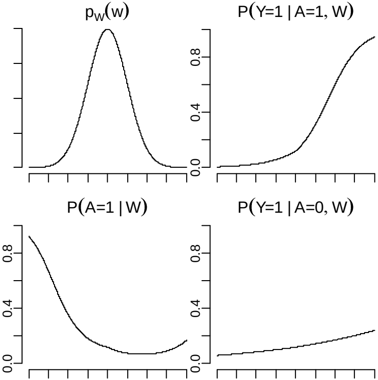
The best prediction function based on A and W:

$$\Psi_p(P) = \underset{\psi \in \boldsymbol{\Psi}}{\text{argmin}} E_P[\{Y - \psi(A, W)\}^2]$$

Statistical Target Parameters

Take a P from $\mathcal{M}$, apply Ψ, get back number.

P



$\Psi_{amd}(P)$

$$\rightarrow \int \{P(Y = 1 | A = 1, W = w) - P(Y = 1 \mid A = 0, W = w)\} \, dP_W(w)$$

Statistical Target Parameters

We can use parametric models to inspire parameters, while ensuring parameters are well-defined everywhere in our model.

For example, suppose $O = (W, A, Y)$, with $W \in \mathbb{R}$, $A \in \mathbb{R}$, and $Y \in \mathbb{R}$.

A linear regression model might motivate interest in the parameter

$$\begin{aligned}
&\Psi_{msm}(P) \\
&= \underset{\beta}{\operatorname{argmin}} \int \Big[E_P\{E_P(Y|A=a, W)\} - (\beta_0 + \beta_1 a) \Big]^2 dP_A(a)
\end{aligned}$$

Statistical Target Parameters

We can also use structural causal models (SCMs) to inspire interesting parameters.

Even if we do not believe the assumptions made by such models, the statistical parameters inspired by these models may still be of interest.

More on SCMs and enriching the interpretation of statistical parameters next time.

Summary: Statistical Models and Target Parameters

1. A statistical model is a collection of probability distributions that contains the true distribution P_0.
2. We will use nonparametric model to mean the model that assumes only i.i.d. observations.
3. We will use semiparametric model to mean any model that assumes more.
4. Parameters are defined as some function of a probability distribution.
5. Parameters should be well-defined for every distribution in the statistical model.
6. Parameters should reflect, to the greatest extent possible, the scientific question of interest.