Section 6

Lecture 2

We make decisions based on "what if" questions...

Some repetition from the first lecture.

- Would starting treatment A prevent a heart attack?
- Is Drug A better than Drug B?
- Would the election campaign increase the number of votes?
- Would university education increase my future earnings?
- What would happen if I went to UNIGE instead of EPFL?

Section 7

More on the definition of causal effects

Effect contrasts

- Additive effect: $\mathbb{E}(Y^{a=1}) - \mathbb{E}(Y^{a=0}) = \mathbb{E}(Y^{a=1} - Y^{a=0})$.
  The additive effect is an average over individual level causal effects. These are marginal quantities.
- Relative effect: $\frac{\mathbb{E}(Y^{a=1})}{\mathbb{E}(Y^{a=0})} \neq \mathbb{E}\left(\frac{Y^{a=1}}{Y^{a=0}}\right)$.
  The relative effect is not an average over individual level causal effects.

Counterfactuals a.k.a. potential outcomes

- We will posit unobserved fixed potential or counterfactual outcomes[3] for each unit[4] under different treatments.[5]

  Hint: It is helpful to think about a counterfactual random variable as a variable that **does exist** in this world, even before interventions take place, but we are not able to observe it.

- We will use superscripts to indicate that a random variable is counterfactual. For example consider a random variable Y. A counterfactual version Y^g is the value Y would have had under an intervention g (also called treatment regime or treatment strategy).
- To get started, in the first lectures, we will consider some simple interventions g which only fixes a binary treatment A to a value $a \in \{0, 1\}$.

---

[3] I will use the terms "counterfactuals" and "potential outcomes". interchangeably.

[4] I will use the terms "unit", "subject" and "individual" interchangeably.

[5] I will use the terms "treatment" and "exposure" interchangeably.

Illustrative experiment (trial) on heart transplant.

Assess the effect of $A \in \{0,1\}$ (1 if heart transplant, 0 otherwise) on $Y \in \{0,1\}$ (1 if dead, 0 otherwise)[6].

	A	Y	Y^0	Y^1
Rheia	0	0	0	?
Kronos	0	1	1	?
Demeter	0	0	0	?
Hades	0	0	0	?
Hestia	1	0	?	0
Poseidon	1	0	?	0
Hera	1	0	?	0
Zeus	1	1	?	1
Artemis	0	1	1	?
Apollo	0	1	1	?
Leto	0	0	0	?
Ares	1	1	?	1
Athena	1	1	?	1
Hephaestus	1	1	?	1
Aphrodite	1	1	?	1
Cyclope	1	1	?	1
Persephone	1	1	?	1
Hermes	1	0	?	0
Hebe	1	0	?	0
Dionysus	1	0	?	0

[6] Miguel A Hernan and James M Robins. *Causal inference: What if?* CRC Boca

Remark on counterfactuals

We have presupposed that:

- $Y^a = Y$ for every unit with $A = a$. In other words, $Y^A = Y$. "Consistency". That is, $Y = I(A = 0)Y^{a=0} + I(A = 1)Y^{a=1}$.

  This "consistency" assumption requires that
  - The intervention on A is well-defined.
    No matter how unit i received treatment a, the outcome Y^a is the same.
  - The counterfactual outcome of unit i does not depend on the treatment values of other units j, that is, "no interference". Otherwise Y_i^a is not well-defined.[7]

[7] This use of consistency is different from the use in estimation.

More on consistency

An example of an ill-defined intervention:
Imagine A is a person's body mass index (BMI). Setting the BMI to a counterfactually different level can happen in many different ways - losing weight by running, loss of appetite due to chain smoking, liposuction etc. Depending on what way the intervention is implemented each time, we will have very different health outcomes, i.e., re-running the experiment will give inconsistent results.
Another example is infectious diseases

Average causal effect

Definition (Average causal effect)

$\mathbb{E}(Y^{a=0})$ vs $\mathbb{E}(Y^{a=1})$.

- Average causal effects can sometimes be identified from data (we will study this extensively).
- In most of this course, average causal effects will be our parameters of interest, i.e. our target estimands.

Effect contrasts

- Additive effect: $\mathbb{E}(Y^{a=1}) - \mathbb{E}(Y^{a=0}) = \mathbb{E}(Y^{a=1} - Y^{a=0})$.
  The additive effect is an average over individual level causal effects. These are marginal quantities.
- Relative effect: $\frac{\mathbb{E}(Y^{a=1})}{\mathbb{E}(Y^{a=0})} \neq \mathbb{E}\left(\frac{Y^{a=1}}{Y^{a=0}}\right)$.
  The relative effect is not an average over individual level causal effects.

Causal effects in the population

More generally, we can consider population causal effects[8]:

Definition (Population causal effect)

A population causal effect can be defined as a contrast of any functional of the marginal distributions of counterfactual outcomes under different interventions.

- For example $\text{VAR}(Y^{a=1}) - \text{VAR}(Y^{a=0})$.
  Remember that we cannot identify $\text{VAR}(Y^{a=1} - Y^{a=0})$.
- From here on, I will often say *causal effect* when I talk about *average causal effect*.

[8] Hernan and Robins, *Causal inference: What if?*

The task of identification

Definition (Identification)

A parameter is said to be **identified** under a particular collection of assumptions if it can be expressed uniquely as a function of the distribution (law) of the observed variables.

That is, a parameter (*estimand*) is identified under a particular collection of assumptions if these assumptions imply that the distribution of the observed data is compatible with a single value of the parameter.

A more formal argument why randomisation is the gold standard

In the previous lecture, I claimed that

- In a randomised experiment, the treatment is assigned independently of all other factors (e.g. by a coin flip or a random number generator).
- In a randomised experiment one of the counterfactual outcomes $Y^{a=0}$ or $Y^{a=1}$ is unobserved.
- However, randomisation ensures that it is *random* whether $Y^{a=0}$ or $Y^{a=1}$ is unobserved, that is,

  $$P(Y^a = y \mid A = 1) = P(Y^a = y \mid A = 0), \forall a \in \{0, 1\}, \forall y \in \mathcal{Y}.$$

  because the treatment assignment is independent of all other factors, including the counterfactual outcomes (Y^a). This conditional independence is called **exchangeability** and will be important in our identification arguments.

Independence notation

Definition (Conditional independence)

$X \perp\!\!\!\perp Y \mid Z \iff F_{X,Y|Z=z}(x,y) = F_{X|Z=z}(x) \cdot F_{Y|Z=z}(y) \; \forall \; x, y, z,$
where $F_{X,Y|Z=z}(x,y) = P(X \leq x, Y \leq y \mid Z = z)$.

We say that X and Y are conditionally independent given Z.
In other words, when $Z = z$ is known, X provides no additional information that allows us to *predict* Y.

Exchangeability (re-visited)

In particular, we can re-write the condition from Slide 59,

$$P(Y^a = y \mid A = 1) = P(Y^a = y \mid A = 0), \forall a \in \{0, 1\}, \forall y \in \mathcal{Y},$$

as

$$Y^a \perp\!\!\!\perp A, \forall a \in \{0, 1\}.$$

Example conditions that ensure identification of causal effects

Suppose that the following 3 conditions hold:

1. $Y^a \perp\!\!\!\perp A, \forall a \in \{0,1\}$ (exchangeability[9]).
2. $P(A = a) > 0 \ \forall a \in \{0,1\}$ (positivity[10]).
3. $Y^a = Y$ for every unit with $A = a$ (consistency[11]).
   that is, $Y = I(A = 0)Y^{a=0} + I(A = 1)Y^{a=1}$.

From (1)-(3), $\mathbb{E}(Y^a) = \mathbb{E}(Y \mid A = a)$.
That is, we have *identified* $\mathbb{E}(Y^a)$ as a functional of observed data.
Assumptions (1)-(3) are external to the data, but – importantly – they hold by design in a perfectly executed experiment.
Just to be clear: The counterfactual independence $Y^a \perp\!\!\!\perp A, \forall a \in \{0,1\}$ does NOT imply the factual independence $Y \perp\!\!\!\perp A$.

---

[9] Also called ignorability.
[10] Also called overlap. Note that this is a feature of the distribution, not the sample.
[11] Similar to the condition SUTVA: Stable Unit Treatment Value Assumption.

Side note: relation to previous statistics courses

- So far you have considered random variables, say, Y.
- Y has a law – i.e. distribution – and you have *inferred*, *i.e. estimated or learned* features of this law: deterministic features of this random variable.
- In the regression courses, you went further and looked at random variables *conditional* on parameters. For example, linear regression is the best (min squared error) linear approximation of Y (or of $\mathbb{E}[Y \mid A]$). where x is a parameter.
- We consider the problem of **identifying** functionals $f(Y^a)$.
  If a functional is identified, then we can use what you have learned so far (and more) **to estimate** these functionals.

Terminology

Remember the difference between the following terms:

- **Estimand** (a parameter of interest, often a causal effect).
- **Estimator** (an algorithm / function that can be applied to data).
- **Estimate** (an output from applying the estimator to data).

We talk about bias of an estimator with respect to an estimand.
That is, the term *bias* (biased / unbiased) is defined with respect to an estimand.

Terminology

estimand

Ingredients

- 150g unsalted butter, plus extra for greasing
- 150g plain chocolate, broken into pieces
- 150g plain flour
- ½ tsp baking powder
- ½ tsp bicarbonate of soda
- 200g light muscovado sugar
- 2 large eggs

Method

1. Heat the oven to 160C/140C fan/gas 3. Grease and base line a 1 litre heatproof glass pudding basin and a 450g loaf tin with baking parchment.

2. Put the butter and chocolate into a saucepan and melt over a low heat, stirring. When the chocolate has all melted remove from the heat.

estimator

estimate

A simple example of estimation of causal effects

Because $\mathbb{E}(Y^a) = \mathbb{E}(Y \mid A = a)$, the simple difference-in-means estimator,

$$\hat{\delta} = \frac{1}{n_1} \sum_{A_i = 1} Y_i - \frac{1}{n_0} \sum_{A_i = 0} Y_i, \;\; n_a = \sum_{i=1}^{n} I(A = a),$$

is an unbiased estimator of the average (additive) causal effect of A in a randomised experiment.

We will discuss estimation in more detail later in this course.

Conditional randomisation

- Let $L \in \{0, 1\}$
  In the heart transplant example, let $L = 1$ if the individual is critically ill, 0 otherwise.
- Suppose A is *conditionally* randomised as a function of L such that $P(A = 1 \mid L = 0) = p_0$ and $P(A = 1 \mid L = 1) = p_1$,
  where $p_0 \neq p_1$ and $p_0, p_1 \in (0, 1)$.

How do we identify $\mathbb{E}(Y^a)$?

Illustrative *conditional* experiment (trial) on heart transplant

	L	A	Y
Rheia	0	0	0
Kronos	0	0	1
Demeter	0	0	0
Hades	0	0	0
Hestia	0	1	0
Poseidon	0	1	0
Hera	0	1	0
Zeus	0	1	1
Artemis	1	0	1
Apollo	1	0	1
Leto	1	0	0
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	1	1
Aphrodite	1	1	1
Cyclope	1	1	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0

In this conditional randomised trial $p_0 = 0.5, p_1 = 0.75$.
Compute an estimator based on the numbers above, and you will find that $\hat{\mathbb{E}}(Y^{a=1}) - \hat{\mathbb{E}}(Y^{a=0}) = 0$.

Identification in a conditional randomised experiment

A is *conditionally* randomised such that $P(A = 1 \mid L = 0) = p_0$ and $P(A = 1 \mid L = 1) = p_1$, where $p_0 \neq p_1$ and $p_0, p_1 \in (0, 1)$.

$Y^a \not\perp\!\!\!\perp A, \forall a \in \{0, 1\}$ (Exchangeability from Slide 63 may fail), but

1. $Y^a \perp\!\!\!\perp A \mid L, \forall a \in \{0, 1\}$ (Exchangeability).
2. $P(A = a \mid L = l) > 0 \ \forall a \in \{0, 1\}, \forall l$ s.t. $P(L = l) > 0$. (positivity).
3. $Y^a = Y$ for every unit with $A = a$ (consistency).

When 1-3 hold, then

$$\mathbb{E}(Y^a) = \sum_l \mathbb{E}(Y \mid L = l, A = a) P(L = l).$$

These conditions hold by design in a conditional randomised experiment.

Identification in a conditional randomised experiment

Proof.

$$
\begin{aligned}
\mathbb{E}(Y^a) &= \sum_l \mathbb{E}(Y^a \mid L = l)P(L = l) \\
&= \sum_l \mathbb{E}(Y^a \mid L = l, A = a)P(L = l) \quad \text{(exchangeability)} \\
&= \sum_l \mathbb{E}(Y \mid L = l, A = a)P(L = l). \quad \text{(positivity and consistency)}
\end{aligned}
$$

□

We say that the 3rd line is an identification formula for $\mathbb{E}(Y^a)$.
This is a special case of a so-called G-formula (or truncation formula).[12]

[12] James M Robins. "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect". In: *Mathematical modelling* 7.9-12 (1986), pp. 1393–1512.

Alternative weighted identification formula

$$\mathbb{E}(Y^a) = \sum_l \mathbb{E}(Y \mid L = l, A = a)\,\Pr(L = l)$$
$$= \mathbb{E}\left[\frac{I(A = a)}{\pi(A \mid L)} Y\right].$$

where $\pi(a \mid l) = P(A = a \mid L = l)$,
Why bother with equivalent expressions?
Because they motivate different *estimators*.

Proof of IPW

Proof.

$$
\begin{aligned}
&\mathbb{E}\left[\frac{I(A=a)}{\pi(A \mid L)} Y\right] \\
=&\mathbb{E}\left[\frac{I(A=a)}{P(A=a \mid L)} Y^a\right] \quad \text{(consistency and positivity)} \\
=&\mathbb{E}\left[\mathbb{E}\left\{\frac{I(A=a)}{P(A=a \mid L)} Y^a \mid L\right\}\right] \\
=&\mathbb{E}\left\{\mathbb{E}\left[\frac{I(A=a)}{P(A=a \mid L)} \mid L\right] \mathbb{E}\left[Y^a \mid L\right]\right\} \text{ (exchangeability)} \\
=&\mathbb{E}\left\{\mathbb{E}\left[Y^a \mid L\right]\right\} = \mathbb{E}\left[Y^a\right].
\end{aligned}
$$

□

Section 8

Lecture 3