Topic: basic causality, taking expectations

1. Causal stories

($3 \times 3 \times 0.5$ points) The following fictional news headlines state a correlation and suggest a causal relationship. However, we know that correlations do not always imply the "obvious" causation. For each headline, come up with three different hypotheses about the causal relationships, potentially including variables that are not mentioned in the headlines. Describe the hypotheses briefly in words and with a graphical representation, such as the ones given in "Explanation 1.1" in the lecture notes.

1. A vaccine is tested in country A and country B. In country A, patients received half a dose of the vaccine and it immunized 90% of patients, while in country B, patients received the full dose and it immunized 60% of patients.
2. Of a population, 10% belong to group A and 90% to group B, but 90% of those arrested for criminal offences belong to group A.
3. High school graduation grades are highly predictive for future career success.

2. Causal and non-causal predictions

($3 \times 2 \times 0.5$ points) In statistics, machine learning and data science, one is often interested in making predictions based on data. Some predictions are of a non-causal nature (for example: predict what the weather will be tomorrow), other predictions are of a causal nature (for example: predict the effect of CO_2 emission on global temperature). State three systems and for each system provide an example of a causal prediction and an example of a non-causal prediction that one could try to make based on data from the system.

3. FDA approval of Aduhelm

($1 + 1$ points) In the brains of Alzheimer's disease patients, amino acids called amyloid beta plaque are found. On June 7th, 2021, the FDA gave accelerated approval to the Alzheimer's drug Aduhelm, releasing the following statement:

"Today, the U.S. Food and Drug Administration approved Aduhelm (aducanumab) for the treatment of Alzheimer's, a debilitating disease affecting 6.2 million Americans. Aduhelm was approved using the accelerated approval pathway, which can be used for a drug for a serious or life-threatening illness that provides a meaningful therapeutic advantage over existing treatments. Accelerated approval can be based on the drug's effect on a surrogate endpoint that is reasonably likely to predict a clinical benefit to patients, with a required post-approval trial to verify that the drug provides the expected clinical benefit.(...)

Researchers evaluated Aduhelm's efficacy in three separate studies representing a total of 3,482 patients. The studies consisted of double-blind, randomized, placebo-controlled dose-ranging studies in patients with Alzheimer's disease. Patients receiving the treatment had significant dose-dependent and time-dependent reduction of amyloid beta plaque, while patients in the control arm of the studies had no reduction of amyloid beta plaque. These results support the accelerated approval of Aduhelm, which is based on the surrogate endpoint of reduction of amyloid beta plaque in the brain—a hallmark of Alzheimer's disease."[1]

1. The FDA bases the accelerated approval on certain assumed causal relations between the variables for treatment with Aduhelm T, the level of amyloid beta plaque in the patient's brain β, and the clinical outcome of the Alzheimer's patient O. Let E be the set of causal relations, where $(A, B) \in E$ if A causes B. What is the set of causal relations between T, β and O assumed by the FDA in the evaluation of this drug?

2. This decision has faced much controversy. For example, Karlawish and Grill write: "Substantial evidence from two randomized, placebo-controlled phase III trials shows that Aduhelm treatment is associated with a reduction in amyloid beta plaque signal in the brains of individuals with Alzheimer's disease. However, these same trials produced insufficient evidence that this reduction slows cognitive and functional decline. Indeed, there is a lack of correlation between changes in amyloid beta plaque and clinical outcomes in patients receiving Aduhelm."[2]

   Propose an alternate set of causal relations E', which could explain the presence of amyloid beta plaque in patients with Alzheimer's disease and the proven effect of Aduhelm on amyloid beta plaque, but where Aduhelm does not influence the clinical outcome.

4. Change of variables

(1 + 1 points)

1. Let X be distributed according to the discrete uniform distribution on $\{-M, -M+1, \dots, N-1, N\}$ where $M, N \in \mathbb{N}$ and $M > N$. Compute the probability mass function of the absolute value $Y = |X|$ w.r.t. counting measure.

2. Let X be distributed according to the standard normal distribution $\mathcal{N}(0, 1)$. Furthermore, let $Y = \exp(X)$. Compute the density of Y w.r.t. Lebesgue measure, where the codomain of Y is $\mathbb{R}^+$.

[1] https://www.fda.gov/drugs/news-events-human-drugs/fdas-decision-approve-new-treatment-alzheimers-disease

[2] https://www.nature.com/articles/s41582-021-00540-6

Ungraded practice exercises

These exercises can be helpful to get acquainted with (or refresh your knowledge of) measure theory.

5. Discrete and Continuous Dice

A discrete die has as sample space the six sides $\mathcal{X} = \{a, b, c, d, e, f\}$ with uniform measure P_D with probability mass function p_D. The value is a random variable $V_D : \mathcal{X} \to \mathbb{R}$ with:

$$V(a) = 1, V(b) = 2, V(c) = 3, V(d) = 4, V(e) = 5, V(f) = 6.$$

A continuous die has as sample space the unit interval $[0, 1]$ with uniform measure P_C with probability density function p_C. The value is a random variable $V_C : [0, 1] \to \mathbb{R} : x \mapsto 5x + 1$. Using the measure theoretical notation introduced in Lecture 1, compute the expected value of the discrete and continuous dice (2+2pts).

6. Proving properties

(1+2+1+1 points)

1. Consider spaces A, B and C with respective sigma algebras $\mathcal{F}, \mathcal{G}$ and $\mathcal{H}$. Show that if the function $f : A \to B$ is $\mathcal{F}$-$\mathcal{G}$-measurable and the function $g : B \to C$ is $\mathcal{G}$-$\mathcal{H}$ measurable, then their composition $g \circ f$ is $\mathcal{F}$-$\mathcal{H}$-measurable.
2. Let $\mathcal{U}$ be an arbitrary set, $(\mathcal{V}, \Sigma)$ a measurable space, and $X : \mathcal{U} \to \mathcal{V}$ an arbitrary function. Show that $\Sigma_X = \{X^{-1}(A) : A \in \Sigma\}$ is a σ-algebra over $\mathcal{U}$. Furthermore, show that Σ_X is the *smallest* sigma algebra that makes the map measurable. More specifically, show that if $\mathcal{F}$ is a sigma algebra for U such that X is $\mathcal{F}$-Σ-measurable, then $\Sigma_X \subseteq \mathcal{F}$.
3. Show that the indicator function $\mathbb{1}_A : \Omega \to \{0, 1\}$ is $\mathcal{B}_\Omega$-$2^{\{0,1\}}$ measurable if $A \in \mathcal{B}_\Omega$.
4. Consider the Lebesgue measure λ on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$. Show that $\lambda(\{1, \frac{1}{2}, \frac{1}{3}, \dots\}) = 0$.

7. Expectation and integration

(1 point) Let $X : \Omega \to \mathbb{R}$ be a random variable. Show from first principles that for simple functions $\mathbb{E}[cX] = c\mathbb{E}[X]$ (i.e. you are **not** allowed to just apply the linearity of the integral, you must show this yourself).
Bonus: Generalize this statement to the general case.