ORIE 6745: Causality and Learning for Intelligent Decision Making August 30, 2017

Lecture 3-4

Lecturer: Nathan Kallus *Scribe: Ashudeep Singh*

1 Recap

We defined counterfactuals using the Potential Outcomes framework.
$Y(z)$: The outcome (for a generic unit) that we would have observed if we applied the treatment z.
In general, we don't observe all of these $\{Y(z) : z \in \mathcal{Z}\}$. We just observe a single factual outcome $Y = Y(Z)$ corresponding to the actually applied treatment Z.
A few things to always keep in mind:

- Association $\neq$ Causation
- Prediction $\neq$ Decision
- $Y|Z = z \overset{d}{\neq} Y(z)$: This means that the outcomes of the treated for a particular treatment may not have the same distribution as the marginal distribution of the outcome for the treatment. For example, if Z is sleeping while wearing shoes or not, Y is getting a headache in the morning, then $Y(z)$ is the distribution of outcomes on the entire population and this may not be the same as $Y|Z = z$ i.e. the distribution of headache for people wearing shoes while sleeping. This makes sense because *alternative explanations* might exist such as drinking in the night causes people to wear shoes while sleeping and also causes headache in the morning.

Later on in the semester, we will study more about creating statistical homogeneity with randomization and controlled trials. Also, we'll talk about *unconfoundedness* i.e. holding everything constant conditional on observables.
For now, we will study the easy case of learning to decide with fully-observed counterfactuals. Observer $Y_i(z) \forall z \in \mathcal{Z}, \forall i = 1, \ldots, n$. We want to find a policy $\pi : \mathcal{X} \to \mathcal{Z}$. Policy risk is defined as $R(\pi) = \mathbb{E}\left[Y(\pi(X)\right]$.
Empirical policy risk: $\hat{R}_n(\pi) = \frac{1}{n} Y_i(\pi(X_i))$.
In the last class, we proved the following:

Theorem 1 (Hoeffding Inequality)**.** *If $V_i \in [a_i, b_i]$ are independent random variables for $i \in [n]$, then*

$$P\left(\left|\frac{1}{n}\sum_i V_i - \frac{1}{n}\mathbb{E}\sum_i V_i\right| \geq \epsilon\right) \leq 2\exp\left(\frac{-2n\epsilon^2}{\frac{1}{n}\sum_{i=1}^n (b_i - a_i)^2}\right).$$

So, if $|Y(z)| \leq c$, then for a fixed π, $\hat{(}R_n)(\pi)$ is tightly concentrated near $R(\pi)$. (Note: This can be generalized for $Y(z)$ that are sub-gaussian (light-tails) instead of bounded.
However we actually want the following;
If we choose $\hat{\pi}_n \in \text{argmin}_{\pi \in \Pi} \hat{R}_n(\pi)$, then we want $D(\Pi) = \sup_{\pi \in \Pi)} \left|\hat{R}_n(\pi) - R(\pi)\right|$ to be small. One reason for this is to use *Empirical Risk minimization*(ERM).

$$\begin{aligned} R(\hat{\pi}_n) &\leq \hat{R}_n(\hat{\pi}_n) + D(\Pi) && \text{(by definition)} \\ &\leq \hat{R}_n(\pi^*) + D(\Pi) && \text{(for } \pi^* \in \Pi \text{ by optimality of } \hat{\pi}_n) \\ &\leq R(\pi^*) + 2D(\Pi) && \text{(by definition)} \end{aligned}$$

Since this holds for any $\pi^* \in \Pi$, this also holds in particular for $\pi^* \in \text{argmin}_{\pi \in \Pi} R(\pi)$. i.e. the best in-class policy.

In this lecture, we will learn different ways to bound $D(\Pi)$ with high probability in terms of the complexity of the class Π.

2 McDiarmid's Inequality

Theorem 2 (McDiarmid's Inequality)**.** *Let $f : \mathcal{V}^n \to \mathbb{R}$ be such that $\forall i$, $v_1, v_2, \ldots, v_n, v_i'$,*

$$|f(v_1, v_2, \ldots, v_i, \ldots, v_n) - f(v_1, v_2, \ldots, v_i', \ldots, v_n)| \leq c_i$$

(we call this bounded difference condition) Let $V_1, V_2, \ldots V_n$ be independent random variables in $\mathcal{V}$, then:

$$\mathbb{P}(f(v_1, \ldots, v_n) - \mathbb{E}f(v_{1:n}) \geq \epsilon) \leq exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^n c_i^2}\right)$$

Proof. Construct a Doob Martingale, $W_i = \mathbb{E}[f(V_{1:n})|V_{1:i}]$.
Notes:

1. $W_0 = \mathbb{E}f(V_{1:n})$
2. $W_n = \mathbb{E}f(V_{1:n})$
3. $\mathbb{E}\left[W_i - W_{i-1}|V_{1:i-1}\right] = 0$
4. $\mathbb{E}\left[e^{t(W_i - W_{i-1})}|V_{1:i-1}\right] \leq e^{t^2 c_i^2/8}$ from Hoeffding's Lemma.

Therefore, for $t > 0$,

$$\begin{aligned}
\mathbb{P}(W_n - W_0) &= \mathbb{P}(e^{t(W_n - W_0)} \geq e^{t\epsilon}) \\
&\leq e^{-t\epsilon}\mathbb{E}\left[e^{t(W_n - W_0)}\right] && \text{(by Markov's inequality)} \\
&= e^{-t\epsilon}\mathbb{E}\left[\Pi_{i=1}^n e^{t(W_i - W_{i-1})}\right] \\
&= e^{-t\epsilon}\mathbb{E}\left[\mathbb{E}\left[\Pi_{i=1}^n e^{t(W_i - W_{i-1})}|V_{1:n-1}\right]\right] && \text{(iterated expectation)} \\
&= e^{-t\epsilon}\mathbb{E}\left[\Pi_{i=1}^n e^{t(W_i - W_{i-1})}\mathbb{E}\left[e^{t(W_n - W_{n-1})}|V_{1:n-1}\right]\right] \\
&\leq e^{-t\epsilon}\mathbb{E}\left[\Pi_{i=1}^n e^{t(W_i - W_{i-1})}e^{t^2 c_n^2/8}\right] \\
&\leq \ldots && \text{(repeating this)} \\
&\leq e^{-t\epsilon}\Pi_{i=1}^n e^{t^2 c_i^2/8} \\
&= exp(-t\epsilon + t^2 \sum_i c_i^2/8) \\
&\text{Let } t = \frac{4\epsilon}{\sum_{i=1}^n c_i^2/8} && \text{(obtained by first order optimality condition)} \\
&\leq -t\epsilon + t^2 \sum_i c_i^2/8 = \frac{-2\epsilon}{\sum_{i=1}^n c_i^2} \quad \square
\end{aligned}$$

$\square$

Suppose $|Y(z)| \leq c$. Then:

$$\Big| \underbrace{\sup_{\pi\in\Pi} \Big| \frac{1}{n}\sum_{j\neq i} Y_j(\pi(x_j)) + \frac{1}{n} Y_i(\pi(x_i) - \mathbb{E}Y(\pi(x))\Big|)}_{\text{(A)}} - \underbrace{\sup_{\pi\in\Pi} \Big| \frac{1}{n}\sum_{j\neq i} Y_j(\pi(x_j)) + \frac{1}{n} Y_i'(\pi(x_i') - \mathbb{E}Y(\pi(x)))\Big|}_{\text{(B)}} \Big|$$

$$\leq \sup_{\pi\in\Pi} \Big| |\text{(A)}| - |\text{(B)}| \Big|$$

$$(\sup_x f(x) - \sup_x g(x) \leq \sup_x f(x) - g(x))$$

$$\leq \sup_{\pi\in\Pi} \Big| \text{(A)} - \text{(B)} \Big| \qquad \text{(Triangle Inequality)}$$

$$= \sup_{\pi\in\Pi} \frac{1}{n} |Y_i(\pi(X_i) = y_i'(\pi(x_i')| \leq 2c/n$$

Therefore, letting $\epsilon = c\sqrt{2log(1/\delta)/n}$, McDiarmid's inequality will give, with probability $\geq 1-\delta$,

$$D(\Pi) \leq \mathbb{E}D(\Pi) + c\sqrt{2log(1/\delta)/n}$$

Let us see what $\mathbb{E}D(\Pi)$ is:

$$\begin{aligned}
\mathbb{E}D(\Pi) &= \mathbb{E}\sup_{\pi\in\Pi} \Big| \frac{1}{n}\sum_i Y_i(\pi_i(x_i)) - \mathbb{E}Y(\pi(X)) \Big| \\
&= \mathbb{E}\sup_{\pi\in\Pi} \Big| \frac{1}{n}\sum_i Y_i(\pi_i(x_i)) - \sum_i \mathbb{E}\left[Y_i'(\pi(X_i')) | X_{1:n}, Y_{1:n} \right] \Big| \qquad () \\
&\leq \mathbb{E}\left[\sup_{\pi\in\Pi} \left| \frac{1}{n}\sum_i Y_i(\pi(X_i)) - \frac{1}{n}\sum_i Y_i'(\pi(X_i')) \right| \right] \\
&= \mathbb{E}\left[\sup_{\pi\in\Pi} \left| \frac{1}{n}\sum_i (Y_i(\pi(X_i)) - Y_i'(\pi(X_i'))) \right| \right] \\
&= \mathbb{E}\left[\sup_{\pi\in\Pi} \left| \frac{1}{n}\sum_i \sigma_i (Y_i(\pi(X_i)) - Y_i'(\pi(X_i'))) \right| \right] \\
&\text{(Let } \sigma_i = \pm 1 \text{ be equal probability, independent Radamacher random variables)}
\end{aligned}$$

Definition 3 (Radamacher Complexity)**.** Given a set of point $\mathcal{A} \subseteq \mathbb{R}^n$ be the Radamacher complexity of $\mathcal{A}$ is:

$$\mathscr{R}(\mathcal{A}) = \frac{1}{2^n} \sum_{\sigma\in\{-1,+1\}^n} \sup_{a\in\mathcal{A}} \frac{1}{n}\sum_{i=1}^n \sigma_i a_i$$

So,

$$\mathbb{E}D(\Pi) \leq 2\mathbb{E}\left[\mathscr{R}(\mathcal{Y}(\Pi))\right]$$

when $\mathcal{Y}(\pi) = \{(Y_1(\pi(X_1)), \ldots, Y_n(\pi(X_n)) : \pi \in \Pi\}$.

Again, if $|Y(z)| \le c$, then $|\mathscr{R}(\mathcal{Y}(\Pi) - \mathscr{R}(\mathcal{Y}'(\Pi))|$ (where $\mathcal{Y}$ is obtained by replacing $X_i, Y_i'(.)$ by $X_i', Y_i'(.)$

$$|\mathscr{R}(\mathcal{Y}(\Pi) - \mathscr{R}(\mathcal{Y}'(\Pi)))| = \left|\mathbb{E}_\sigma\left[\underbrace{\sup_{\pi\in\Pi}\left(\frac{1}{n}\sum_{j\neq i}\sigma_j Y_j(\pi(x_j)) + \frac{1}{n}\sigma_i Y_i y(\pi(x_i))\right)}_{\text{(A)}} - \underbrace{\sup_{\pi\in\Pi}\left(\frac{1}{n}\sum_{j\neq i}\sigma_j Y_j(\pi(x_j)) + \frac{1}{n}\sigma_i Y_i'(\pi(x_i'))\right)}_{\text{(B)}}\right]\right|$$

$$\le \mathbb{E}_\sigma\left|\sup_{\pi\in\Pi}\text{(A)} - \sup_{\pi\in\Pi}\text{(B)}\right| \le \mathbb{E}_\sigma \sup_{\pi\in\Pi}|\text{(A)} - \text{(B)}|$$

$$= \mathbb{E}_\sigma \sup_{\pi\in\Pi}|\sigma_i|\left|\frac{1}{n}(Y_i(\pi(x_i)) - Y_i'(\pi(X_i')))\right|$$

$$\le \frac{2c}{n}$$

So by McDiarmid's, with probability $\ge 1-\delta$, $\mathbb{E}\mathscr{R}_n(\mathcal{Y}_n(\Pi)) \le \mathscr{R}_n(\mathcal{Y}_n(\Pi)) + c\sqrt{\frac{2\log 1/\delta}{n}}$.
So far:

$$D(\Pi) \le \mathscr{R}_n(\mathcal{Y}_n(\Pi)) + 2c\sqrt{\frac{2\log 2/\delta}{n}}$$

Next: Interpret $\mathscr{R}_n(\mathcal{Y}(\Pi))$ as the complexity of the policy class Π.

Theorem 4 (An extension of Ledoux-Talegrand Lipschitz comparison lemmma)**.** *Suppose $\mathcal{Z} \subseteq \mathbb{R}^d$,*
Fix $X_i, Y_i(\cdot), i = 1, \dots, n$
Suppose each $Y_i(.)$ is L-Lipschitz i.e. $Y_i(z) - Y_i(z') \le L||z - z'||\infty$, for all i.
Let,

$$\mathcal{A}_n^{(k)}(\Pi) = \{((\pi(X_1))_k, \dots, (\pi(X_n))_k) : \pi \in \Pi\}$$

Then,

$$\mathscr{R}_n(\mathcal{Y}_n(\Pi)) \le L\sum_{i=1}^d \mathscr{R}_n(\mathcal{A}_n^{(k)}(\Pi))$$

Proof. Without loss of generality, $L = 1$ (otherwise, divide LHS by L).
We will show the following (stronger than the theorem):

$$\mathbb{E}_\sigma \sup_{\pi\in\Pi}\frac{1}{n}\sum_i \sigma_i Y_i \Pi(X_i) \le \mathbb{E}_{\sigma\in\{-1,1\}^{n\times d}} \sup_{\pi\in\Pi}\frac{1}{n}\sum_{i=1}^n\sum_{k=1}^d \sigma_{ik}(\pi(X_i))_k$$

Suppose we have that for any $\mathcal{S} \subseteq \mathbb{R}\times\mathcal{Z}$ and a 1-Lipchitz $Q : \mathcal{Z} \to \mathbb{R}$,

$$\mathbb{E}_\sigma \sup_{t,z\in\mathcal{S}}(t \in \sigma Q(z)) \le \mathbb{E}_\sigma \sup_{t,z\in\mathcal{S}}(t + \sum_k \sigma_k z_k) \qquad (*)$$

Then:

$$\mathbb{E}_\sigma \sup_{\pi\in\Pi}\frac{1}{n}\sum_i \sigma_i Y_i \Pi(X_i) = \mathbb{E}_{\sigma_1,\dots,\sigma_{n-1}}\left[\mathbb{E}_{\sigma_n}\left[\sup_{\substack{\pi\in\Pi \\ t=\frac{1}{n}\sum_{i=1}^{k-1}\sigma_i Y_i(\pi(x_i)) \\ z=\pi(x_n)}} t+\sigma_n Y_n(z)|\sigma_{1:n-1}\right]\right] \le \cdots \le \mathbb{E}_\sigma \sup_{\pi\in\Pi}\frac{1}{n}\sum_i\sum_k \sigma_{ik}(\pi(x_i))_k$$

Let's show $(*)$ now:
Fix any $(t^{(+1)}, z^{(+1)}), (t^{(+1)}, z^{(+1)}) \in \mathcal{S}$.
Let $s^* = \pm 1, k^* = 1, \dots d$ be

$$||z^{(+1)} - z^{(-1)}||_\infty = s^*(z_{k^*}^{(+1)} - z_{k^*}^{(-1)})$$

$$
\begin{aligned}
RHS\textcircled{*} &\geq \mathbb{E}_{\sigma_{k^*}}\left[\sup_{t,z\in\mathcal{S}} \mathbb{E}_{\sigma_k:k\neq k^*}\left[t+\sum_k \sigma_k z_k|\sigma_{k^*}\right]\right] \\
&\geq \frac{1}{2}\sigma_{\sigma_k:k\neq k^*}\left[t^{(+s^*)}+z_{k^*}^{+s^*}+\sum_{k\neq k^*}\sigma_k z_k^{+s^*}\right] \\
&\quad + \frac{1}{2}\sigma_{\sigma_k:k\neq k^*}\left[t^{(-s^*)}-z_{k^*}^{-s^*}+\sum_{k\neq k^*}\sigma_k z_k^{+s^*}\right] \\
&= \frac{1}{2}\left(t^{(-1)}+t^{(+1)}+||z^{(+1)}-z^{(-1)}||_\infty\right) \\
&\geq \frac{1}{2}\left(t^{(+1)}+\psi(z^{(+1)})\right)+\frac{1}{2}\left(t^{(-1)}+\psi(z^{(-1)})\right)
\end{aligned}
$$

Taking suprema over $t^{(\pm 1)}, z^{(\pm 1)}$, we set the LHS of $\textcircled{*}$. □

All together, with probability $\geq 1-\delta$,

$$
D(\Pi) \leq L\sum_{k=1}^{d}\mathscr{R}_n(\mathcal{A}_n^{(k)}(\Pi)) + 2c\sqrt{\frac{2\log(2/\delta)}{n}}
$$

2.1 Example 1: Supervised Learning

Supervised regression with bounded outcomes, $Y_i(z) = (L_i - z)^2$ is 2c-Lipschitz $z \in \mathbb{R}$ univariate. So:

$$
D(\Pi) \leq \underbrace{\mathscr{R}_n(\mathcal{A}_n(\Pi))}_{\text{What's this? Let's see!}} + 2c\sqrt{\frac{2\log(2/\delta)}{n}}
$$

One kind of bound:
If $||X||_2 \leq M$, $\Pi = \{X \to \beta^T X : ||\beta||_2 \leq R\}$, then:

$$
\begin{aligned}
\mathscr{R}_n(\mathcal{A}_n(\Pi)) &= \mathbb{E}_\sigma \sup_{||\beta||_2\leq R}\frac{1}{n}\sum_{i=1}^n \sigma_i\beta^T X_i \\
&= \mathbb{E}_\sigma \sup_{||\beta||_2\leq R}\beta^T(\frac{1}{n}\sum_i \sigma_i X_i) \\
&= \mathbb{E}_\sigma\left[R.||\frac{1}{n}\sigma_i\sigma_i X_i||_2\right] && (\sup_{||x||_2\leq 1} v^T x = ||x||_2) \\
&\leq R\sqrt{\mathbb{E}_\sigma||\frac{1}{n}\sum_i \sigma_i X_i||_2^2} && (\text{ Parallelogram law: } \tfrac{1}{2}||X_1+X_2||_2^2 = ||X_1||^2+||X_2||^2) \\
&= \frac{R}{n}\sqrt{\sum_{i=1}^n ||X_i||_2^2} \leq R\frac{\sqrt{nM^2}}{n} = \frac{RM}{\sqrt{n}} \text{With high probability} && \forall\pi\in\Pi
\end{aligned}
$$

So, whp, $\mathscr{R}(\pi) \leq \hat{\mathscr{R}}_n(\pi) + 2c\frac{RM}{\sqrt{n}} + 2c\sqrt{\frac{2\log(2/\delta)}{n}}$

Leads to Structural Risk Minimization: Choose $\hat{\pi_n}$ to minimize $\hat{R}_n(x \mapsto \beta^T x) + \lambda||\beta||_2$.

3 VC-Dimension

Now, we will talk about another kind of bound: VC-Dimension.
Suppose $\mathcal{Z} = \{\pm 1\}$, e.g.:

- predict positive/negative classification
- give or don't give treatment (actual decision)

Then $Y(\cdot)$ is 2c-Lipschitz. What about $\mathscr{R}_n(\mathcal{A}_n(\Pi))$?

Lemma 5 (Massart's lemma)**.** *Let $\mathcal{A} \subseteq [-M, M]^n$. Then:*

$$\mathscr{R}_n(\mathcal{A}) \leq M\sqrt{\frac{2}{n}\log|\mathcal{A}|}$$

Proof. Let $\lambda > 0$. Then:

$$\begin{aligned}
e^{\lambda\mathscr{R}_n(\mathcal{A})} &\leq \mathbb{E}_\sigma\left[exp(\lambda \sup_{a\in\mathcal{A}} \frac{1}{n}\sum_i \sigma_i a_i)\right] \\
&= \mathbb{E}_\sigma\left[\sup_{a\in\mathcal{A}} exp(\lambda\frac{1}{n}\sum_i \sigma_i a_i\right] \\
&\leq \mathbb{E}_\sigma\left[\sum_{a\in\mathcal{A}} exp(\frac{\lambda}{n}\sum_i \sigma_i a_i)\right] \\
&= \sum_{a\in\mathcal{A}} \Pi_{i=1}^n \mathbb{E}_{\sigma_i} exp(\frac{\lambda}{n}\sigma_i a_i) \\
&= \sum_{a\in\mathcal{A}} \Pi_{i=1}^n \left(\frac{1}{2}exp(\frac{\lambda}{n}a_i) + \frac{1}{2}exp(\frac{-\lambda}{n}a_i)\right) \\
&\leq \sum_{a\in\mathcal{A}} \Pi_{i=1}^n e^{\lambda^2 a_i^2/2n^2} && (cosh(x) \leq e^{x^2/2}) \\
&\leq \sum_{a\in\mathcal{A}} \Pi_{i=1}^n e^{\lambda^2 M^2/2n^2} \\
&\leq \sum_{a\in\mathcal{A}} e^{\lambda^2 M^2/2n} \\
&= |\mathcal{A}| e^{\lambda^2 M^2/2n} \\
\Rightarrow \mathscr{R}_n(\mathcal{A}) &\leq \frac{log|\mathcal{A}|}{\lambda} + \frac{\lambda M^2}{2n} \text{Set} \lambda = \sqrt{2n\log|\mathcal{A}|}/M \\
\Rightarrow \mathscr{R}_n(\mathcal{A}) &\leq \sqrt{2\log|\mathcal{A}|/n}.M.D
\end{aligned}$$

□

Definition 6. $|\mathcal{A}(\Pi)|$ is called the growth function.

Definition 7. The VC-dimension of $\Pi \subseteq [\mathcal{X} \to \{-1, +1\}]$, $VC(\Pi)$, is the largest number n s.t. $\forall x_1, x_2, \ldots, x_n \in \mathcal{X}$. We have:

$$|\mathcal{A}_n(\Pi)| = 2^n$$

called "Π shatters $x_1, \ldots, x_n$" i.e. any labeling of ± 1 on the set of points can be described using the model class Π.

Proposition 8. $\Pi = \{x \to sign(\beta^T x) : \beta \in \mathbb{R}^p\}, x \in \mathbb{R}^p$ *has VC-dimension p.*

Proof. To show $VC(\Pi) \geq p$: find p points that we can shatter. Consider $X_1 = (1, 0, \ldots, 0), X_2 = (0, 1, \ldots, 0), \ldots X_p = (0, \ldots, 1)$.
Let $z \in \{-1, 1\}^p$ be given. Set $\beta_i = z_i$, we get

$$(\pi(x_1), \pi(x_2), \ldots, \pi(x_p)) = (z_i, \ldots, z_p)$$

To show $VC(\pi) \leq p$: show that we cannot shatter $p + 1$ points.
Let $X_1, X_2, \ldots, X_{p+1}$ be given. They cannot be linearly independent. So $\exists \lambda \neq 0, \sum_i \lambda_i X_i = 0$

$$\Rightarrow \exists \lambda, j : X_j = \sum_{i \neq j} \lambda_i X_i$$

Let $\beta \in \mathbb{R}^p$ and suppose:

$$sign(\beta^T X_i) = sign(a_i), \forall i \neq j$$
$$\Rightarrow sign(\beta^T X_j) = sign(\sum_{i \neq j} a_i \beta^T X_i) = \pm 1$$

□

3.1 Why VC dimension?

Lemma 9 (Sauer's Lemma)**.** *If $VC(\Pi) \leq v$, then $|\mathcal{A}(\Pi)| \leq \sum_{i=1}^{v} \binom{n}{i} \leq \left(\frac{en}{v}\right)^v$*

Till now, we have studied Radamacher Complexity, Lipschitz comparison, McDiarmid's inequality, Massart's Lemma and Sauer's lemma.
For binary decision policies and bounded outcomes,

$$\begin{aligned} D(\Pi) &= \sup_{\pi \in \Pi} |\hat{\mathscr{R}}_n(\pi) - \mathscr{R}(\pi)| \\ &\leq 2c\sqrt{2\log(2/\delta)/n} + 4c\sqrt{\frac{2\log n}{n} VC(\pi)} \end{aligned}$$

Note: We can extend this to get:

- rid of $\log n$ (Chaining argument and see Pollard reading)
- replace bounded requirement by subgaussian (lecture 2)
- deal with non-binary decision

4 Further

Usually counterfactuals are not observed, we only observed factuals $Y = Y(z)$. We need to understand the difference between $Y|Z = z$ vs. $Y(z)$. In the next class, we learn about controlled experiments.