Machine Learning

Justin Grimmer

University of Chicago

February 28th, 2018

Donald J. Trump @realDonaldTrump · Feb 17

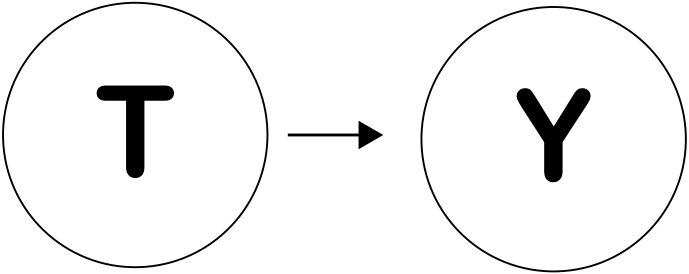
General McMaster forgot to say that the results of the 2016 election were not impacted or changed by the Russians and that the only Collusion was between Russia and Crooked H, the DNC and the Dems. Remember the Dirty Dossier, Uranium, Speeches, Emails and the Podesta Company!

42K 27K 96K


Shelby Hall
Engineering and Computing Sciences

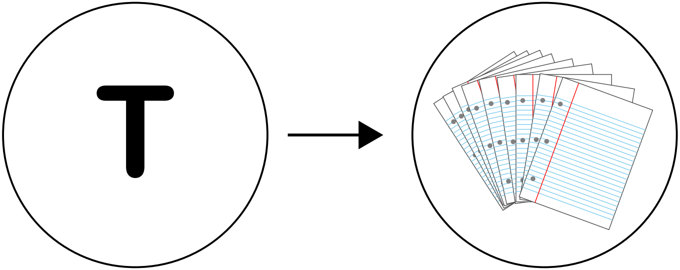
The Causal Inference Problem

- Two roles: text as outcome and text as treatment



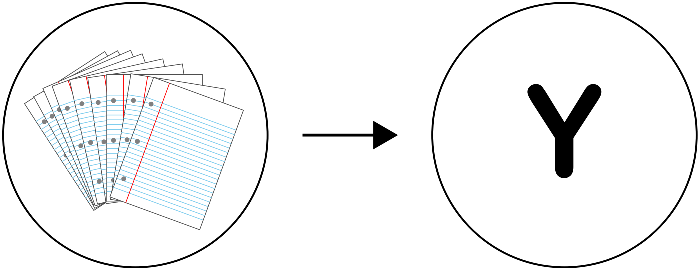
The Causal Inference Problem

- Two roles: text as outcome and text as treatment



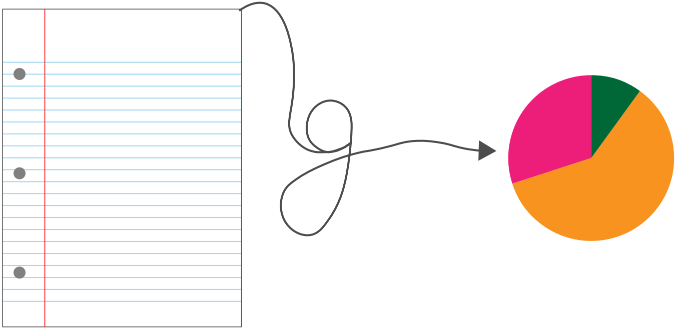
The Causal Inference Problem

- Two roles: text as outcome and text as treatment



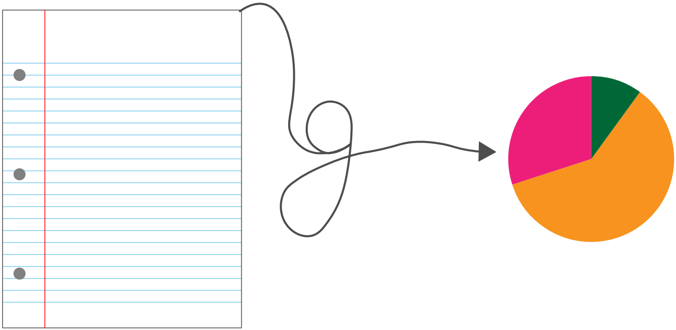
The Causal Inference Problem

- Two roles: text as outcome and text as treatment
- Any text analysis requires mapping from high to low dimensions



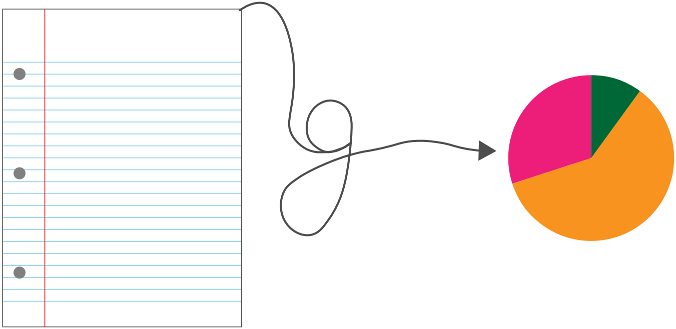
The Causal Inference Problem

- Two roles: text as outcome and text as treatment
- Any text analysis requires mapping from high to low dimensions
- Need to look at the data to create a good mapping



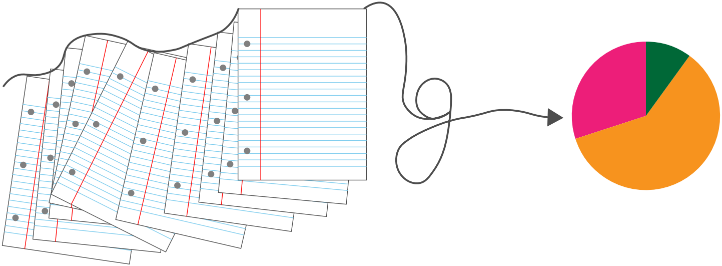
The Causal Inference Problem

- Two roles: text as outcome and text as treatment
- Any text analysis requires mapping from high to low dimensions
- Need to look at the data to create a good mapping
  - $\rightarrow$ potential for problems like overfitting



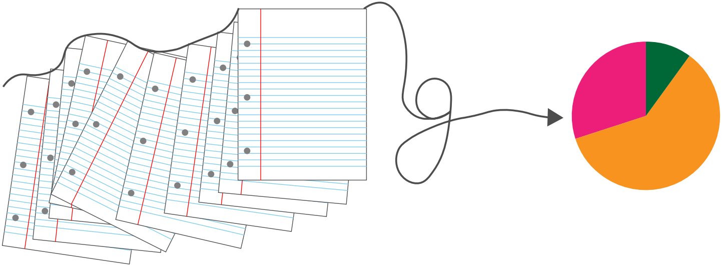
The Causal Inference Problem

- Two roles: text as outcome and text as treatment
- Any text analysis requires mapping from high to low dimensions
- Need to look at the data to create a good mapping
  - $\rightarrow$ potential for problems like overfitting
- One subtle problem: Analyst-Induced SUTVA violation



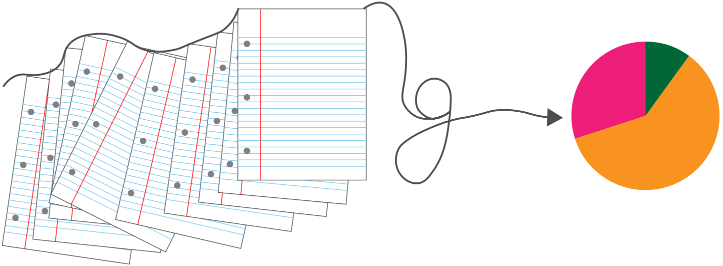
The Causal Inference Problem

- Two roles: text as outcome and text as treatment
- Any text analysis requires mapping from high to low dimensions
- Need to look at the data to create a good mapping
  - $\rightarrow$ potential for problems like overfitting
- One subtle problem: Analyst-Induced SUTVA violation
  - $\rightarrow$ outcome depends on every unit's treatment assignment



The Causal Inference Problem

- Two roles: text as outcome and text as treatment
- Any text analysis requires mapping from high to low dimensions
- Need to look at the data to create a good mapping
  - $\rightarrow$ potential for problems like overfitting
- One subtle problem: Analyst-Induced SUTVA violation
  - $\rightarrow$ outcome depends on every unit's treatment assignment
  - $\rightarrow$ SUTVA violation induced by the analyst



What's a social scientist to do?

What's a social scientist to do?

Framework for causal inference in texts with discovery (Athey and Wager, Chernozhukov et al., Cranmer and Desmarais, Fafchamps and Labonne, Fong and Grimmer):

What's a social scientist to do?

Framework for causal inference in texts with discovery (Athey and Wager, Chernozhukov et al., Cranmer and Desmarais, Fafchamps and Labonne, Fong and Grimmer):

A) (Not) Pre-Analysis Plan:

What's a social scientist to do?

Framework for causal inference in texts with discovery (Athey and Wager, Chernozhukov et al., Cranmer and Desmarais, Fafchamps and Labonne, Fong and Grimmer):

A) (Not) Pre-Analysis Plan:

- Specify g function *a priori* based on sample documents or some background knowledge.

What's a social scientist to do?

Framework for causal inference in texts with discovery (Athey and Wager, Chernozhukov et al., Cranmer and Desmarais, Fafchamps and Labonne, Fong and Grimmer):

A) (Not) Pre-Analysis Plan:
   - Specify g function *a priori* based on sample documents or some background knowledge.

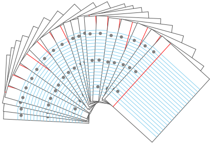
B) Train-Test Split (Our Plan):

What's a social scientist to do?

Framework for causal inference in texts with discovery (Athey and Wager, Chernozhukov et al., Cranmer and Desmarais, Fafchamps and Labonne, Fong and Grimmer):

A) (Not) Pre-Analysis Plan:
   - Specify g function *a priori* based on sample documents or some background knowledge.

B) Train-Test Split (Our Plan):
   1) Explicitly set aside a training set for discovery.



What's a social scientist to do?

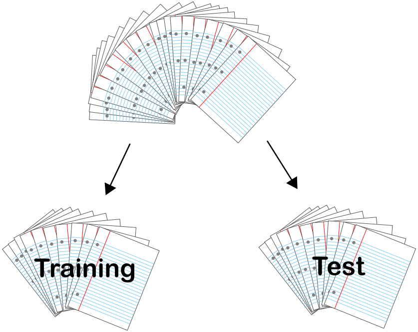
Framework for causal inference in texts with discovery (Athey and Wager, Chernozhukov et al., Cranmer and Desmarais, Fafchamps and Labonne, Fong and Grimmer):

A) (Not) Pre-Analysis Plan:

- Specify g function *a priori* based on sample documents or some background knowledge.

B) Train-Test Split (Our Plan):

1) Explicitly set aside a training set for discovery.



What's a social scientist to do?

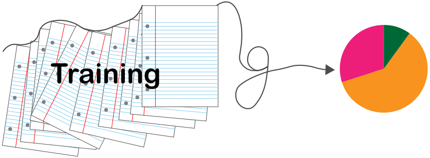
Framework for causal inference in texts with discovery (Athey and Wager, Chernozhukov et al., Cranmer and Desmarais, Fafchamps and Labonne, Fong and Grimmer):

A) (Not) Pre-Analysis Plan:

- Specify g function *a priori* based on sample documents or some background knowledge.

B) Train-Test Split (Our Plan):

1) Explicitly set aside a training set for discovery.
2) Use this training set to develop a g function that maps high-dimensional data to measurements.

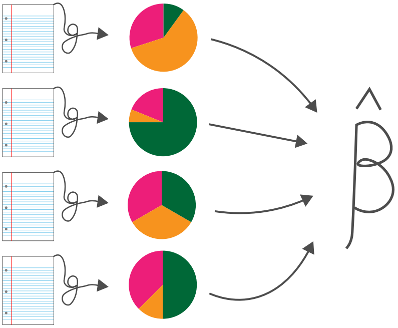


What's a social scientist to do?

Framework for causal inference in texts with discovery (Athey and Wager, Chernozhukov et al., Cranmer and Desmarais, Fafchamps and Labonne, Fong and Grimmer):

A) (Not) Pre-Analysis Plan:
   - Specify g function *a priori* based on sample documents or some background knowledge.

B) Train-Test Split (Our Plan):
   1) Explicitly set aside a training set for discovery.
   2) Use this training set to develop a g function that maps high-dimensional data to measurements.
   3) Given g, estimate the causal effect in test set



What's a social scientist to do?

Framework for causal inference in texts with discovery (Athey and Wager, Chernozhukov et al., Cranmer and Desmarais, Fafchamps and Labonne, Fong and Grimmer):

A) (Not) Pre-Analysis Plan:
   - Specify g function *a priori* based on sample documents or some background knowledge.

B) Train-Test Split (Our Plan):
   1) Explicitly set aside a training set for discovery.
   2) Use this training set to develop a g function that maps high-dimensional data to measurements.
   3) Given g, estimate the causal effect in test set

Train-Test allows for discovery while avoiding possibilities of fishing.

Two Running Examples: Treatment and Outcome

- Trump tweets $\rightsquigarrow$ partisan reactions
- Presidents "Going Public" $\rightsquigarrow$ media coverage

How do people react to Trump's rhetoric?

Donald J. Trump
@realDonaldTrump

Following

Little Adam Schiff, who is desperate to run for higher office, is one of the biggest liars and leakers in Washington, right up there with Comey, Warner, Brennan and Clapper! Adam leaves closed committee hearings to illegally leak confidential information. Must be stopped!

4:39 AM - 5 Feb 2018

31,930 Retweets **99,706** Likes

48K 32K 100K

Tweet 1:

Why would Kim Jong-un insult me by calling me “old," when I would NEVER call him “short and fat?" Oh well, I try so hard to be his friend-and maybe someday that will happen!

Tweet 2:

Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before. Fake News needs the competition!

Tweet 1:

`Why would Kim Jong-un insult me by calling me “old," when I would NEVER call him “short and fat?" Oh well, I try so hard to be his friend-and maybe someday that will happen!`

Tweet 2:

`Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before. Fake News needs the competition!`

Observe difference in evaluations of biographies

Tweet 1:

`Why would Kim Jong-un insult me by calling me "old," when I would NEVER call him "short and fat?" Oh well, I try so hard to be his friend-and maybe someday that will happen!`

Tweet 2:

`Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before. Fake News needs the competition!`

Observe difference in evaluations of biographies $\rightsquigarrow$ Difficult to generalize underlying features (treatments) that drive response

Tweet 1:

`Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before. Fake News needs the competition!`

Tweet $1'$:

`Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before.` ~~Fake~~ `News needs the competition!`

Tweet 1:

`Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before. Fake News needs the competition!`

Tweet $1'$:

`Steve Bannon will be a tough and smart new voice at @BreitbartNews...maybe even better than ever before.` ~~`Fake`~~ `News needs the competition!`

Randomly assign 1, $1'$ and assess response $\rightsquigarrow$ are we interested in effect of one word?

Tweet 1:

Negotiations on DACA have begun. Republicans want to make a deal and Democrats say they want to make a deal. Wouldn't it be great if we could finally, after so many years, solve the DACA puzzle. This will be our last chance, there will never be another opportunity! March 5th.

Tweet 2:

Negotiations on DACA have begun. Republicans want to make a deal and Democrats say they want to make a deal. Wouldn't it be great if we could finally, after so many years, solve the DACA puzzle. I will use my office to negotiate a fair deal for the dreamers! March 5th.

Tweet 1:

Negotiations on DACA have begun. Republicans want to make a deal and Democrats say they want to make a deal. Wouldn't it be great if we could finally, after so many years, solve the DACA puzzle. This will be our last chance, there will never be another opportunity! March 5th.

Tweet 2:

Negotiations on DACA have begun. Republicans want to make a deal and Democrats say they want to make a deal. Wouldn't it be great if we could finally, after so many years, solve the DACA puzzle. I will use my office to negotiate a fair deal for the dreamers! March 5th.

Latent Representation (Codebook) $\leadsto$ true whether hand coded, supervised, or unsupervised

Text-Based Intervention

- Assume "Interesting" treatments (coding) must be known in advance

Text-Based Intervention

- Assume "Interesting" treatments (coding) must be known in advance
- Discovery of treatments may (often/usually) happen after viewing data

Text-Based Intervention

- Assume "Interesting" treatments (coding) must be known in advance
- Discovery of treatments may (often/usually) happen after viewing data
- Explicit discovery phase in research

Automatically discover treatments

$+$

Estimate marginal effects

Three Key Steps

Three Key Steps

1) Theory: conditions to identify marginal effects of latent treatments (Average Marginal Component Effect (AMCE) is identified)

Three Key Steps

1) Theory: conditions to identify marginal effects of latent treatments (Average Marginal Component Effect (AMCE) is identified)
2) Method for discovering features (treatments)

Three Key Steps

1) Theory: conditions to identify marginal effects of latent treatments (Average Marginal Component Effect (AMCE) is identified)
2) Method for discovering features (treatments)
3) Method for estimating marginal effect for discovered features (treatments)

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments

- Let $\boldsymbol{Z}_i$ be i's binary feature vector

- Ex: $\boldsymbol{Z}_i = (0, 0, 1, 1, 0)$

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments
- Let $\boldsymbol{Z}_i$ be i's binary feature vector
- Ex: $\boldsymbol{Z}_i = (0, 0, 1, 1, 0)$

$$\text{AMCE}_k = \int_{\boldsymbol{Z}_{-k}} \mathbb{E}\left[Y(Z_k = 1, \boldsymbol{Z}_{-k}) - Y(Z_k = 0, \boldsymbol{Z}_{-k})\right] m(\boldsymbol{Z}_{-k}) d\boldsymbol{Z}_{-k}$$

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments
- Let $\boldsymbol{Z}_i$ be i's binary feature vector
- Ex: $\boldsymbol{Z}_i = (0, 0, 1, 1, 0)$

$$\text{AMCE}_k = \int_{\boldsymbol{Z}_{-k}} \mathbb{E}\left[Y(Z_k = 1, \boldsymbol{Z}_{-k}) - Y(Z_k = 0, \boldsymbol{Z}_{-k})\right] m(\boldsymbol{Z}_{-k}) d\boldsymbol{Z}_{-k}$$

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments
- Let $\boldsymbol{Z}_i$ be i's binary feature vector
- Ex: $\boldsymbol{Z}_i = (0, 0, 1, 1, 0)$

$$\text{AMCE}_k \quad = \quad \int_{\boldsymbol{Z}_{-k}} \mathbb{E}\left[Y(Z_k = 1, \boldsymbol{Z}_{-k}) - Y(Z_k = 0, \boldsymbol{Z}_{-k})\right] m(\boldsymbol{Z}_{-k}) d\boldsymbol{Z}_{-k}$$

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments
- Let $\boldsymbol{Z}_i$ be i's binary feature vector
- Ex: $\boldsymbol{Z}_i = (0, 0, 1, 1, 0)$

$$\text{AMCE}_k = \int_{\boldsymbol{Z}_{-k}} \mathbb{E}\left[Y(Z_k = 1, \boldsymbol{Z}_{-k}) - Y(Z_k = 0, \boldsymbol{Z}_{-k})\right] m(\boldsymbol{Z}_{-k}) d\boldsymbol{Z}_{-k}$$

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments
- Let $\boldsymbol{Z}_i$ be i's binary feature vector
- Ex: $\boldsymbol{Z}_i = (0, 0, 1, 1, 0)$

$$\text{AMCE}_k = \int_{\boldsymbol{Z}_{-k}} \mathbb{E}\left[Y(Z_k = 1, \boldsymbol{Z}_{-k}) - Y(Z_k = 0, \boldsymbol{Z}_{-k})\right] m(\boldsymbol{Z}_{-k}) d\boldsymbol{Z}_{-k}$$

Conjoint With Discovered Treatments

Identifying the Marginal Effects of Latent Treatments

- Average Marginal Component Effect (AMCE): Isolate effect of one treatment, averaging over other treatments
- Let $\boldsymbol{Z}_i$ be i's binary feature vector
- Ex: $\boldsymbol{Z}_i = (0, 0, 1, 1, 0)$

$$\text{AMCE}_k = \int_{\boldsymbol{Z}_{-k}} \mathbb{E}\left[Y(Z_k = 1, \boldsymbol{Z}_{-k}) - Y(Z_k = 0, \boldsymbol{Z}_{-k})\right] m(\boldsymbol{Z}_{-k}) d\boldsymbol{Z}_{-k}$$

Conjoint With Discovered Treatments (or) Discover Features that Drive Response in A/B Test

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function (assume known for now): text $\rightsquigarrow$ treatments in text ($\boldsymbol{Z}_i \equiv g(\boldsymbol{X}_i)$)

  $\boldsymbol{Z}_i$ is a low-dimensional rep of $\boldsymbol{X}_i$, describing treatments

Proposition 1

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function (assume known for now): text $\rightsquigarrow$ treatments in text ($\boldsymbol{Z}_i \equiv g(\boldsymbol{X}_i)$)

  $\boldsymbol{Z}_i$ is a low-dimensional rep of $\boldsymbol{X}_i$, describing treatments

Assume:

Proposition 1

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function (assume known for now): text $\rightsquigarrow$ treatments in text ($\boldsymbol{Z}_i \equiv g(\boldsymbol{X}_i)$)

  $\boldsymbol{Z}_i$ is a low-dimensional rep of $\boldsymbol{X}_i$, describing treatments

Assume:

1) No "spillover" (SUTVA, Rubin 1986: $\tilde{Y}_i(\boldsymbol{X}) = \tilde{Y}_i(\boldsymbol{X}_i)$)

Proposition 1

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function (assume known for now): text $\rightsquigarrow$ treatments in text ($\boldsymbol{Z}_i \equiv g(\boldsymbol{X}_i)$)

  $\boldsymbol{Z}_i$ is a low-dimensional rep of $\boldsymbol{X}_i$, describing treatments

Assume:

1) No "spillover" (SUTVA, Rubin 1986: $\tilde{Y}_i(\boldsymbol{X}) = \tilde{Y}_i(\boldsymbol{X}_i)$)
2) Random assignment of texts ($\tilde{Y}_i(\boldsymbol{X}_i) \perp\!\!\!\perp \boldsymbol{X}_i$ for all i)

Proposition 1

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function (assume known for now): text $\rightsquigarrow$ treatments in text ($\boldsymbol{Z}_i \equiv g(\boldsymbol{X}_i)$)

  $\boldsymbol{Z}_i$ is a low-dimensional rep of $\boldsymbol{X}_i$, describing treatments

Assume:

1) No "spillover" (SUTVA, Rubin 1986: $\tilde{Y}_i(\boldsymbol{X}) = \tilde{Y}_i(\boldsymbol{X}_i)$)
2) Random assignment of texts ($\tilde{Y}_i(\boldsymbol{X}_i) \perp\!\!\!\perp \boldsymbol{X}_i$ for all i)
3) Sufficiency: For all $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ such that $g(\boldsymbol{X}) = g(\boldsymbol{X}^{'})$ then $E[\tilde{Y}_i(\boldsymbol{X})] = E[\tilde{Y}_i(\boldsymbol{X}^{'})]$.

Proposition 1

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function (assume known for now): text $\rightsquigarrow$ treatments in text ($\boldsymbol{Z}_i \equiv g(\boldsymbol{X}_i)$)

  $\boldsymbol{Z}_i$ is a low-dimensional rep of $\boldsymbol{X}_i$, describing treatments

Assume:

1) No "spillover" (SUTVA, Rubin 1986: $\tilde{Y}_i(\boldsymbol{X}) = \tilde{Y}_i(\boldsymbol{X}_i)$)
2) Random assignment of texts ($\tilde{Y}_i(\boldsymbol{X}_i) \perp\!\!\!\perp \boldsymbol{X}_i$ for all i)
3) Sufficiency: For all $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ such that $g(\boldsymbol{X}) = g(\boldsymbol{X}^{'})$ then $E[\tilde{Y}_i(\boldsymbol{X})] = E[\tilde{Y}_i(\boldsymbol{X}^{'})]$.
4) Common support: all combinations of treatments have non-zero probability ($f(\boldsymbol{Z}_i) > 0$ for all $\boldsymbol{Z}_i \in$ Range $g(\cdot)$)

Proposition 1

Identifying the AMCE

- An individual sees a text ($\boldsymbol{X}_i$: text seen by i)
- Function (assume known for now): text $\rightsquigarrow$ treatments in text ($\boldsymbol{Z}_i \equiv g(\boldsymbol{X}_i)$)

  $\boldsymbol{Z}_i$ is a low-dimensional rep of $\boldsymbol{X}_i$, describing treatments

Assume:

1) No "spillover" (SUTVA, Rubin 1986: $\tilde{Y}_i(\boldsymbol{X}) = \tilde{Y}_i(\boldsymbol{X}_i)$)
2) Random assignment of texts ($\tilde{Y}_i(\boldsymbol{X}_i) \perp\!\!\!\perp \boldsymbol{X}_i$ for all i)
3) Sufficiency: For all $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ such that $g(\boldsymbol{X}) = g(\boldsymbol{X}^{'})$ then $E[\tilde{Y}_i(\boldsymbol{X})] = E[\tilde{Y}_i(\boldsymbol{X}^{'})]$.
4) Common support: all combinations of treatments have non-zero probability ($f(\boldsymbol{Z}_i) > 0$ for all $\boldsymbol{Z}_i \in$ Range $g(\cdot)$)

Proposition 1

Assumptions 1-4 are sufficient to identify the $AMCE_k$ for arbitrary k.

Discovering Treatments and Estimating Marginal Effects

Discovery of Treatments from Text Corpora

Discovery of Treatments from Text Corpora

1) (Assume) Randomly assign texts, $\boldsymbol{X}_i$, to respondents

Discovery of Treatments from Text Corpora

1) (Assume) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent

Discovery of Treatments from Text Corpora

1) (Assume) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) (Randomly) divide texts and responses into training and test set

Discovery of Treatments from Text Corpora

1) (Assume) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) (Randomly) divide texts and responses into training and test set
   a) Avoid technical issues with using entire sample (Analyst-induced SUTVA violations)

Discovery of Treatments from Text Corpora

1) (Assume) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) (Randomly) divide texts and responses into training and test set
    a) Avoid technical issues with using entire sample (Analyst-induced SUTVA violations)
    b) Ensure we avoid overfitting

Discovering Interesting Treatments

Discovering function from texts to treatments $g()$

Discovering Interesting Treatments

Discovering function from texts to treatments $g()$

- Use both documents and responses to discover the function

Discovering Interesting Treatments

Discovering function from texts to treatments $g()$

- Use both documents and responses to discover the function
- Topic and Supervised Topic models workhorse text models (Blei, Ng, and Jordan 2003; Blei and McAuliffe, 2007)

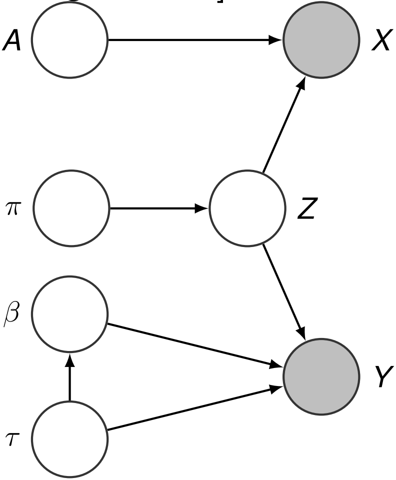
Discovering Interesting Treatments

Discovering function from texts to treatments $g()$

- Use both documents and responses to discover the function
- Topic and Supervised Topic models workhorse text models (Blei, Ng, and Jordan 2003; Blei and McAuliffe, 2007)

Treatments on simplex imply marginalization impossible $\rightsquigarrow$ increase in one category implies decrease in other category

The Supervised Indian Buffet Process (sIBP, distinct [though related] to Quadrianto et al 2013)



Text and response depend on latent treatments

- Treatment assignment

$$z_{i,k} \sim \text{Bernoulli}(\pi_k)$$
$$\pi_k \sim \prod_{m=1}^{k} \eta_m$$
$$\eta_m \sim \text{Beta}(\alpha, 1)$$

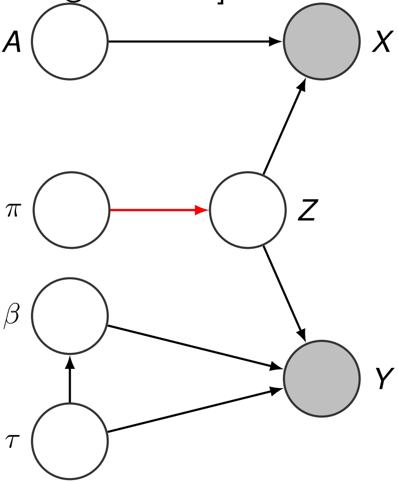
- Document Creation:

$$\boldsymbol{X}_i \sim \text{MVN}(\boldsymbol{Z}_i \boldsymbol{A}, \sigma_X^2 I_D)$$
$$\boldsymbol{A}_k \sim \text{MVN}(\boldsymbol{0}, \sigma_A^2 I_D)$$

- Response:

$$Y_i \sim \text{MVN}(Z_i \boldsymbol{\beta}, \tau^{-1})$$
$$\boldsymbol{\beta}|\tau \sim \text{MVN}(\boldsymbol{0}, \tau^{-1} I_K)$$
$$\tau \sim \text{Gamma}(a, b)$$

The Supervised Indian Buffet Process (sIBP, distinct [though related] to Quadrianto et al 2013)



Text and response depend on latent treatments

- Treatment assignment

$$
\begin{aligned}
z_{i,k} &\sim \text{Bernoulli}(\pi_k) \\
\pi_k &\sim \prod_{m=1}^{k} \eta_m \\
\eta_m &\sim \text{Beta}\left(\alpha, 1\right)
\end{aligned}
$$

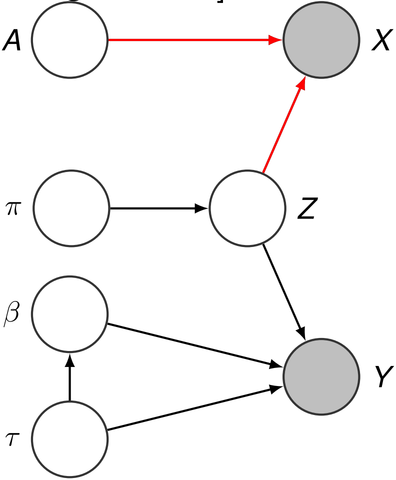
- Document Creation:

$$
\begin{aligned}
\boldsymbol{X}_i &\sim \text{MVN}(\boldsymbol{Z}_i \boldsymbol{A}, \sigma_X^2 I_D) \\
\boldsymbol{A}_k &\sim \text{MVN}(\boldsymbol{0}, \sigma_A^2 I_D)
\end{aligned}
$$

- Response:

$$
\begin{aligned}
Y_i &\sim \text{MVN}(Z_i \boldsymbol{\beta}, \tau^{-1}) \\
\boldsymbol{\beta}|\tau &\sim \text{MVN}(\boldsymbol{0}, \tau^{-1} I_K) \\
\tau &\sim \text{Gamma}(a, b)
\end{aligned}
$$

The Supervised Indian Buffet Process (sIBP, distinct [though related] to Quadrianto et al 2013)



Text and response depend on latent treatments

- Treatment assignment

$$z_{i,k} \sim \text{Bernoulli}(\pi_k)$$
$$\pi_k \sim \prod_{m=1}^{k} \eta_m$$
$$\eta_m \sim \text{Beta}(\alpha, 1)$$

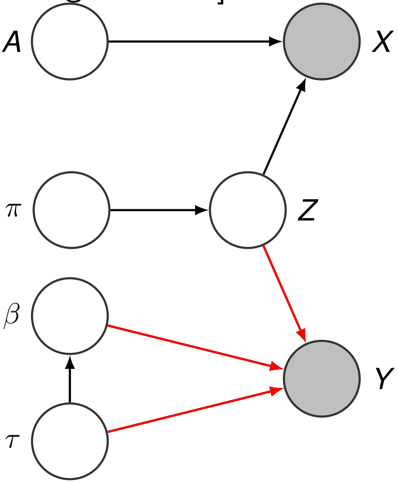
- Document Creation:

$$\boldsymbol{X}_i \sim \text{MVN}(\boldsymbol{Z}_i \boldsymbol{A}, \sigma_X^2 I_D)$$
$$\boldsymbol{A}_k \sim \text{MVN}(\mathbf{0}, \sigma_A^2 I_D)$$

- Response:

$$Y_i \sim \text{MVN}(Z_i \beta, \tau^{-1})$$
$$\beta|\tau \sim \text{MVN}(\mathbf{0}, \tau^{-1} I_K)$$
$$\tau \sim \text{Gamma}(a, b)$$

The Supervised Indian Buffet Process (sIBP, distinct [though related] to Quadrianto et al 2013)



Text and response depend on latent treatments

- Treatment assignment

$$\begin{aligned} z_{i,k} &\sim \text{Bernoulli}(\pi_k) \\ \pi_k &\sim \prod_{m=1}^{k} \eta_m \\ \eta_m &\sim \text{Beta}\,(\alpha, 1) \end{aligned}$$

- Document Creation:

$$\begin{aligned} \boldsymbol{X}_i &\sim \text{MVN}(\boldsymbol{Z}_i\boldsymbol{A}, \sigma_X^2 I_D) \\ \boldsymbol{A}_k &\sim \text{MVN}(\boldsymbol{0}, \sigma_A^2 I_D) \end{aligned}$$

- Response:

$$\begin{aligned} Y_i &\sim \text{MVN}(Z_i\boldsymbol{\beta}, \tau^{-1}) \\ \boldsymbol{\beta}|\tau &\sim \text{MVN}(\boldsymbol{0}, \tau^{-1} I_K) \\ \tau &\sim \text{Gamma}(a, b) \end{aligned}$$

Discovery of Treatments from Text Corpora

1) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) Divide texts and responses into training and test set
4) In training set: Discover mapping from texts to treatments

Discovery of Treatments from Text Corpora

1) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) Divide texts and responses into training and test set
4) In training set: Discover mapping from texts to treatments
    a) Apply supervised Indian Buffet Process (sIBP) to documents and responses to infer latent treatments in texts

Discovery of Treatments from Text Corpora

1) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) Divide texts and responses into training and test set
4) In training set: Discover mapping from texts to treatments
    a) Apply supervised Indian Buffet Process (sIBP) to documents and responses to infer latent treatments in texts
    b) Model selection via nonparametric process, quantitative fit, and qualitative assessment

Discovery of Treatments from Text Corpora

1) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) Divide texts and responses into training and test set
4) In training set: Discover mapping from texts to treatments
    a) Apply supervised Indian Buffet Process (sIBP) to documents and responses to infer latent treatments in texts
    b) Model selection via nonparametric process, quantitative fit, and qualitative assessment
5) In test set: infer treatments and measure their effect

Discovery of Treatments from Text Corpora

1) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) Divide texts and responses into training and test set
4) In training set: Discover mapping from texts to treatments
    a) Apply supervised Indian Buffet Process (sIBP) to documents and responses to infer latent treatments in texts
    b) Model selection via nonparametric process, quantitative fit, and qualitative assessment
5) In test set: infer treatments and measure their effect
    a) Use sIBP trained on training set to infer latent treatments on test set documents (without conditioning on test set responses)

Discovery of Treatments from Text Corpora

1) Randomly assign texts, $\boldsymbol{X}_i$, to respondents
2) Obtain responses $\boldsymbol{Y}_i$ for each respondent
3) Divide texts and responses into training and test set
4) In training set: Discover mapping from texts to treatments
   a) Apply supervised Indian Buffet Process (sIBP) to documents and responses to infer latent treatments in texts
   b) Model selection via nonparametric process, quantitative fit, and qualitative assessment
5) In test set: infer treatments and measure their effect
   a) Use sIBP trained on training set to infer latent treatments on test set documents (without conditioning on test set responses)
   b) Estimate effect of treatments with regression, with a bootstrap procedure to estimate uncertainty

Recovering Treatments from Heterogeneous Experiment

Recovering Treatments from Heterogeneous Experiment

- Credit Claiming Project (Grimmer, Messing and Westwood (2017)): Modular (large-factorial experiment) to assess credit claiming effects

Recovering Treatments from Heterogeneous Experiment

Credit Claiming Condition

Headline: Representative (redacted) |stageTitle |moneyTitle |typeTitle
Body: Representative (redacted), |partyMain, |alongMain |stageMain |moneyMain
Rep. (redacted) said "This money |stageQuote typeQuote"

|**stageTitle**:[will request/requested/secured]
|**moneyTitle**:[$50 thousand/$20 million]
|**typeTitle** : [to purchase safety equipment for local firefighters/to purchase sa
police/to repave local roads, to beautify local parks/for medical equipment at the loc
help build a state of the art gun range]
|**partyMain** : [Democrat/Republican]
|**alongMain** : [(No text)/and Senator (redacted), a Democrat/ and Senator (redac
|**stageMain** : [will request/requested/secured]
|**moneyMain**: [$50 thousand/ $20 million]
|**typeMain**: [to purchase safety equipment for local firefighters/to purchase safety
lice/to repave local roads, to beautify local parks/for medical equipment at the loca
help build a state of the art gun range]
|**stageQuote** : [would help/would help/will help]
|**typeQuote**: [our brave firefighters stay safe as they protect our businesses and
officers stay safe as they protect our property from criminals/keep our roads in saf
ensuring that our local economy will continue to grow/create parks that add valu

Recovering Treatments from Heterogeneous Experiment

Headline: Representative (redacted) secured $20 Million to purchase safety equipment for local firefighters
Body: Representative (redacted), Democrat, secured $20 Million to purchase safety equipment for local firefighters. Rep. (redacted) said "This money will help our brave firefighters stay safe as they protect our businesses and homes"

Recovering Treatments from Heterogeneous Experiment

- Credit Claiming Project (Grimmer, Messing and Westwood (2017)): Modular (large-factorial experiment) to assess credit claiming effects
- Train/Test split response, text as if natural language (50%/50%)

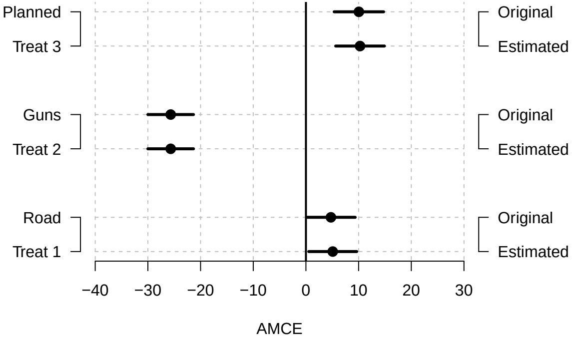
Recovering Treatments from Heterogeneous Experiment

- Credit Claiming Project (Grimmer, Messing and Westwood (2017)): Modular (large-factorial experiment) to assess credit claiming effects
- Train/Test split response, text as if natural language (50%/50%)

	Type Treatment From Experiment					
Discovered	Planned Parenthood	Parks	Gun Range	Roads	Police	Fire
Treat 1	0	4	0	127	0	0
Treat 2	0	0	122	0	0	0
Treat 3	119	4	0	0	0	0

Recovering Treatments from Heterogeneous Experiment

- Credit Claiming Project (Grimmer, Messing and Westwood (2017)): Modular (large-factorial experiment) to assess credit claiming effects
- Train/Test split response, text as if natural language (50%/50%)



Trump Tweets

`YouGov`: survey response to trump tweets

Trump Tweets

`YouGov`: survey response to trump tweets

Donald J. Trump
@realDonaldTrump

Following

Little Adam Schiff, who is desperate to run for higher office, is one of the biggest liars and leakers in Washington, right up there with Comey, Warner, Brennan and Clapper! Adam leaves closed committee hearings to illegally leak confidential information. Must be stopped!

4:39 AM - 5 Feb 2018

31,930 Retweets **99,706** Likes

48K 32K 100K

Trump Tweets

`YouGov`: survey response to trump tweets

- Survey Equal # Republicans, Democrats, Independents: read Trump tweet + evaluate (`Great, Good, OK, Bad, Terrible`)

Trump Tweets

`YouGov`: survey response to trump tweets

- Survey Equal # Republicans, Democrats, Independents: read Trump tweet + evaluate (`Great, Good, OK, Bad, Terrible`)
- Aggregate, create scale $[-200, 200]$

Trump Tweets

YouGov: survey response to trump tweets

- Survey Equal # Republicans, Democrats, Independents: read Trump tweet + evaluate (Great, Good, OK, Bad, Terrible)
- Aggregate, create scale $[-200, 200]$
- Modify sIBP: Shared treatments, discover heterogeneous effects

Trump Tweets

YouGov: survey response to trump tweets

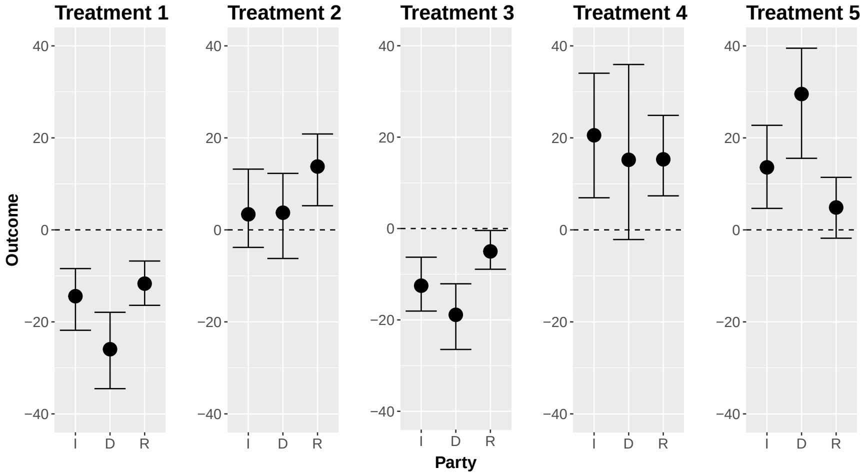
- Survey Equal # Republicans, Democrats, Independents: read Trump tweet + evaluate (Great, Good, OK, Bad, Terrible)
- Aggregate, create scale $[-200, 200]$
- Modify sIBP: Shared treatments, discover heterogeneous effects
- Train (66%), Test (33%), Clustered by tweet

Trump Tweets

`YouGov`: survey response to trump tweets

- Survey Equal # Republicans, Democrats, Independents: read Trump tweet + evaluate (`Great, Good, OK, Bad, Terrible`)
- Aggregate, create scale $[-200, 200]$
- Modify sIBP: Shared treatments, discover heterogeneous effects
- Train (66%), Test (33%), Clustered by tweet

Treatment 1	Treatment 2	Treatment 3	Treatment 4	Treatment 5
fake	cuts	obamacare	flotus	prime
news	strange	senators	behalf	minister
media	tax	repeal	anthem	korea
cnn	luther	healthcare	melania	north
election	stock	replace	nfl	stock
story	market	republican	flag	market
nbc	alabama	vote	prayers	china
stories	reform	republicans	bless	executive
hillary	record	senate	ready	prayers
clinton	high	north	players	order



Treatment 1
Treatment 2
Treatment 3
Treatment 4
Treatment 5
Outcome
40
20
0
−20
−40
I
D
R
Party

R Package: `textEffect`

Appendix

Formal Argument for Train/Test Split

Learn $g \rightsquigarrow \hat{g}$

Formal Argument for Train/Test Split

Learn $g \rightsquigarrow \hat{g}$

- $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ are randomizations, with $\boldsymbol{X}_i = \boldsymbol{X}_i^{'}$ (but with at least one $\boldsymbol{X}_j \neq \boldsymbol{X}_j^{'}$)

Formal Argument for Train/Test Split

Learn $g \rightsquigarrow \hat{g}$

- $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ are randomizations, with $\boldsymbol{X}_i = \boldsymbol{X}_i^{'}$ (but with at least one $\boldsymbol{X}_j \neq \boldsymbol{X}_j^{'}$)
- $\boldsymbol{Y}(\boldsymbol{X}), \boldsymbol{Y}(\boldsymbol{X}^{'})$ are potential outcomes.

Formal Argument for Train/Test Split

Learn $g \rightsquigarrow \hat{g}$

- $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ are randomizations, with $\boldsymbol{X}_i = \boldsymbol{X}_i^{'}$ (but with at least one $\boldsymbol{X}_j \neq \boldsymbol{X}_j^{'}$)
- $\boldsymbol{Y}(\boldsymbol{X}), \boldsymbol{Y}(\boldsymbol{X}^{'})$ are potential outcomes.

Define function:

Formal Argument for Train/Test Split

Learn $g \rightsquigarrow \hat{g}$

- $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ are randomizations, with $\boldsymbol{X}_i = \boldsymbol{X}_i^{'}$ (but with at least one $\boldsymbol{X}_j \neq \boldsymbol{X}_j^{'}$)
- $\boldsymbol{Y}(\boldsymbol{X}), \boldsymbol{Y}(\boldsymbol{X}^{'})$ are potential outcomes.

Define function:

$$\hat{g}(\boldsymbol{X}_i) \quad = \quad \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X}))$$

Formal Argument for Train/Test Split

Learn $g \rightsquigarrow \hat{g}$

- $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ are randomizations, with $\boldsymbol{X}_i = \boldsymbol{X}_i^{'}$ (but with at least one $\boldsymbol{X}_j \neq \boldsymbol{X}_j^{'}$)
- $\boldsymbol{Y}(\boldsymbol{X}), \boldsymbol{Y}(\boldsymbol{X}^{'})$ are potential outcomes.

Define function:

$$\hat{g}(\boldsymbol{X}_i) \quad = \quad \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X}))$$

Problem:

$$\hat{g}(\boldsymbol{X}_i) = \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X})) \quad \neq \quad \hat{g}(\boldsymbol{X}_i^{'}) = \hat{g}(\boldsymbol{X}_i^{'}, \boldsymbol{Y}(\boldsymbol{X}^{'}))$$

Formal Argument for Train/Test Split

Learn $g \rightsquigarrow \hat{g}$

- $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ are randomizations, with $\boldsymbol{X}_i = \boldsymbol{X}_i^{'}$ (but with at least one $\boldsymbol{X}_j \neq \boldsymbol{X}_j^{'}$)
- $\boldsymbol{Y}(\boldsymbol{X}), \boldsymbol{Y}(\boldsymbol{X}^{'})$ are potential outcomes.

Define function:

$$\hat{g}(\boldsymbol{X}_i) \quad = \quad \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X}))$$

Problem:

$$\hat{g}(\boldsymbol{X}_i) = \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X})) \quad \neq \quad \hat{g}(\boldsymbol{X}_i^{'}) = \hat{g}(\boldsymbol{X}_i^{'}, \boldsymbol{Y}(\boldsymbol{X}^{'}))$$

Solutions:

Formal Argument for Train/Test Split

Learn $g \rightsquigarrow \hat{g}$

- $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ are randomizations, with $\boldsymbol{X}_i = \boldsymbol{X}_i^{'}$ (but with at least one $\boldsymbol{X}_j \neq \boldsymbol{X}_j^{'}$)
- $\boldsymbol{Y}(\boldsymbol{X}), \boldsymbol{Y}(\boldsymbol{X}^{'})$ are potential outcomes.

Define function:

$$\hat{g}(\boldsymbol{X}_i) \quad = \quad \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X}))$$

Problem:

$$\hat{g}(\boldsymbol{X}_i) = \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X})) \quad \neq \quad \hat{g}(\boldsymbol{X}_i^{'}) = \hat{g}(\boldsymbol{X}_i^{'}, \boldsymbol{Y}(\boldsymbol{X}^{'}))$$

Solutions:

1) No responses/define before: $\hat{g}(\boldsymbol{X}_i) = \hat{g}(\boldsymbol{X}_i, \boldsymbol{X})$ Miss treatments

Formal Argument for Train/Test Split

Learn $g \rightsquigarrow \hat{g}$

- $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ are randomizations, with $\boldsymbol{X}_i = \boldsymbol{X}_i^{'}$ (but with at least one $\boldsymbol{X}_j \neq \boldsymbol{X}_j^{'}$)
- $\boldsymbol{Y}(\boldsymbol{X}), \boldsymbol{Y}(\boldsymbol{X}^{'})$ are potential outcomes.

Define function:

$$\hat{g}(\boldsymbol{X}_i) \quad = \quad \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X}))$$

Problem:

$$\hat{g}(\boldsymbol{X}_i) = \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X})) \quad \neq \quad \hat{g}(\boldsymbol{X}_i^{'}) = \hat{g}(\boldsymbol{X}_i^{'}, \boldsymbol{Y}(\boldsymbol{X}^{'}))$$

Solutions:

1) No responses/define before: $\hat{g}(\boldsymbol{X}_i) = \hat{g}(\boldsymbol{X}_i, \boldsymbol{X})$ Miss treatments
2) Assume Invariant to randomizations$\rightsquigarrow$ potential for fishing

Formal Argument for Train/Test Split

Learn $g \rightsquigarrow \hat{g}$

- $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ are randomizations, with $\boldsymbol{X}_i = \boldsymbol{X}_i^{'}$ (but with at least one $\boldsymbol{X}_j \neq \boldsymbol{X}_j^{'}$)
- $\boldsymbol{Y}(\boldsymbol{X}), \boldsymbol{Y}(\boldsymbol{X}^{'})$ are potential outcomes.

Define function:

$$\hat{g}(\boldsymbol{X}_i) \quad = \quad \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X}))$$

Problem:

$$\hat{g}(\boldsymbol{X}_i) = \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X})) \quad \neq \quad \hat{g}(\boldsymbol{X}_i^{'}) = \hat{g}(\boldsymbol{X}_i^{'}, \boldsymbol{Y}(\boldsymbol{X}^{'}))$$

Solutions:

1) No responses/define before: $\hat{g}(\boldsymbol{X}_i) = \hat{g}(\boldsymbol{X}_i, \boldsymbol{X})$ Miss treatments
2) Assume Invariant to randomizations$\rightsquigarrow$ potential for fishing
3) Train/Test Split: (Randomly) Divide responses into S_{train} and S_{test}

Formal Argument for Train/Test Split

Learn $g \rightsquigarrow \hat{g}$

- $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ are randomizations, with $\boldsymbol{X}_i = \boldsymbol{X}^{'}_i$ (but with at least one $\boldsymbol{X}_j \neq \boldsymbol{X}^{'}_j$)
- $\boldsymbol{Y}(\boldsymbol{X}), \boldsymbol{Y}(\boldsymbol{X}^{'})$ are potential outcomes.

Define function:

$$\hat{g}(\boldsymbol{X}_i) \quad = \quad \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X}))$$

Problem:

$$\hat{g}(\boldsymbol{X}_i) = \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X})) \quad \neq \quad \hat{g}(\boldsymbol{X}^{'}_i) = \hat{g}(\boldsymbol{X}^{'}_i, \boldsymbol{Y}(\boldsymbol{X}^{'}))$$

Solutions:

1) No responses/define before: $\hat{g}(\boldsymbol{X}_i) = \hat{g}(\boldsymbol{X}_i, \boldsymbol{X})$ Miss treatments
2) Assume Invariant to randomizations$\rightsquigarrow$ potential for fishing
3) Train/Test Split: (Randomly) Divide responses into S_{train} and S_{test}
    - Learn g using S_{train} (Analgous to pre-test)

Formal Argument for Train/Test Split

Learn $g \rightsquigarrow \hat{g}$

- $\boldsymbol{X}$ and $\boldsymbol{X}^{'}$ are randomizations, with $\boldsymbol{X}_i = \boldsymbol{X}_i^{'}$ (but with at least one $\boldsymbol{X}_j \neq \boldsymbol{X}_j^{'}$)
- $\boldsymbol{Y}(\boldsymbol{X}), \boldsymbol{Y}(\boldsymbol{X}^{'})$ are potential outcomes.

Define function:

$$\hat{g}(\boldsymbol{X}_i) \quad = \quad \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X}))$$

Problem:

$$\hat{g}(\boldsymbol{X}_i) = \hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X})) \quad \neq \quad \hat{g}(\boldsymbol{X}_i^{'}) = \hat{g}(\boldsymbol{X}_i^{'}, \boldsymbol{Y}(\boldsymbol{X}^{'}))$$

Solutions:

1) No responses/define before: $\hat{g}(\boldsymbol{X}_i) = \hat{g}(\boldsymbol{X}_i, \boldsymbol{X})$ Miss treatments
2) Assume Invariant to randomizations$\rightsquigarrow$ potential for fishing
3) Train/Test Split: (Randomly) Divide responses into S_{train} and S_{test}
   - Learn g using S_{train} (Analgous to pre-test)
   - Infer treatment in S_{test} $\hat{g}(\boldsymbol{X}_i, \boldsymbol{Y}(\boldsymbol{X})_{\text{train}})$