Bayesian Imputation for Potential Outcomes

Ryan T. Moore

Graduate Methods Master Class
11 February 2005
Cambridge MA

Fundamental Problem of Causal Inference

Fundamental Problem of Causal Inference

Since applying both treatment and control to a particular unit at a particular point in time is not possible, we must make assumptions about the unobserved potential outcomes.

Fundamental Problem of Causal Inference

Since applying both treatment and control to a particular unit at a particular point in time is not possible, we must make assumptions about the unobserved potential outcomes. For the treated group, this means imputing the potential outcome under control. For the control group, this means imputing the potential outcome under treatment.

Fundamental Problem of Causal Inference

Since applying both treatment and control to a particular unit at a particular point in time is not possible, we must make assumptions about the unobserved potential outcomes. For the treated group, this means imputing the potential outcome under control. For the control group, this means imputing the potential outcome under treatment.

Unit	X	T	Y(1)	Y(0)
1	...	1	1	?
2	...	1	4	?
3	...	1	3	?
4	...	0	?	10
5	...	0	?	1
6	...	0	?	5

Alternatives for Imputing Missing Potential Outcomes

Alternatives for Imputing Missing Potential Outcomes

- Assume the treatment effect is zero, and fill in the observed potential outcome for each unobserved potential outcome:

Alternatives for Imputing Missing Potential Outcomes

- Assume the treatment effect is zero, and fill in the observed potential outcome for each unobserved potential outcome:

Unit	X	T	Y(1)	Y(0)
1	...	1	1	*1*
2	...	1	4	*4*
3	...	1	3	*3*
4	...	0	*10*	10
5	...	0	*1*	1
6	...	0	*5*	5

Alternatives for Imputing Missing Potential Outcomes

- Assume the treatment effect is zero, and fill in the observed potential outcome for each unobserved potential outcome:

Unit	X	T	Y(1)	Y(0)
1	...	1	1	*1*
2	...	1	4	*4*
3	...	1	3	*3*
4	...	0	*10*	10
5	...	0	*1*	1
6	...	0	*5*	5

This defines the causal effect we calculate as zero, but permits calculating the randomization p-value.

Alternatives for Imputing Missing Potential Outcomes

Alternatives for Imputing Missing Potential Outcomes

- Perform some sort of matching, and assume that matched pairs have identical potential outcomes:

Alternatives for Imputing Missing Potential Outcomes

- Perform some sort of matching, and assume that matched pairs have identical potential outcomes:

Unit	X	T	Y(1)	Y(0)
1	. . .	1	1	*10*
2	. . .	1	4	*1*
3	. . .	1	3	*5*
4	. . .	0	*1*	10
5	. . .	0	*4*	1
6	. . .	0	*3*	5

Alternatives for Imputing Missing Potential Outcomes

- Perform some sort of matching, and assume that matched pairs have identical potential outcomes:

Unit	X	T	Y(1)	Y(0)
1	...	1	1	*10*
2	...	1	4	*1*
3	...	1	3	*5*
4	...	0	*1*	10
5	...	0	*4*	1
6	...	0	*3*	5

Here the causal effect is estimated to be $-\frac{8}{3}$.

Alternatives for Imputing Missing Potential Outcomes

Alternatives for Imputing Missing Potential Outcomes

- Estimate a model using the covariates and treatment indicator to predict the missing potential outcome.

Alternatives for Imputing Missing Potential Outcomes

- Estimate a model using the covariates and treatment indicator to predict the missing potential outcome. Let's just do a linear model, but let's sample in a Bayesian way.

Alternatives for Imputing Missing Potential Outcomes

- Estimate a model using the covariates and treatment indicator to predict the missing potential outcome. Let's just do a linear model, but let's sample in a Bayesian way.
- To do so:

Alternatives for Imputing Missing Potential Outcomes

- Estimate a model using the covariates and treatment indicator to predict the missing potential outcome. Let's just do a linear model, but let's sample in a Bayesian way.

- To do so:
  - Calculate the usual OLS regression coefficients.

Alternatives for Imputing Missing Potential Outcomes

- Estimate a model using the covariates and treatment indicator to predict the missing potential outcome. Let's just do a linear model, but let's sample in a Bayesian way.

- To do so:
  - Calculate the usual OLS regression coefficients.
  - Draw a value for σ^2 from some prior (uninformative, if we prefer).

Alternatives for Imputing Missing Potential Outcomes

- Estimate a model using the covariates and treatment indicator to predict the missing potential outcome. Let's just do a linear model, but let's sample in a Bayesian way.

- To do so:
  - Calculate the usual OLS regression coefficients.
  - Draw a value for σ^2 from some prior (uninformative, if we prefer).
  - Draw a value for $\sigma^2 | Y_{obs}, X$ from its posterior.

Alternatives for Imputing Missing Potential Outcomes

- Estimate a model using the covariates and treatment indicator to predict the missing potential outcome. Let's just do a linear model, but let's sample in a Bayesian way.

- To do so:
  - Calculate the usual OLS regression coefficients.
  - Draw a value for σ^2 from some prior (uninformative, if we prefer).
  - Draw a value for $\sigma^2 | Y_{obs}, X$ from its posterior.
  - Draw a vector $\beta | \sigma^2, Y_{obs}, X$ from $\beta \sim MVN(\hat{\beta}_{OLS}, \sigma^2 (X'X)^{-1})$. (A flat prior on β generates the MVN, since the MVN is centered at $\hat{\beta}_{OLS}$).

Alternatives for Imputing Missing Potential Outcomes

- Estimate a model using the covariates and treatment indicator to predict the missing potential outcome. Let's just do a linear model, but let's sample in a Bayesian way.

- To do so:
  - Calculate the usual OLS regression coefficients.
  - Draw a value for σ^2 from some prior (uninformative, if we prefer).
  - Draw a value for $\sigma^2 | Y_{obs}, X$ from its posterior.
  - Draw a vector $\beta | \sigma^2, Y_{obs}, X$ from $\beta \sim MVN(\hat{\beta}_{OLS}, \sigma^2 (X'X)^{-1})$. (A flat prior on β generates the MVN, since the MVN is centered at $\hat{\beta}_{OLS}$).
  - Find the missing potential outcomes for the treated by calculating $[X\ 0]\beta + \hat{\epsilon}$, where $\hat{\epsilon} \sim N(0, \sigma^2)$.

Alternatives for Imputing Missing Potential Outcomes

- Estimate a model using the covariates and treatment indicator to predict the missing potential outcome. Let's just do a linear model, but let's sample in a Bayesian way.

- To do so:
  - Calculate the usual OLS regression coefficients.
  - Draw a value for σ^2 from some prior (uninformative, if we prefer).
  - Draw a value for $\sigma^2 | Y_{obs}, X$ from its posterior.
  - Draw a vector $\beta | \sigma^2, Y_{obs}, X$ from $\beta \sim MVN(\hat{\beta}_{OLS}, \sigma^2(X'X)^{-1})$. (A flat prior on β generates the MVN, since the MVN is centered at $\hat{\beta}_{OLS}$).
  - Find the missing potential outcomes for the treated by calculating $[X\ 0]\beta + \hat{\epsilon}$, where $\hat{\epsilon} \sim N(0, \sigma^2)$.
  - Find the missing potential outcomes for the control by calculating $[X\ 1]\beta + \hat{\epsilon}$, where $\hat{\epsilon} \sim N(0, \sigma^2)$.

Alternatives for Imputing Missing Potential Outcomes

- Estimate a model using the covariates and treatment indicator to predict the missing potential outcome. Let's just do a linear model, but let's sample in a Bayesian way.

- To do so:
  - Calculate the usual OLS regression coefficients.
  - Draw a value for σ^2 from some prior (uninformative, if we prefer).
  - Draw a value for $\sigma^2 | Y_{obs}, X$ from its posterior.
  - Draw a vector $\beta | \sigma^2, Y_{obs}, X$ from $\beta \sim MVN(\hat{\beta}_{OLS}, \sigma^2 (X'X)^{-1})$. (A flat prior on β generates the MVN, since the MVN is centered at $\hat{\beta}_{OLS}$).
  - Find the missing potential outcomes for the treated by calculating $[X\ 0]\beta + \hat{\epsilon}$, where $\hat{\epsilon} \sim N(0, \sigma^2)$.
  - Find the missing potential outcomes for the control by calculating $[X\ 1]\beta + \hat{\epsilon}$, where $\hat{\epsilon} \sim N(0, \sigma^2)$.
  - Calculate the ATE: $E[Y(1) - Y(0)]$.

Alternatives for Imputing Missing Potential Outcomes

- Estimate a model using the covariates and treatment indicator to predict the missing potential outcome. Let's just do a linear model, but let's sample in a Bayesian way.

- To do so:
  - Calculate the usual OLS regression coefficients.
  - Draw a value for σ^2 from some prior (uninformative, if we prefer).
  - Draw a value for $\sigma^2 | Y_{obs}, X$ from its posterior.
  - Draw a vector $\beta | \sigma^2, Y_{obs}, X$ from $\beta \sim MVN(\hat{\beta}_{OLS}, \sigma^2 (X'X)^{-1})$. (A flat prior on β generates the MVN, since the MVN is centered at $\hat{\beta}_{OLS}$).
  - Find the missing potential outcomes for the treated by calculating $[X\ 0]\beta + \hat{\epsilon}$, where $\hat{\epsilon} \sim N(0, \sigma^2)$.
  - Find the missing potential outcomes for the control by calculating $[X\ 1]\beta + \hat{\epsilon}$, where $\hat{\epsilon} \sim N(0, \sigma^2)$.
  - Calculate the ATE: $E[Y(1) - Y(0)]$.
  - Repeat this process k times for posterior estimates of β, SE_β, σ^2, the ATE, and the SE_{ATE}.

Unit	X	T	Y(1)	Y(0)
1	. . .	1	1	*5.33*
2	. . .	1	4	*5.33*
3	. . .	1	3	*5.33*
4	. . .	0	*2.67*	10
5	. . .	0	*2.67*	1
6	. . .	0	*2.67*	5

(Here I use only the treatment indicator as a regressor, so by construction I find the ATE is equal to the coefficient on the treatment indicator.)

Alternatives for Imputing Missing Potential Outcomes

Alternatives for Imputing Missing Potential Outcomes

- This time, we relax the constraint that the treatment effect is equal for the treatment and control groups.

Alternatives for Imputing Missing Potential Outcomes

- This time, we relax the constraint that the treatment effect is equal for the treatment and control groups.
- Estimate two different models, one for treatment, one for control.

Alternatives for Imputing Missing Potential Outcomes

- This time, we relax the constraint that the treatment effect is equal for the treatment and control groups.
- Estimate two different models, one for treatment, one for control. β_{tr} estimates how the covariates affect the outcome for the X values of the treated group.

Alternatives for Imputing Missing Potential Outcomes

- This time, we relax the constraint that the treatment effect is equal for the treatment and control groups.

- Estimate two different models, one for treatment, one for control. β_{tr} estimates how the covariates affect the outcome for the X values of the treated group. β_{co} estimates how the covariates affect the outcome for the X values of the treated group.

Alternatives for Imputing Missing Potential Outcomes

- This time, we relax the constraint that the treatment effect is equal for the treatment and control groups.

- Estimate two different models, one for treatment, one for control. β_{tr} estimates how the covariates affect the outcome for the X values of the treated group. β_{co} estimates how the covariates affect the outcome for the X values of the treated group. Especially if the X values for the two groups do not overlap, we may expect and estimate very different values for β_{tr} and β_{co}.

Alternatives for Imputing Missing Potential Outcomes

- This time, we relax the constraint that the treatment effect is equal for the treatment and control groups.

- Estimate two different models, one for treatment, one for control. β_{tr} estimates how the covariates affect the outcome for the X values of the treated group. β_{co} estimates how the covariates affect the outcome for the X values of the treated group. Especially if the X values for the two groups do not overlap, we may expect and estimate very different values for β_{tr} and β_{co}.

- We generate β_{tr} and β_{co} vectors as before.

Alternatives for Imputing Missing Potential Outcomes

- This time, we relax the constraint that the treatment effect is equal for the treatment and control groups.

- Estimate two different models, one for treatment, one for control. β_{tr} estimates how the covariates affect the outcome for the X values of the treated group. β_{co} estimates how the covariates affect the outcome for the X values of the treated group. Especially if the X values for the two groups do not overlap, we may expect and estimate very different values for β_{tr} and β_{co}.

- We generate β_{tr} and β_{co} vectors as before.

- Find the missing potential outcomes for the treated by calculating $[X_{\mathrm{tr}}]\beta_{\mathrm{co}}$.

Alternatives for Imputing Missing Potential Outcomes

- This time, we relax the constraint that the treatment effect is equal for the treatment and control groups.
- Estimate two different models, one for treatment, one for control. β_{tr} estimates how the covariates affect the outcome for the X values of the treated group. β_{co} estimates how the covariates affect the outcome for the X values of the treated group. Especially if the X values for the two groups do not overlap, we may expect and estimate very different values for β_{tr} and β_{co}.
- We generate β_{tr} and β_{co} vectors as before.
- Find the missing potential outcomes for the treated by calculating $[X_{tr}]\beta_{co}$.
- Find the missing potential outcomes for the control by calculating $[X_{co}]\beta_{tr}$.

Alternatives for Imputing Missing Potential Outcomes

- This time, we relax the constraint that the treatment effect is equal for the treatment and control groups.
- Estimate two different models, one for treatment, one for control. β_{tr} estimates how the covariates affect the outcome for the X values of the treated group. β_{co} estimates how the covariates affect the outcome for the X values of the treated group. Especially if the X values for the two groups do not overlap, we may expect and estimate very different values for β_{tr} and β_{co}.
- We generate β_{tr} and β_{co} vectors as before.
- Find the missing potential outcomes for the treated by calculating $[X_{tr}]\beta_{co}$.
- Find the missing potential outcomes for the control by calculating $[X_{co}]\beta_{tr}$.
- Calculate the ATE: $E[Y(1) - Y(0)]$.

Alternatives for Imputing Missing Potential Outcomes

- This time, we relax the constraint that the treatment effect is equal for the treatment and control groups.
- Estimate two different models, one for treatment, one for control. β_{tr} estimates how the covariates affect the outcome for the X values of the treated group. β_{co} estimates how the covariates affect the outcome for the X values of the treated group. Especially if the X values for the two groups do not overlap, we may expect and estimate very different values for β_{tr} and β_{co}.
- We generate β_{tr} and β_{co} vectors as before.
- Find the missing potential outcomes for the treated by calculating $[X_{tr}]\beta_{co}$.
- Find the missing potential outcomes for the control by calculating $[X_{co}]\beta_{tr}$.
- Calculate the ATE: $E[Y(1) - Y(0)]$.
- Repeat this process k times for posterior estimates of the two β's, and the ATE.

Unit	X	T	Y(1)	Y(0)
1	. . .	1	1	*5.33*
2	. . .	1	4	*5.33*
3	. . .	1	3	*5.33*
4	. . .	0	*2.67*	10
5	. . .	0	*2.67*	1
6	. . .	0	*2.67*	5

(Here I use only an intercept as the regressor in the two regressions, since the treatment indicator is perfectly colinear with a column of ones. Thus, I get the same results as before. In general, with substantive covariates, this will not be the case.)

An Illustration

An Illustration

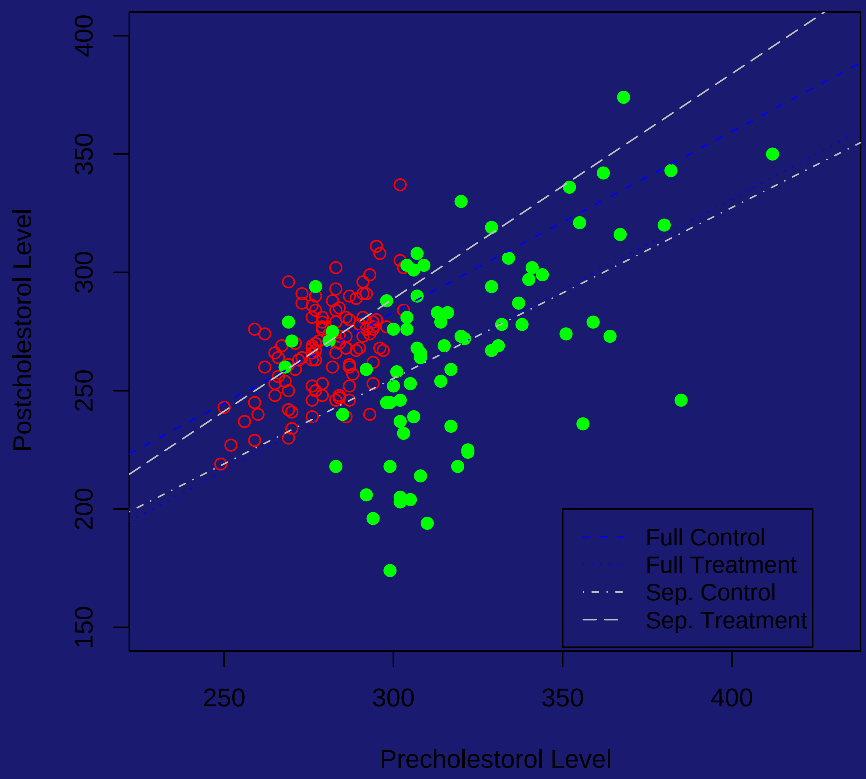
In the figure below, the pretreatment covariate is "Precholestorol Level" and the outcome is "Postcholestorol Level".

An Illustration

In the figure below, the pretreatment covariate is "Precholestorol Level" and the outcome is "Postcholestorol Level". The treatment and control groups only partially overlap on the covariate.

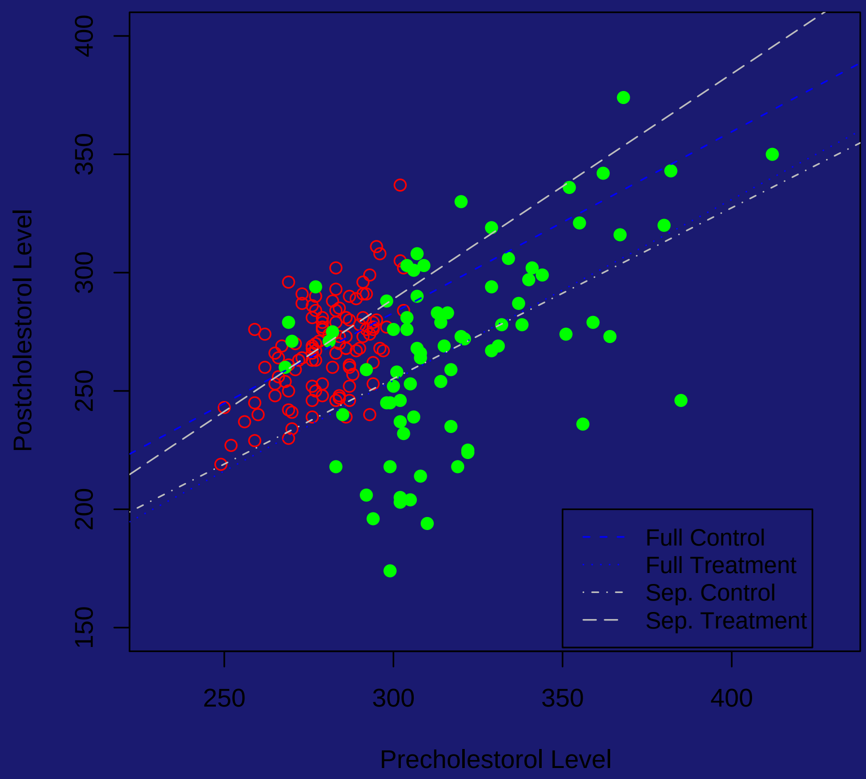
An Illustration

In the figure below, the pretreatment covariate is “Precholestorol Level” and the outcome is “Postcholestorol Level”. The treatment and control groups only partially overlap on the covariate.



An Illustration

In the figure below, the pretreatment covariate is "Precholestorol Level" and the outcome is "Postcholestorol Level". The treatment and control groups only partially overlap on the covariate.



Treated (green disks) and Control (red circles). Full and separate regression lines indicate that the ATE as modeled here does not appear to have the same relationship in the two subpopulations. "Full" lines indicate single regression; "Sep" lines indicate separate regressions for the treatment and control groups.

Final Thoughts

Final Thoughts

- Causal inference requires imputation of missing potential outcomes.

Final Thoughts

- Causal inference requires imputation of missing potential outcomes.
- Choices must be made about how imputation will occur.

Final Thoughts

- Causal inference requires imputation of missing potential outcomes.
- Choices must be made about how imputation will occur.
- Causal estimates can best be attributed to treatment when covariates of treatment and control groups overlap.

Final Thoughts

- Causal inference requires imputation of missing potential outcomes.
- Choices must be made about how imputation will occur.
- Causal estimates can best be attributed to treatment when covariates of treatment and control groups overlap.
- Especially when covariates don't overlap, estimate relationships between covariates and treatment and outcomes as (possibly) different in treatment and control groups.

Final Thoughts

- Causal inference requires imputation of missing potential outcomes.
- Choices must be made about how imputation will occur.
- Causal estimates can best be attributed to treatment when covariates of treatment and control groups overlap.
- Especially when covariates don't overlap, estimate relationships between covariates and treatment and outcomes as (possibly) different in treatment and control groups.
- Bayesian & MCMC methods allow for sampling from distributions that may be mathematically intractable, that include multilevel relationships, and that include large numbers of parameters.

Final Thoughts

- Causal inference requires imputation of missing potential outcomes.
- Choices must be made about how imputation will occur.
- Causal estimates can best be attributed to treatment when covariates of treatment and control groups overlap.
- Especially when covariates don't overlap, estimate relationships between covariates and treatment and outcomes as (possibly) different in treatment and control groups.
- Bayesian & MCMC methods allow for sampling from distributions that may be mathematically intractable, that include multilevel relationships, and that include large numbers of parameters.

- Next week: Alexis Diamond on matching algorithms, genetic code, etc. (Early start due to teaching.)

Final Thoughts

- Causal inference requires imputation of missing potential outcomes.
- Choices must be made about how imputation will occur.
- Causal estimates can best be attributed to treatment when covariates of treatment and control groups overlap.
- Especially when covariates don't overlap, estimate relationships between covariates and treatment and outcomes as (possibly) different in treatment and control groups.
- Bayesian & MCMC methods allow for sampling from distributions that may be mathematically intractable, that include multilevel relationships, and that include large numbers of parameters.

- Next week: Alexis Diamond on matching algorithms, genetic code, etc. (Early start due to teaching.)
- Then: Diff-in-diffs & IV, Selection models and alternatives, EI intro and cutting edge, time series, statistical graphics, zero-inflated models, counterfactual analysis and the convex hull, . . .