Univariate Matching

September 27, 2012

Mechanical vs Scientific Tasks

- Two tasks for matching inference:

Mechanical vs Scientific Tasks

- Two tasks for matching inference:
  1. Create the matched strata.

Mechanical vs Scientific Tasks

- Two tasks for matching inference:
  1. Create the matched strata.
  2. Do these matches look comparable?

Mechanical vs Scientific Tasks

- Two tasks for matching inference:
  1. Create the matched strata.
  2. Do these matches look comparable?

- Ultimately, we are asking if our mechanical operations are sufficient for identification of a treatment effect.

Mechanical vs Scientific Tasks

- Two tasks for matching inference:
  1. Create the matched strata.
  2. Do these matches look comparable?

- Ultimately, we are asking if our mechanical operations are sufficient for identification of a treatment effect.
- As Rosenbaum says, "The second task is not a mechanical but rather a scientific task, one that can be controversial and difficult to bring to a rapid and definitive closure; this task is, therefore, more challenging, and hence more interesting."

Notation

- Let $\mathbf{X}_i$ refer to a set of observed covariates for person i

Notation

- Let $\mathbf{X}_i$ refer to a set of observed covariates for person i
- let u_i refer to an unobserved covariate for subject i

Notation

- Let $\mathbf{X}_i$ refer to a set of observed covariates for person i
- let u_i refer to an unobserved covariate for subject i
- T_i refers to treatment assignment (1 for treatment, 0 for control)

Notation

- Let $\mathbf{X}_i$ refer to a set of observed covariates for person i
- let u_i refer to an unobserved covariate for subject i
- T_i refers to treatment assignment (1 for treatment, 0 for control)
- Y_{i1} refers to potential outcome under treatment, and Y_{i0} refers to potential outcome under control

Notation

- Let $\mathbf{X}_i$ refer to a set of observed covariates for person i
- let u_i refer to an unobserved covariate for subject i
- T_i refers to treatment assignment (1 for treatment, 0 for control)
- Y_{i1} refers to potential outcome under treatment, and Y_{i0} refers to potential outcome under control
- There are N subjects

A Model of Treatment Assignment

- In the population before matching, subject i received treatment with probability π_i, independently of other subjects, where π_i may vary from one person to the next and is *unknown*.

  More precisely:

$$\pi_i = Pr(T_i = 1 | Y_{i1}, Y_{i0}, \mathbf{X}_i, u_i)$$

  Assuming SUTVA, the joint probability distribution is:

$$Pr(T_1 = t_1, \ldots, T_N = t_n | Y_{11}, Y_{10}, \mathbf{X}_1, u_1, \ldots, Y_{n1}, Y_{n0}, \mathbf{X}_n, u_n)$$
$$= \prod_{i=1}^{N} \pi_i^{t_i} (1 - \pi_i)^{1 - t_i}$$

The Ideal Match

- Find two subjects, k and j, with $T_k + T_j = 1$

The Ideal Match

- Find two subjects, k and j, with $T_k + T_j = 1$
- If $\pi_k = \pi_j$, we can pair these two subjects and call them a matched-pair $\{k, j\}$

The Ideal Match

- Find two subjects, k and j, with $T_k + T_j = 1$
- If $\pi_k = \pi_j$, we can pair these two subjects and call them a matched-pair $\{k, j\}$

- (Note, we impose $0 < \pi_j < 1$ by assumption, and $\{k, j\} \in S$, for our S sample)

The Ideal Match

- Find two subjects, k and j, with $T_k + T_j = 1$
- If $\pi_k = \pi_j$, we can pair these two subjects and call them a matched-pair $\{k, j\}$

- (Note, we impose $0 < \pi_j < 1$ by assumption, and $\{k, j\} \in S$, for our S sample)
- We don't observe u_k or u_j, and we either observe Y_{k1} or Y_{j1} (but not both) and either Y_{k0} or Y_{j0}.

The Ideal Match

- Find two subjects, k and j, with $T_k + T_j = 1$
- If $\pi_k = \pi_j$, we can pair these two subjects and call them a matched-pair $\{k, j\}$

- (Note, we impose $0 < \pi_j < 1$ by assumption, and $\{k, j\} \in S$, for our S sample)
- We don't observe u_k or u_j, and we either observe Y_{k1} or Y_{j1} (but not both) and either Y_{k0} or Y_{j0}.

- Create a matched pair with $\pi_k = \pi_j$ and $T_k + T_j = 1$, then what would this give us?

Treatment Odds

$$Pr(T_k = 1, T_j = 0 | Y_{k1}, Y_{k0}, \mathbf{X}_k, u_k, Y_{j1}, Y_{j0}, \mathbf{X}_j, u_j, T_k + T_j = 1)$$

$$= \frac{Pr(T_k = 1, T_j = 0 | Y_{k1}, Y_{k0}, \mathbf{X}_k, u_k, Y_{j1}, Y_{j0}, \mathbf{X}_j, u_j)}{Pr(T_k + T_j = 1 | Y_{k1}, Y_{k0}, \mathbf{X}_k, u_k, Y_{j1}, Y_{j0}, \mathbf{X}_j, u_j)}$$

$$= \frac{\pi_k^{1+0}(1-\pi_k)^{(1-1)+(1-0)}}{\pi_k^{1+0}(1-\pi_k)^{(1-1)+(1-0)} + \pi_j^{0+1}(1-\pi_j)^{(1-0)+(1-1)}}$$

$$= \frac{\pi_k(1-\pi_k)}{\pi_k(1-\pi_k) + \pi_j(1-\pi_j)} = \frac{1}{2}, \quad \text{since } \pi_k = \pi_j$$

Treatment Odds

- It is important to note that this is always true if treatment is assigned by the fair flip of a fair coin
- OR independent flips of a group of biased coins where the same biased coin is used when subject k and subject j have the same observable characteristics (assuming no coins have a probability of 0 or 1)

Exact Matching

- If the naive model is true, then exactly match on $\mathbf{X}$
- Dimensionality problem with multivariate $\mathbf{X}$ in finite samples

Propensity Score

- The propensity score is defined as the conditional probability of treatment, $T = 1$ given the observed covariates $\mathbf{X}$.

$$e(\mathbf{X}) = Pr(T = 1|\mathbf{X})$$

Propensity Score

- The propensity score is defined as the conditional probability of treatment, $T = 1$ given the observed covariates $\mathbf{X}$.

$$e(\mathbf{X}) = Pr(T = 1|\mathbf{X})$$

- Balancing property on $\mathbf{X}$ is always true (with A2)
  - T and $\mathbf{X}$ are conditionally independent given the propensity score.

$$Pr\{\mathbf{X}|T = 1, e(\mathbf{X})\} = Pr\{\mathbf{X}|T = 0, e(\mathbf{X})\} \Leftrightarrow T \perp\!\!\!\perp \mathbf{X}|e(\mathbf{X})$$

  - Treated and control units with the same propensity score have the same distribution of the *observed* characteristics.

Propensity Score

- The propensity score is defined as the conditional probability of treatment, $T = 1$ given the observed covariates $\mathbf{X}$.

$$e(\mathbf{X}) = Pr(T = 1|\mathbf{X})$$

- Balancing property on $\mathbf{X}$ is always true (with A2)
  - T and $\mathbf{X}$ are conditionally independent given the propensity score.

$$Pr\{\mathbf{X}|T = 1, e(\mathbf{X})\} = Pr\{\mathbf{X}|T = 0, e(\mathbf{X})\} \Leftrightarrow T \perp\!\!\!\perp \mathbf{X}|e(\mathbf{X})$$

  - Treated and control units with the same propensity score have the same distribution of the *observed* characteristics.
- Within a given matched pair, it is not necessary that subject k and subject j have the same values of $\mathbf{X}$, only that they have the same propensity score, $e(\mathbf{X}_k) = e(\mathbf{X}_j)$.

Propensity Score

- We often estimate the propensity score, coming up with an estimate $\hat{e}(\mathbf{X})$ to produce balance on the observed covariates $\mathbf{X}$.

Propensity Score

- We often estimate the propensity score, coming up with an estimate $\hat{e}(\mathbf{X})$ to produce balance on the observed covariates $\mathbf{X}$.
- If the naive model *were* true, then from the propensity score we could get ignorable treatment assignment. We could produce the "ideal match" from the propensity score, since it just reduces our dimensionality of $\mathbf{X}$.

Propensity Score

- We often estimate the propensity score, coming up with an estimate $\hat{e}(\mathbf{X})$ to produce balance on the observed covariates $\mathbf{X}$.
- If the naive model *were* true, then from the propensity score we could get ignorable treatment assignment. We could produce the "ideal match" from the propensity score, since it just reduces our dimensionality of $\mathbf{X}$.
- If the naive model holds, then $\pi_i = e(\mathbf{X})$, so matching on the propensity score is matching on π_i. In the naive model:

$$T \perp\!\!\!\perp Y_{i1}, Y_{i0}, u_i | \mathbf{X} \Rightarrow T \perp\!\!\!\perp Y_{i1}, Y_{i0}, u_i | e(\mathbf{X})$$

Welders and DNA

Welders get exposed to chromium and nickel, substances that can cause inappropriate links between DNA and proteins. Costa, Zhitkovich, and Toniolo measured DNA-protein cross-links in samples of white blood cells from 21 railroad arc welders exposed to chromium and nickel and from 26 unexposed controls. All 47 subjects were male. In their data ... there are three covariates, namely age, race and current smoking behavior. The response is a measure of DNA-protein cross-links.

Before matching, we get the following descriptive statistics for the means of the two groups:

```
          control      treat
age    42.6923077 38.2380952
black   0.1923077  0.0952381
smoker  0.3461538  0.5238095
```

Estimate the Propensity Score

How do we estimate the propensity score? Often we use a linear logit model (this bounds our propensity score between 0 and 1). The propensity is then estimated by:

$$log\left(\frac{e(\mathbf{X}_i)}{1-e(\mathbf{X}_i)}\right) = \zeta_0 + \zeta_1 age_i + \zeta_2 black_i + \zeta_3 smoker_i$$

$\hat{e}(\mathbf{X}_i)$ are the fitted values from this model

So, we run the following code in R to find our propensity score:

```
pscore <- glm(treat ~ age + black + smoker,
              family = binomial(link = logit),
              data = data)$fitted.values
```

Some Considerations

1. What distance metric do we use?

Some Considerations

1. What distance metric do we use?
2. Do we match with replacement or without replacement?

Some Considerations

1. What distance metric do we use?
2. Do we match with replacement or without replacement?
3. What do we do with ties?

Some Considerations

1. What distance metric do we use?
2. Do we match with replacement or without replacement?
3. What do we do with ties?
4. What do we consider a "good" match?

Distance Metrics

- In order to figure out what the "closest" match is, we have to decide what our metric for the distance between observations k and j.

Distance Metrics

- In order to figure out what the "closest" match is, we have to decide what our metric for the distance between observations k and j.
- Since we are only matching on one covariate, in this case the propensity score, we can use the squared distance between the two estimated propensity scores.

$$d = (\hat{e}(\mathbf{X}_k) - \hat{e}(\mathbf{X}_j))^2$$

Distance Metrics

- In order to figure out what the "closest" match is, we have to decide what our metric for the distance between observations k and j.
- Since we are only matching on one covariate, in this case the propensity score, we can use the squared distance between the two estimated propensity scores.

$$d = (\hat{e}(\mathbf{X}_k) - \hat{e}(\mathbf{X}_j))^2$$

- This will punish large differences more than small distances. Alternatively, we could use the absolute value of the distance between the estimated propensity scores.

Distance Metrics

- In order to figure out what the "closest" match is, we have to decide what our metric for the distance between observations k and j.
- Since we are only matching on one covariate, in this case the propensity score, we can use the squared distance between the two estimated propensity scores.

$$d = (\hat{e}(\mathbf{X}_k) - \hat{e}(\mathbf{X}_j))^2$$

- This will punish large differences more than small distances. Alternatively, we could use the absolute value of the distance between the estimated propensity scores.
- Nearest-neighbor matching matches the closest control unit to each treated unit (in the case of ATT) or the closest treated unit to each control unit (ATC), given the distance metric

With or Without Replacement

- If we match without replacement, then once we match a control unit, we take it out of the pool of potential matches for all remaining treated units.

With or Without Replacement

- If we match without replacement, then once we match a control unit, we take it out of the pool of potential matches for all remaining treated units.
- It is important to notice that if we do this, then depending on the order of the controls and the algorithm we use to sort through them, we may get different matches.

With or Without Replacement

- If we match without replacement, then once we match a control unit, we take it out of the pool of potential matches for all remaining treated units.
- It is important to notice that if we do this, then depending on the order of the controls and the algorithm we use to sort through them, we may get different matches.
- If we match with replacement, then this means that after a control gets matched to a treated unit, it goes back into the pool of potential matches for the remaining treated units. This means that a control unit could be matched to multiple treated units.

With or Without Replacement

- If we match without replacement, then once we match a control unit, we take it out of the pool of potential matches for all remaining treated units.
- It is important to notice that if we do this, then depending on the order of the controls and the algorithm we use to sort through them, we may get different matches.
- If we match with replacement, then this means that after a control gets matched to a treated unit, it goes back into the pool of potential matches for the remaining treated units. This means that a control unit could be matched to multiple treated units.
- In general, we'd like to match with replacement to make sure that we get the "best" match every time.

Ties

- The case may arise that when we look for matches to a given treated unit i, there are two control units that are the same distance from i based on our distance metric d.

Ties

- The case may arise that when we look for matches to a given treated unit i, there are two control units that are the same distance from i based on our distance metric d.
- What do we do?

Ties

- The case may arise that when we look for matches to a given treated unit i, there are two control units that are the same distance from i based on our distance metric d.
- What do we do?
- Flip a coin

Ties

- The case may arise that when we look for matches to a given treated unit i, there are two control units that are the same distance from i based on our distance metric d.
- What do we do?
- Flip a coin
- Allow a tie: we match both control units to treated unit i, but we give each of these controls a weight of $\frac{1}{2}$ in our matched data set (in effect, we average the control units)

What do we consider a good match?

- What if the closest control unit to treated unit i has a large distance, d. We may want to say that treated unit i cannot be matched because there is no control unit that is "close" to it.

What do we consider a good match?

- What if the closest control unit to treated unit i has a large distance, d. We may want to say that treated unit i cannot be matched because there is no control unit that is "close" to it.
- To do this, we would enforce a caliper, which says that if there is no "nearest neighbor" to treated unit k, defined as being within a certain distance of j, we say that we cannot match treated unit k.

$$|\hat{e}(\mathbf{X}_k) - \hat{e}(\mathbf{X}_j)| > w$$

Where, if the distance is greater than the caliper w, we set the distance to infinity.

What do we consider a good match?

- What if the closest control unit to treated unit i has a large distance, d. We may want to say that treated unit i cannot be matched because there is no control unit that is "close" to it.
- To do this, we would enforce a caliper, which says that if there is no "nearest neighbor" to treated unit k, defined as being within a certain distance of j, we say that we cannot match treated unit k.

$$|\hat{e}(\mathbf{X}_k) - \hat{e}(\mathbf{X}_j)| > w$$

Where, if the distance is greater than the caliper w, we set the distance to infinity.

- When we drop treated observations, we are changing what we are estimating, but it is no longer ATT ...