Joint Matching and Treatment Effect Estimation in Causal Inference through Bayesian Fellegi-Sunter Methods

Brian Kundinger

Duke University

May 11, 2023

Table of Contents

Overview of Approaches

- Exact Matching
- Mahalonobis Distance
- Propensity Score
- Almost Exact Matching

Mahalonobis Distance

- One method of matching has been to calculate distance in the covariate space through

$$d(X_1, X_2) = \sqrt{|X_1 - X_2| S^{-1} |X_1 - X_2|} \tag{1}$$

- An algorithm then matches each unit in control to closest unit in treated

Propensity Score

- Another method is to estimate propensity scores $p(Z_n = 1|X_n)$ for each unit in the data, and match units based on this score.

Pitfalls of Both Approaches

- The pros and cons of each method are summarized very well on this Stack Exchange page
- Essentially, propensity scores can end up matching units with that have similar probability of treatment, but have very different covariate values
- Conversely, Mahalonobis can end up matching units that are similar on features that are irrelevant to the outcome of interest

Additional Thoughts

- Sekhot (2011) recommends including the propensity score into the Malahalonobis distance. This is just practical advice, but it forms the foundation of thoughts to come
- This paper estimates PS, identifies all units in the treatment have PS within a specified caliper, and then chooses the unit with the smallest Mahalanobis distance. No uncertainty quantification
- Alvarez and Levin (2021) (Section 2, pages 5 - 7) use a Bayesian model to create draws of the propensity scores. For set of scores, they use a deterministic algorithm to match records. They do not use any other covariates

Table of Contents

Intuition

- The machinery of Fellegi Sunter record linkage is very good at identifying duplicate records. It should also be good at identifying separate individuals that are similar.
- FS does matching purely on the reliability and uniqueness (m and u parameters) of fields of information. It does not take into account any relationship with an outcome variable. Conceptually, this is very similar to the Mahalonobis distance, but allows for uncertainty quantification.
- If we include propensity scores into FS, the model will place high weight on units with "similar" scores. Then, the FS machinery for the other fields should find the units that are most similar.

Nuts and Bolts

- Data is $\{y_n, z_n, x_n\}_{n=1}^{N}$ where y_n is the outcome, z_n is the treatment indicator, and x_n is a vector of F fields of covariates.
- Split data into a datasets T with records $i \in \{1, \ldots, n_T\}$ for treated units and C with records $j \in \{1, \ldots, n_C\}$.
- At each step of the Gibbs Sampler, calculate propensity score $PS(n) = P(Z_n = 1|X)$ for each data point (logistic or probit regression)

Nuts and Bolts

- Define comparison field

$$\Gamma_0 = \begin{cases} 1 & \text{PS(i) and PS(j) are within specified caliper} \\ 0 & \text{otherwise} \end{cases}$$

- Conduct Fellegi Sunter Matching on new full set of comparison vectors $\Gamma^* = (\Gamma_0, \gamma)$. Note γ is fixed data, but the Γ are random.

Big Picture

- Each iteration of the Gibbs sampler will provide a matched dataset, from which you can calculate an MCMC sample of the treatment effect (TE) or other quantities of interest
- We obtain measures of match quality as usual, through the posterior samples of the linkage structure
- I believe this to be first causal matching method to produce matchings probabilistically
- The proposed method is fully model based, with no deterministic algorithms. It propogates uncertainty between the matching procedure and the treatment effect estimation.

Table of Contents

Model (Basic)

$$
\begin{aligned}
Z_n &= x_n'\beta + \epsilon \\
\beta &\sim N(0, \tau) \\
\epsilon &\sim N(0, \sigma^2) \\
\mathcal{L}(\Delta, m, u \mid \Gamma^*) &= \prod_{i=1}^{n_T} \prod_{j=1}^{n_C} \prod_{f=0}^{F} \prod_{l=1}^{L_f} \left[m_{fl}^{\Delta_{ij}=1} u_{fl}^{\Delta_{ij}=0} \right]^{I(\gamma_{ij}^f = l)} \\
m, u &\sim \text{Dirichlet} \\
\Delta &\sim \text{Desired Record Linkage Prior}
\end{aligned}
$$

Different record linkage priors can induce one-to-one, many-to-one, and many-to-many matches. Any one of them can work within this framework.

Model (With Variable Selection)

- In causal matching, it is especially important not to match on irrelevant features. Therefore, we can use a spike and slab prior on β, and only use fields where $\beta_f \neq 0$ in the linkage.

$$\beta_j \sim \pi_0 \delta_0 + (1 - \pi_0) N(0, \tau_j)$$

$$\mathcal{L}(\Delta, m, u \mid \Gamma^*, \beta) = \prod_{i=1}^{n_T} \prod_{j=1}^{n_C} \prod_{f=0}^{F} \left[\prod_{l=1}^{L_f} \left[m_{fl}^{\Delta_{ij}=1} u_{fl}^{1-\Delta_{ij}=1} \right]^{I(\gamma_{ij}^f = l)} \right]^{\beta_f \neq 0}$$

Full Conditional for β

Reexpress the likelihood as

$$\mathcal{L}(\Delta, m, u, \beta \mid \Gamma^*) = \prod_{i=1}^{n_T} \prod_{j=1}^{n_C} \prod_{f=0}^{F} c_f \prod_{l=1}^{L_f} \left(\frac{m_{fl}}{u_{fl}} \right)^{I(\gamma_{ij}^f = l) I(\Delta_{ij}=1) I(\beta_f \neq 0)}$$

$$= \prod_{i=1}^{n_T} \prod_{j=1}^{n_C} \prod_{f=0}^{F} c_f \prod_{l=1}^{L_f} w_{fl}^{I(\gamma_{ij}^f = l) I(\Delta_{ij}=1) I(\beta_f \neq 0)}$$

where $w_{fl} = \frac{m_{fl}}{u_{fl}}$. Let $M_{fl} = \sum_{i,j} I(\Delta_{ij} = 1) I(\gamma_{ij}^f = l)$.
Then, conditional on Δ, we have

$$\mathcal{L}(m, u, \beta \mid \Gamma^*, \Delta) = \prod_{i=1}^{n_T} \prod_{j=1}^{n_C} \prod_{f=0}^{F} \left[c_f \prod_{l=1}^{L_f} w_{fl}^{M_{fl}} \right]^{I(\beta_f \neq 0)}$$

Full Conditional for β

The quantity $\prod_{l=1}^{L_f} w_{fl}^{M_{fl}}$ is measure of how strong field f is in the linking procedure. And it shows up in the full conditional for β_f, showing how the regression and matching procedure are linked.

Table of Contents

Caveats

- Using a comparison vector made from the propensity score may technically violate the conditional independence assumption of FS
- Propensity and Mahalonobis matching both have theoretic garuntees. I don't know right now if this method would have them
- I know very little about causal inference! This would be a large undertaking

Questions

- What do you all think of this idea?
- Jerry, how well versed are you in matching for causal inference? To what extent to should I seek guidance/collaboration from Fan Li or Alex Volfovsky?

Possible Theoretical Question

- I'm really struck by the conceptual similarity between Mahalanobis distance and the FS weights. If you make some assumptions (categorical data, one-to-one matching, known parameters), it might be possible to prove a correspondence between the two.

Possibly Better Approach

- The FS model assumes that agreement levels are independent conditional on the matching status of the "record" pair. This is often violated, with minimal issues.
  However, the propensity score is very much not independent from other agreement levels. It may therefore to incorporate proximity in PS not as a feature, but rather as a *latent class* for the agreement vector.
  See Xu et. al (2019) for more on this.
  I think the first method would work well, and presents enough innovation to stand on its own. I can do this latent class stuff in an extension.