

Propensity Score Matching

by Valentina Andrade

23-12-2020

Definición

Es una técnica estadística para estudios observacionales, que realiza coincidencia o pareamiento intentando estimar el efecto de una intervención, tratamiento o política en base a covariables que predicen que este recibe un tratamiento.

Esta técnica intenta reducir el sesgo producido debido a la diferencia de resultados que pueden tener el grupo que recibió el tratamiento (g_t) del grupo que no lo recibió (g_c), sesgo que sería producido pues la selección o decisión de administrar el tratamiento se debe a las características propias de los grupos. En consecuencia, las diferencias de los grupos se deberían más a un sesgo por características (confusión de variables) que a una diferencia producida por el tratamiento en sí.

Este sesgo ocurre principalmente pues las mediciones son observacionales y no provienen de un *experimento aleatorio*, donde la aleatorización permite la estimación objetiva de los efectos de tratamiento. Empero, esta técnica permitiría la inferencia causal y corrección del sesgo de selección.

Matching

Consiste en coincidir las características únicas que distinguen a los grupos de control (g_c) y tratamiento (g_t), con el objetivo de hacer a estos dos lo más parecidos posible. Es frecuente que los grupos no compartan características sustantivas (superposición sustancial), entonces evidentemente se generará un error importante. Por ejemplo, puede ocurrir que los grupos de comparación seleccionados para el tratamiento tengan cierto grupo de características “mejores” o que los hacen más propensos y que el grupo de control tenga las “peores” o menos propensos, haciendo que la estimación del modelo muestre parámetros sesgados.

Scores

La diferencia sustancial entre el simple pareamiento (*matching*) con el *matching* por puntajes es que PSM cuenta con una probabilidad predicha a la pertenencia del grupo. Estas probabilidades predichas, que luego son los *puntajes* (*scores*), se obtienen de modelos (eg. logísticos) que permiten construir un grupo contrafactual.

La estrategia de ocupar las probabilidades predichas también se utiliza para hacer coincidir o parear ya sea como covariables independientes o ya sea, con las otras variables “coincidentes” (pero que no coinciden sustancialmente, produciendo el sesgo por superposición que comentábamos arriba).

Propensity scores

Se define como la probabilidad de una unidad (eg. persona, empresa, sindicato) que se le asigna a un tratamiento particular, dado un conjunto de covariables observadas. Su función es reducir el sesgo de selección al equiparar o “estandarizar” grupos basados en estas covariables.

1.Realizar regresión logística

Matemáticamente, sea T un tratamiento binario $[0,1]$, Y un resultado, y las covariables X_n , entonces

- $p(x) = Pr(T = 1|X = x)$, es la **probabilidad condicional

2. Predicción de la puntuación

Sea $Y(0)$ e $Y(1)$ los resultados potenciales bajo control y tratamiento, respectivamente. La asignación del tratamiento no estará condicionalmente confundida si los resultados potenciales son independientes del tratamiento condicionado por las variables previas (X). Es decir,

$T \perp Y(0), Y(1)|X$, donde \perp denota independencia estadística.

$T \perp Y(0), Y(1)|p(x)$, donde unconfoundedness mantiene

Procedimiento

1.Realizar una regresión logística:

- Variable dependiente: $y = 1$, si participar; $y = 0$, de lo contrario.
- Elegir las variables de “confusión” adecuadas (variables hipotéticamente asociadas con el tratamiento y el resultado).
- Obtener la puntuación de propensión: probabilidad predicha (p) o $\log(\frac{p}{(1-p)})$.

2.Emparejar cada participante a uno o más de los no participantes según el puntaje de propensión:

- Vecino más cercano similar
- Emparejamiento por calibre (caliper)
- Emparejamiento según el valor Mahalanobis junto con PSM
- Emparejamiento por estratificación
- Emparejamiento por diferencias en diferencias (kernel y los pesos lineales locales)

3.Verificar que las covariables están equilibradas entre el grupo de tratamiento y el de comparación, de la nueva muestra emparejada o ponderada

4.Realizar un análisis multivariado sobre la base de la nueva muestra

- Utilizar los análisis apropiados para diferentes muestras no independientes

Discusión

Ventajas

- Mediante el uso de una combinación lineal de las covariables para una única puntuación, que equilibra los grupos de tratamiento y control en un gran número de covariables sin perder un gran número de observaciones.

Desventajas

- **Identificación del modelo:** Ahora bien, si las unidades en el tratamiento y control se equilibraron en un gran número de covariables uno a la vez, serían necesarios un gran número de observaciones para superar el “problema de dimensionalidad” por lo que la introducción de una nueva covariable de equilibrio aumenta el número mínimo necesario de observaciones en la muestra geoméricamente.
- **Método observacional:** las covariables son observadas, y por ello los factores que probablemente si afectan a la asignación al tratamiento, pero que no pueden ser observados no pueden ser contabilizados en el procedimiento correspondiente.
- **Representatividad muestral:** por un lado, requiere grandes muestras y, por otro, con superposición sustancial entre los grupos de tratamiento y control.
- **Sesgo por uso de puntajes:** de la mano con los problemas ocurridos por sesgo de variable latente, hay quienes plantean (Pearl, 2000) que el igualar puntajes en las variables observadas produce un sesgo al confundir esos puntajes con las variables latentes.

Referencias

Rosenbaum, Paul R.; Rubin, Donald B. (1983). «The central role of the propensity score in observational studies for causal effects». *Biometrika* 70 (1): 41-55. doi:10.1093/biomet/70.1.41

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*, Cambridge University Press.

PSM

- Youtube.

Steps

Pan and Bai, 2015

1. Estimate propensity score

Logistics or probit regresion of the treatment condition on the vector of covariates

2. Matching

Types of matching

- a) Similar
 - **Nearest neighbor matching:** match treatment unit i to nearest comparison unit j (i.e, match units with the most similar propensity scores)
- b) Distance
 - **Caliper matching:** match treatment unit i to a comparison unit j within a range (caliper)
 - **Radius matching:** like caliper but matches a treatment unit i , to multiple comparison units within a band
 - **Others:** Mahalanobis

Post-hoc

- **Greedy and optimal matching:** as soon as a match is made between two observations kept even if a better match could be found later in the data whereas an optimal matching the algorithm is able to reconsider the match and perhaps match two units that have closer propensity scores
- **Replacement:** With replacement, means an observation can actually be matched more than one time, it is put back in the pool of possible matches whereas; without replacement, means that once a match has been made that the observation is taken out and not able to be matched with other units and the data

3. Evaluate quality of matching

The goal of the matching techniques is to achieve balance between the treatment and comparison group on observable traits (i.e, the groups look the same on observables beforehand). We can evaluate the quality of the match with several different approaches

Approches

- a. Compare means (*t-test*)
- b. Calculate standardized bias for each covariate (for balance)
- c. Percent bias reduction (*PBR*) (for balance)
- d. Graphical comparison (histograms, box-plots, etc.)

4. Evaluate outcomes

Evaluate the effect of an intervention or policy. We can do this again with several ways

1. **Compare means of matched samples:** if our algorithm do a good job, our analysis are more closely approximates a randomized control trial
2. **Run a regression on the matched sample controlling for unbalanced covariates**
3. **Propensity score adjustment:** running a regression controlling for propensity score

Limitations of PSM

1. **Omitted variable** bias may still be an issue. Observational method still be a problem (eg. Two people say they went to vote in the plebiscite but, in some way, it is likely that this answer is hiding other variables such as political position or others associated with the desirability of the vote.)
2. **Independence conditional on the covariates:** it assumes independence conditional on the covariates. In some sense, this is little different than assumption of *OLS*.

Q&A

- Diferencia con diff in diff

PSM - CMOYA

- Propensity Score Matching (*PSM*) is a method for preprocessing data to improve causal inferences in observational data (Pearl, 2012; Morgan and Winship, 2014). The goal is to reduce imbalance in the empirical distribution of the pre-treatment confounders between the treated and control groups (Suart, 2010, p.13).

In a general, the method balance the data by pruning some observations through matching the more likely PSM will degrade inference (using predict probabilities of the logistic regression in step 1)

Goals

Critics

- increasing imbalance
- model dependence: Lowering imbalance reduces the degree of model dependence in the statistical estimation of causal effects (Ho et al., 2007; Imai, King and Stuart, 2008; Iacus, King and Porro, 2011b)
- Leads inefficiency biased estimates and research discretion. The second and third, will be produce by PSM since it leaves researchers making qualitative choices among larger sets of causal estimates.
- Hypothesis: If one's data are so imbalanced that making valid causal inferences from it without heavy modeling assumptions is impossible, then the paradox we identify is avoidable and PSM will reduce imbalance but then the data are probably not very useful for causal inference by any method. Thus, PSM will be productive to other uses (see Pro), because mathematical theorems construct by Rosenbaum and Rubin (1984) are correct and useful, but not **suitable** and relevant to finite sample matching analyses.
- Evidence: Interaction between propensity scores and matching.

Pro

Thus, PSM will be productive to other uses, such regression adjustment (Vansteelandt and Daniel, 2014), inverse weighting (Robins, Hernan and Brumback, 2000), stratification (Rosenbaum and Rubin, 1984), and some uses of the propensity score within other methods (e.g. Diamond and Sekhon, 2012; Imai and Ratkovic, 2014).

Revisar

- Confounders
- Fully blocked randomized experiments vs completely randomized experiments (Rubin and Thomas, 2000)