

# Predicción y Causalidad

Alvaro J. Riascos Villegas  
Universidad de los Andes y Quantil

Septiembre de 2017

# Contenido

1 Introducción

2 Ejemplos

# Introducción

- Kleinberg, Ludwig, Mullainathan, Obermeyer. 2015. Prediction Policy Problems. AER.
- En ocasiones la identificación de un efecto casual es irrelevante.
- Este artículo introduce un marco conceptual para pensar la relación entre el problema de predicción y causalidad.

# Introducción

- Sea  $y$  la variable de interés que no sabemos cómo depende de  $x_0$ , que es exógena (política económica) y  $x$ .
- El objetivo es maximizar una función conocida  $\Pi(x_0, y)$ .
- La decisión depende de:

$$\frac{d\Pi}{dx_0} = \frac{\partial\Pi}{\partial x_0}(x_0, y) + \frac{\partial\Pi}{\partial y}(x_0, y) \frac{\partial y}{\partial x_0}$$

- 1 Si bien  $\Pi$  es conocido el efecto de  $x_0$  depende de  $y$  (un problema de predicción).
  - 2 El segundo término depende de como  $x_0$  afecta a  $y$  (un problema de causalidad).
- Obérvese que ambos efectos dependen de la predicción  $y$ .
  - Luego un tomador de decisiones debe resolver ambos problemas.

# Introducción: Ejemplo

- Suponga que la variable de interés  $y$  es si llueve o no. En un problema la acción es hacer un rito para que no llueva. En otro problema es decidir si llevar o no una sombrilla.
- La función objetivo puede ser la utilidad un domingo que se va al parque.
- El problema de hacer un rito para que no llueva es un problema de causalidad:

$$\frac{\partial \Pi}{\partial x_0}(x_0, y) = 0$$

- El problema de llevar una sombrilla es un problema puro de predicción:

$$\frac{\partial y}{\partial x_0} = 0$$

# Introducción: Ejemplo

- El problema de predicción es relevante porque y afecta directamente la utilidad (en el ejemplo anterior, la utilidad de una sombrilla depende de si llueve o no).
- Obsérvese que la dependencia es del pronóstico  $\hat{y}$ , luego lo que importa es predecir bien, no que los coeficientes sean no sesgados o que exista un efecto casual.

# Contenido

1 Introducción

2 Ejemplos

# Ejemplos

- La utilidad de un reemplazo de cadera depende de el reemplazo ( $x_0$ ) y cuanto tiempo vaya vivir una persona con el reemplazo ( $y$ ).
- Presumiblemente, el reemplazo no afecta el tiempo que va vivir una persona. Este es un problema puro de predicción.
- El 1,3 % de los beneficiarios de Medicare en el 2010 se hicieron reemplazo de cadera.
- De esto el 1,4 % se mueren el primer mes (posiblemente por complicaciones). 4,2 % se muere entre el mes 1 y 12. Esto es más o menos la mortalidad en un año de Medicare.
- Debemos concluir que el reemplazo es útil?



- La pregunta correcta es, si entre aquellos que es predecible que van a morir con una probabilidad alta (alto riesgo de muerte), se podría hacer algo para obtener mayores beneficios netos para la sociedad.

TABLE 1—RISKIEST JOINT REPLACEMENTS

Predicted mortality percentile	Observed mortality rate	Futile procedures averted	Futile spending (\$ mill.)
1	0.435 (0.028)	1,984	30
2	0.422 (0.028)	3,844	58
5	0.358 (0.027)	8,061	121
10	0.242 (0.024)	10,512	158
20	0.152 (0.020)	12,317	185
30	0.136 (0.019)	16,151	242

*Notes:* We predict 1–12 month mortality using an  $L_1$  regularized logistic regression trained on 65,395 Medicare beneficiaries undergoing joint replacement in 2010, using 3,305 claims-based variables and 51 state indicators.  $\lambda$  was tuned using ten-fold cross-validation in the training set. In columns 1 and 2 we sort a hold-out set of 32,695 by predicted risk into percentiles (column 1) and calculate actual 1–12 month mortality (column 2). Columns 3 and 4 show results of a simulation exercise: we identify a population of eligibles (using published Medicare guidelines: those who had multiple visits to physicians for osteoarthritis and multiple claims for physical therapy or therapeutic joint injections) who did not receive replacement and assign them a predicted risk. We then substitute the high risk surgeries in each row with patients from this eligible distribution for replacement, starting at *median* predicted risk. Column 3 counts the futile procedures averted (i.e., replaced with non-futile procedures) and column 4 quantifies the dollars saved in millions by this substitution.