

ENTREGA FINAL

Anzony Quispe

Introducción

Muchos efectos causales o efectos de políticas públicas vienen definidos por modelos lineales de alta dimensión ($p > n$) ya sea porque la fuente de información tiene esa cantidad de variables o porque se utiliza regresores contruados, o ambos. Esta cantidad de parámetros (covariables) generar problemas en la estimación del parámetro. Luego, las alternativas para lograr su estimación, herramientas de *machine learning* como Lasso, introducen sesgo y dejamos de alcanzar una estimación generan consistente y asintóticamente normal.

Frente a ello, surge la necesidad explotar bases de datos de alta dimensionalidad y seguir obteniendo estimaciones asintóticamente normales y consistentes. Chernozhukov et al. (2018) logran eliminar el sesgo generado por la regularización y overfitting propios de los modelos de *Machine Learning*. Este proceso de *debiasing* (eliminación del sesgo) varía de acuerdo a la característica de la variable de tratamiento (binaria, continua, efectos de política incrementales, etc). Estas distintas formas de la variable de interés hacen más complicado el proceso de *debiasing*; sin embargo, Chernozhukov et al. (2022) explotan el teorema de representación de Riesz para no tener que enfrentar de manera distinta el proceso *debiasing*, sino aprovechar nuevamente las bondades de *machine Learning* y mantener el proceso de *debiasing*. Esta última forma de estimación no es de forma cerrada, sino que es una aproximación.

La metodología ya ha sido implementada usando Redes Neuronales y *Random Forests*. Sin embargo, aún no se ha generado un paquete de libre uso para su implementación. Esta tesis busca generar como producto un paquete en Python, R y Julia, poniendo

más énfasis al primer lenguaje, que implemente esta metodología y además buscará hacer un estudio de simulación que muestre la validación de la metodología y compare con otros métodos de estimación como DoubleMLIRM y rlassoATE.

Modelo Descriptivo

El modelo sigue al paper de Automatic Debiased Machine Learning (Chernozhukov et al., 2022). Esta es una extensión del trabajo primario de Chernozhukov et al. (2014). Intentaré explicar cómo el método consigue estimar de manera consistente un parámetro de dimensión finita ante la presencia de parámetros de alta dimensión siguiendo un Modelo Parcial Lineal tal como lo hace el autor originalmente. Veremos que la estimación logra eliminar el sesgo por regularización y sobreajuste que son propios de los modelos de *machine Learning*. Luego, señalaremos que este proceso de *debiasing* puede ser extendido para cualquier forma de variable de interés explotando el teorema de Riesz.

Los métodos de estimación de *Machine Learning* son particularmente precisos al predecir en un contexto de alta dimensión, es decir, cuando los parámetros son muchos, incluso más que las observaciones en la base de datos. Esta propiedad viene del hecho que utilizan procesos de regularización que hace posible que el aprendizaje sea posible en los modelos. La penalización de algunas variables mejora la predicción porque reduce varianza y añade sesgo en menor proporción, en consecuencia, una estimación más precisa. Además, estos métodos por lo general tienen cierto grado de overfitting. Pese a que algunos creadores de estos algoritmos señalan que no es así (e.g. *Random Forest*), no han demostrado la ausencia de overfitting.

El overfitting y la regularización a una estimación sesgada del parámetro a estimar. Sin embargo, los autores encuentran que este sesgo puede ser eliminado si utilizamos momentos Neyman-Orthogonal y cross-fitting. El primero estima el modelo ortogonalizando el tratamiento con el resto de las variables y quedándose únicamente

con el residuo de la regresión. Un proceso muy parecido al que seguimos en Frisch–Waugh–Lovell. Para este proceso utilizamos un modelo ML y lo ajustamos con una porción auxiliar de la data. De esta parte proviene el nombre de *double machine learning*, puesto que estamos estimando dos modelos, uno para la regresión principal y otro en la ortogonalización. Este proceso extra permite eliminar el sesgo generado por la regularización del modelo original, en otras palabras, logra que el parámetro sea menos sensible al resto de parámetros distractores.

Sin embargo, aún después de eliminar este sesgo queda por resolver el sesgo inducido por *overfitting*. Para esto necesitamos asumir que las observaciones son independientes y que el modelo ML principal y auxiliar fue ajustado utilizando únicamente las observaciones auxiliares. Por lo tanto, cuando condicionamos en la muestra auxiliar, este último término tiende a 0 en probabilidad. Con ambos modelos ajustados, usamos la data principal y obtenemos una estimación del parámetro, luego repetimos el proceso solo intercambiando la data y obtenemos otra estimación del parámetro, finalmente, obtenemos el promedio de ambos parámetros que sí es consistente.

Este proceso señalado arriba puede ser resuelto utilizando la ortogonalización de Neyman y el teorema de representación de Riesz. Ambos permiten visualizar el término que quita el sesgo de parametrización (*debiasing term*). La literatura ha intentado caracterizar este término para luego estimarlo; sin embargo, este se puede estimar directamente. El teorema de representación de Riesz permite estimar este término con cualquier modelo de *Machine Learning* ya que expresa al término *debiasing* como el resultado de un proceso de minimización del error entre el momento del tratamiento y el termino *debiasing*, que a su vez es el error entre el sesgo de regularización verdadero y el estimado.

Finalmente, usando *Automatic Debiased Machine Learning* no necesitamos caracterizar el *debiasing* term sino que nos aproximamos a él utilizando un modelo de *machine Learning*. Eliminamos el sesgo siguiendo la metodología señalada arriba.

Modelo Matemático

En esta sección comentaré acerca del modelo matemático que seguiré en la tesis. Este modelo está basado en el trabajo acerca de Automatic Debiased Machine Learning (Chernozhukov et al., 2022). Este trabajo tiene por objetivo la estimación del parámetro θ_0 , en un contexto de alta dimensionalidad.

$$\theta_0 = E(m(W; g_0))$$

En esta ecuación tenemos que $W := (Y, Z)$ y $g_0(Z) := E[Y|Z]$. Tenemos que Y es el outcome de interés y $Z = (T, X)$ contiene a las covariables y el tratamiento de interés. Como lo expresamos en el modelo descriptivo, no podemos obtener la estimación directamente de θ_0 a través de un modelo ML debido al sesgo generado por regularización y overfitting. Para ello, necesitamos de procesos de *debiasing* como el explicado por Chernozhukov et al. (2018). Sin embargo, en trabajos más recientes (Chernozhukov et al. (2021)) esta primera ecuación puede ser expresada como la multiplicación de dos términos bajo el teorema de representación de Riesz.

$$\theta_0 = E(m(W; g_0)) = E(\alpha_0(Z)g_0(Z))$$

Donde $\alpha_0(Z)$ existe si y solo si $E(m(W; g_0))$ es una función lineal continua de g_0 . Se denomina a $\alpha_0(Z)$ como *Riesz representer* (RR). Luego, este término es útil para generar términos que quiten el sesgo de la función de momento $m(W; g_0)$. Para el caso mostrado arriba, se muestra que podemos estimar el parámetro de la siguiente ecuación.

$$\hat{\theta} = E_n[m(W; \hat{g}) + \hat{\alpha}(Z)(Y - \hat{g}(Z))]$$

Entonces, cuando el producto de los ratios de convergencia de las *nuisance functions* es mayor que $n^{0.5}$, el estimador converge en distribución. Finalmente, si g_0 está correctamente estimado, el requerimiento del ratio de $\hat{\alpha}$ es menos estricto.

Estos términos permiten quitar el sesgo de la estimación de $\hat{\theta}$. La literatura ha ido construyendo estos términos de forma independiente para diferentes casos. Sin embargo, Chernozhukov et al. (2021) muestra que este término puede ser estimado como una función de minimización que evita la derivación de una forma específica del término, especificación del término.

$$\begin{aligned}\alpha_0 &= \arg \min_{\alpha} E[(\alpha(Z) - \alpha_0(Z))^2] \\ \alpha_0 &= \arg \min_{\alpha} E[\alpha(Z)^2 - 2m(W; \alpha)] \\ \hat{\alpha} &= \arg \min_{\alpha \in A_n} E[\alpha(Z)^2 - 2m(W; \alpha)]\end{aligned}$$

Entonces, siguiendo el teorema representación de Riesz Podemos obtener el término *debiating* sin derivar su forma análitica, y por el proceso de minimización podemos tener un intercambio entre sesgo y varianza.

En este trabajo de investigación se buscará implementar esta metodología que aún no está disponible para Python ni R ni Julia. Luego, se aplicará en un estudio de simulación para su validación y se comparará con otras metodologías (rlassoATE, DoubleMLIRM).

Plan

El plan seguirá los siguientes pasos:

1. Revisión de literature y presentación a los mentores acerca de la metodología a seguir. Este paso busca profundizar en el entendimiento de la metodología y sus posibles aplicaciones.
2. Replicar los códigos hechos por Victor Quintanas en el [siguiente link](#). Esto con el fin de familiarizarnos con los métodos computacionales previamente seguidos.
3. Generar los códigos, clases para [DoubleML](#) y añadirlo a su set de herramientas que ya tienen para estimación en grandes dimensiones.
4. Comparar los resultados de la metodología con otros procesos de estimación como rlassoATE y DoubleMLIRM.

Mentores

Los mentores ya están encontrados y son la profesora Verónica Diman y Walter Sosa. Con la primera planeo discutir y profundizar las bases teóricas del teorema de representación de Riesz y con Walter quiero discutir los resultados de las estimaciones. El proceso de generación del paquete está a mi cargo y planeo discutirlo con Martin Spindler y Philip Bach creadores del repositorio DoubleML.

Referencias

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.

Chernozhukov, V., Newey, W. K., Quintas-Martinez, V., & Syrgkanis, V. (2021). Automatic debiased machine learning via neural nets for generalized linear regression. arXiv preprint arXiv:2104.14737.

Chernozhukov, V., Newey, W. K., & Singh, R. (2022). Automatic debiased machine learning of causal and structural effects. *Econometrica*, 90(3), 967-1027.

Chernozhukov, V., Newey, W., Quintas-Martínez, V. M., & Syrgkanis, V. (2022, June). Riesznet and forestriesz: Automatic debiased machine learning with neural nets and random forests. In *International Conference on Machine Learning* (pp. 3901-3914). PMLR.