Introdução à Econometria Semiparamétrica

Aula 3 - Estimação Semiparamétrica

Luis A. F. Alvarez

10 de outubro de 2024

conteúdo...

Bibliografia I

Angrist, Joshua D. (1998). "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants". Em: *Econometrica* 66.2, pp. 249–288. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/2998558 (acesso em 10/10/2024).

Goldsmith-Pinkham, Paul, Peter Hull e Michal Kolesár (2024). *Contamination Bias in Linear Regressions*. arXiv: 2106.05024 [econ.EM]. URL: https://arxiv.org/abs/2106.05024.

Exemplo

- Considere uma população de interesse em que definimos um tratamento individual binário, denotado por uma variável aleatória, $D \in \{0, 1\}$ e um resultado de interesse Y.
  - Os resultados potenciais, que descrevem o que acontece com um indivíduo caso ele seja alocado ao tratamento ou não, são dados por $(Y(0), Y(1))$, de modo que o resultado observado é dado por $Y = DY(0) + (1 - D)Y(1)$ e o efeito da política é $Y(1) - Y(0)$.
- Sejam X um vetor de características observáveis, tais que seja razoável supor que:

$$Y(0) \perp D|X$$

Estimação do ATT

- Sob a hipótese de identificação anterior, se supomos que $\mathbb{P}[D = 1] > 0$ e a seguinte hipótese de suporte comum (*overlap*):

$$\exists \epsilon > 0, \quad \mathbb{P}[\mathbb{P}[D = 0|X] \geq \epsilon] = 1$$

- Então é possível identificar o efeito médio do tratamento nos tratados,

$$\mathbb{E}[Y(1) - Y(0)|D = 1] = \frac{\mathbb{E}[DY]}{\mathbb{E}[D]} - \mathbb{E}\left[\frac{1}{1 - \mathbb{P}[D = 1|X = 1]}(1 - D)Y\right]$$

- O problema é que a hipótese de suporte comum pode ser irrazoável em alguns contextos.

Combinações convexas do ATT

- Considere, como alternativa, o estimando β^* que resolve

$$(\beta^*, g) = \text{argmin}_{b \in \mathbb{R}, h \in \mathcal{H}} \mathbb{E}[(Y - bD - h(X))^2],$$

onde $\mathcal{H}$ é um sub-espaço fechado de $L_2(\mathbb{P}_X)$.

- Nesse caso, é possível mostrar que, se $\mathbb{P}[D = 1|X] \in \mathcal{H}$ ou $\mathbb{E}[Y(0)|X] \in \mathcal{H}$, então:

$$\beta^* = \frac{\mathbb{E}[\mathbb{E}[Y(1) - Y(0)|X](D - v(X))^2]}{\mathbb{E}[(D - v(X))^2]}$$

onde $v(X) = \text{argmin}_{h \in \mathcal{H}} \mathbb{E}[(D - v(X))^2]$.

- Resultado é extensão direta de Angrist (1998).
- Estimando é uma combinação convexa de ATTs, dando mais peso para pontos do suporte com melhor sobreposição.
    - Combinação convexa mais fácil de se estimar eficientemente (Goldsmith-Pinkham, Hull e Kolesár, 2024), sob algumas hipóteses.

Estimando β^*

- Com base no resultado anterior e uma amostra aleatória da população, poderíamos tentar estimar β^* resolvendo o análogo amostral do problema.
  - Para implementar a classes "complexas" $\mathcal{H}$, podemos alternar o estimador de MQO dos resíduos $(Y_i - \tilde{g}(X_i))$ em D_i e o estimador que projeta $Y_i - \tilde{\beta} D_i$ em $\mathcal{H}$ até convergência.
- Note, entretanto, que para a representação anterior valer, devemos escolher uma classe suficiente expressiva para representar ou $\mathbb{E}[Y(0)|X]$ ou $\mathbb{P}[D = 1|X]$.
  - Além disso, se temos muitos possíveis controles, mas somente um subconjunto deve ser relevante para explicar a seleção ao tratamento, deveríamos utilizar métodos de alta dimensão válidos sob esparsidade aproximada.
- Estimador resultante é dado por:

$$\hat{\beta} = \frac{1}{n} \frac{\sum_{i=1}^{n} D_i(Y_i - \hat{g}(X_i))}{\mathbb{E}_n D^2},$$

onde $\mathbb{E}_n X$ é abreviatura para $\frac{1}{n}\sum_{i=1}^{n} X_i$.

Aproximação assintótica do estimador

$$\sqrt{n}\left(\hat{\beta}_0 - \beta^*\right) =$$

$$\underbrace{\left(\frac{1}{n}\sum_{i\in I} D_i^2\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i\in I} D_i U_i}_{=:a} + \underbrace{\left(\frac{1}{n}\sum_{i\in I} D_i^2\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i\in I} D_i\left(g\left(X_i\right) - \hat{g}\left(X_i\right)\right)}_{=:b}$$

onde $U_i = Y_i - \beta^* D_i - g(X_i)$

- Termo a é bem comportado em amostras grandes $a \xrightarrow{d} N(0, \sigma^2)$.
- No entanto, termo b é dado por: $b = \left(E\left[D_i^2\right]\right)^{-1} \frac{1}{\sqrt{n}} \sum_{i\in I} m_0\left(X_i\right)\left(g_0\left(X_i\right) - \hat{g}_0\left(X_i\right)\right) + o_P(1)$ esse termo não é bem comportado para estimadores não paramétricos modernos.