Introdução à Econometria Semiparamétrica

Aula 2 - Estimação Não Paramétrica Moderna

Luis A. F. Alvarez

7 de outubro de 2024

Recapitulando a estimação por séries

- Recorde-se do análogo populacional

$$\hat{h} \in \min_{s \in \mathcal{H}} \mathbb{E}[(Y_i - s(\boldsymbol{X}_i))^2]\,,$$

  onde $\mathcal{H}$ é um sub-espaço de $L_2(\mathbb{P}_{\boldsymbol{X}})$ "simples" (dimensão finita).
- Na análise de séries, $\mathcal{H} = \Theta_{J_n}$, onde $(\Theta_j)_{j \in \mathbb{N}}$ é uma sequência crescente de espaços com propriedades de aproximação global.
  - A escolha de J_n na prática visava a operar o *trade-off* viés-variância de modo a produzir um estimador com boas propriedades.
    - Por exemplo, para o *spline* cúbico, a escolha ótima em termos de velocidade de convergência do estimador é: $J_n \propto (\log(n)/n)^{s/(s+d)}$

Estimação não paramétrica moderna

- Os métodos da literatura que se convencionou chamar aprendizagem estatística (ou aprendizagem de máquina, em seu braço mais computacional) também partem do problema populacional.

$$\hat{h} \in \min_{s \in \mathcal{H}} \mathbb{E}[(Y_i - s(\boldsymbol{X}_i))^2] \, ,$$

- O que esses problemas adicionam, em relação à estimação clássica por séries?
  - Classes $\mathcal{H}$ de funções que incorporam não linearidade ("expressividade") de um jeito "inteligente", com "menor" complexidade (estimadores de $\downarrow$ variância) que métodos de séries.
  - De modo relacionado, métodos de seleção da complexidade da aproximação utilizada que, implícita ou explicitamente, operam no trade-off viés-variância de modo eficiente.

Estimação não paramétrica em altas dimensões

- Alguns dos métodos de aprendizagem estatística também são bastante úteis em ambiente de alta dimensionalidade.
  - Conceitualmente, ambientes de alta dimensionalidade são aqueles em que a aproximação assintótica mais adequada para representar o processo gerador é uma em que a dimensão de $\boldsymbol{X}$, d_n, diverge com $n \to \infty$, com possivelmente $d_n >> n$.

- Veremos que são as restrições, implícitas ou explícitas, na expressividade das classes $\mathcal{H}$ usadas por métodos de aprendizagem estatística, que garantem seu bom comportamento em ambientes de alta dimensionalidade.
  - O bom funcionamento prático desses métodos decorre, pois, de essas restrições servirem de boa aproximação para o processo gerador verdadeiro.

- Essas restrições levam a um bom funcionamento dos métodos mesmo quando $d_n > n$, evitando o problema do sobreajuste de métodos clássicos.

Problema do sobreajuste

- Considere um contexto em que temos n observações independentes do par $(Y, \boldsymbol{X})$, onde: Y é uma resposta escalar de interesse e $\boldsymbol{X}$ é um vetor de $k < n$ controles
  - Suponha que a matriz de desenho $\mathbb{X}_{n\times k} = [\boldsymbol{X}_1, \ldots \boldsymbol{X}_n]'$ apresenta posto k, e que $\mathbb{E}[Y_i|\boldsymbol{X}_i] = \gamma'\boldsymbol{X}_i$.
- Considere gerar $n-k$ vetores de controles adicionais Z_j, $j = k+1, \ldots n$, sorteando-os independentemente dos dados e entre si, de uma, $\mathcal{N}(0,1)$.
- Seja $\hat{\beta}$ o estimador de MQO de Y em X; e $(\tilde{\beta}, \tilde{\gamma})$ o estimador de MQO de Y em X e os Z_j.
  - Qual estimador tem o melhor ajuste na amostra?
- Considere realizar uma previsão de Y, com base nos estimadores da amostra, e num novo ponto $\boldsymbol{X}^*$, independente das demais observações?
  - Qual estimador esperamos que funcionará melhor, em termos de erro quadrático médio?

Problema de sobreajuste

- O exemplo anterior mostra, num cenário extremo, que estimadores baseados na minimização do risco empírico podem apresentar comportamento bastante indesejável quando a dimensão dos controles é alta.
  - Quando o número de controles k é moderadamente grande em comparação a n, estimador pode exibir um excelente ajuste dentro da amostra, mas funcionar bastante mal em termos de aproximar $\mathbb{E}[Y|\boldsymbol{X}]$, o objeto de interesse.
    - Estimador se ajusta inclusive ao erro idiossincrático ϵ_i dos $Y_i = \mathbb{E}[Y|\boldsymbol{X}_i] + \epsilon_i$ usados na estimação, produzindo alta variância.
- Esse problema é especialmente acentuaoa na estimação por séries, em que a dimensão do vetor utilizado na estimação, (número de elementos da base de Θ_{J_n}) , cresce exponencialmente no número de entradas de $\boldsymbol{X}$.

Regularização e sobreajuste

- Uma solução, na estimação por séries, é considerar o seguinte problema regularizado.

$$\min_{s \in \Theta_{\bar{J}}} \sum_{i=1}^{n} (y_i - s(\boldsymbol{X}_i))^2 + \lambda \Phi(s)\,,$$

  onde $\bar{J}$ pode ser relativamente "grande", e Φ é uma função que denota a "complexidade"de um candidato s.
  - $\lambda > 0$ dá o peso relativo da penalização, vis-à-vis ajuste na amostra (*trade-off* viés-variância).
- **Exemplo:** para *splines* cúbicos, pode-se tomar $\Phi(s) = \int (s''(x))^2 dx$.
  - Nesse caso, se

Bibliografia I

Achdou, Yves et al. (2022). "Income and wealth distribution in macroeconomics: A continuous-time approach". Em: *The review of economic studies* 89.1, pp. 45–86.

Alvarez, Luis e Cristine Pinto (2023). *A maximal inequality for local empirical processes under weak dependence*. arXiv: 2307.01328 [econ.EM]. URL: https://arxiv.org/abs/2307.01328.

Armstrong, Timothy B e Michal Kolesár (2020). "Simple and honest confidence intervals in nonparametric regression". Em: *Quantitative Economics* 11.1, pp. 1–39.

Belloni, Alexandre, Victor Chernozhukov, Denis Chetverikov e Iván Fernández-Val (2019). "Conditional quantile processes based on series or many regressors". Em: *Journal of Econometrics* 213.1, pp. 4–29.

Bibliografia II

- Belloni, Alexandre, Victor Chernozhukov, Denis Chetverikov e Kengo Kato (2015). "Some new asymptotic theory for least squares series: Pointwise and uniform results". Em: *Journal of Econometrics* 186.2, pp. 345–366.
- Benhabib, Jess e Alberto Bisin (2018). "Skewed wealth distributions: Theory and empirics". Em: *Journal of Economic Literature* 56.4, pp. 1261–1291.
- Calonico, Sebastian, Matias D Cattaneo e Max H Farrell (2018). "On the effect of bias estimation on coverage accuracy in nonparametric inference". Em: *Journal of the American Statistical Association* 113.522, pp. 767–779.
- — (2022). "Coverage error optimal confidence intervals for local polynomial regression". Em: *Bernoulli* 28.4.
- Calonico, Sebastian, Matias D Cattaneo e Rocio Titiunik (2014). "Robust nonparametric confidence intervals for regression-discontinuity designs". Em: *Econometrica* 82.6, pp. 2295–2326.

Bibliografia III

- Cattaneo, Matias D, Max H Farrell e Yingjie Feng (2020). "Large sample properties of partitioning-based series estimators". Em: *The Annals of Statistics* 48.3, pp. 1718–1741.
- Chen, Xiaohong (2007). "Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models". Em: ed. por James J. Heckman e Edward E. Leamer. Vol. 6. Handbook of Econometrics. Elsevier, pp. 5549–5632. DOI: https://doi.org/10.1016/S1573-4412(07)06076-X. URL: https://www.sciencedirect.com/science/article/pii/S157344120706076X.

Bibliografia IV

Chen, Xiaohong, Timothy Christensen e Sid Kankanala (mar. de 2024). "Adaptive Estimation and Uniform Confidence Bands for Nonparametric Structural Functions and Elasticities". Em: *The Review of Economic Studies*, rdae025. ISSN: 0034-6527. DOI: 10.1093/restud/rdae025. eprint: https://academic.oup.com/restud/advance-article-pdf/doi/10.1093/restud/rdae025/57021176/rdae025.pdf. URL: https://doi.org/10.1093/restud/rdae025.

Chen, Xiaohong e Timothy M. Christensen (2015). "Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions". Em: *Journal of Econometrics* 188.2. Heterogeneity in Panel Data and in Nonparametric Analysis in honor of Professor Cheng Hsiao, pp. 447–465. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2015.03.010. URL: https://www.sciencedirect.com/science/article/pii/S0304407615000792.

Bibliografia V

Einmahl, Uwe e David M. Mason (2005). “Uniform in bandwidth consistency of kernel-type function estimators”. Em: *The Annals of Statistics* 33.3, pp. 1380–1403. DOI: 10.1214/009053605000000129. URL: https://doi.org/10.1214/009053605000000129.

Fan, Jianqing e Irene Gijbels (mar. de 1996). *Local polynomial modelling and its applications*. en. Chapman & Hall/CRC Monographs on Statistics and Applied Probability. Philadelphia, PA: Chapman & Hall/CRC.