
Estimação Não-Paramétrica da Função de Distribuição: EDF vs AKDF e AKDF vs AKDF

Carlos TENREIRO – `tenreiro@mat.uc.pt`
Departamento de Matemática, FCTUC

Resumo: Se X_1, \dots, X_n são variáveis aleatórias reais independentes e absolutamente contínuas com densidade comum f , o estimador habitualmente considerado da sua função de distribuição F é a função de distribuição empírica (EDF). Motivados pelos trabalhos de Rosenblatt (Ann. Math. Statist. 27 (1956) 832) e de Parzen (Ann. Math. Statist. 33 (1962) 1065) sobre o estimador do núcleo da densidade f , autores como Tiago de Oliveira (Rev. Fac. Ciências Lisboa 9 (1963) 111), Nadaraya (Theory Probab. Appl. 9 (1964) 497) e Watson e Leadbetter (Sankhyä, Ser. A 26 (1964) 101), aparentemente de forma independente, propõem uma nova classe de estimadores de F . Para $x \in \mathbb{R}$, o estimador do núcleo de F (KDF) é definido por

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \bar{K}_{h_n}(x - X_i),$$

onde, para $u \in \mathbb{R}$,

$$\bar{K}_{h_n}(u) = \int_{]-\infty, u]} K_{h_n}(v) dv,$$

com $K_{h_n}(\cdot) = K(\cdot/h_n)/h_n$, K um núcleo em \mathbb{R} , i.e., uma função integrável tal que $\int_{\mathbb{R}} K(u) du = 1$, e (h_n) uma sucessão de números reais estritamente positivos convergindo para zero quando $n \rightarrow +\infty$. Para um núcleo fixo, a janela h_n é habitualmente escolhida em função das observações X_1, \dots, X_n . Tal procedimento conduz a uma classe mais vasta de estimadores de F , com $h_n = A_n(X_1, \dots, X_n)$ uma sucessão de funções mensuráveis, ditos estimadores automáticos do núcleo (AKDF).

Tendo por base o trabalho de Shirahata e Chu (Ann. Inst. Statist. Math. 44 (1992) 579), começamos por estabelecer um desenvolvimento assintótico para o erro quadrático integrado (ISE)

$$\int_{\mathbb{R}} \{\hat{F}_n(x) - F(x)\}^2 dF(x),$$

para estimadores automáticos do núcleo cujas sucessões (A_n) satisfazem a condição $A_n/h_n - 1 = o_p(1)$, para alguma sucessão determinista (h_n) nas condições acima descritas. Com base num tal desenvolvimento, descrevemos o comportamento assintótico do ISE em função de (A_n) , e comparamos, no sentido do ISE, não só os estimadores EDF e AKDF, bem como estimadores automáticos do núcleo definidos a partir dum mesmo núcleo e de janelas distintas.

Os resultados obtidos permitem concluir, em particular, que, de um ponto de vista assintótico, a escolha da janela (A_n) não tem na estimação da função de distribuição

o papel de relêvo que assume no contexto da estimação da densidade de probabilidade pelo método do núcleo. Estimadores automáticos do núcleo cuja sucessão determinista associada não é ótima no sentido do erro quadrático médio integrado (MISE), podem ser indistinguíveis, no sentido do ISE, de estimadores cuja sucessão determinista associada é ótima no sentido do MISE.
