Selecção multietápica da janela para o estimador do núcleo da função de distribuição

Ana Paula Martins
Universidade da Beira Interior, Departamento de Matemática

Carlos Tenreiro[1]
Universidade de Coimbra, Departamento de Matemática

Resumo: Neste artigo propomos versões multietápicas do método *plug-in* para a escolha da janela do estimador do núcleo da função de distribuição introduzido por Altman e Léger [*J. Statist. Plann. Inference* 46, 1995, 195–214]. Um estudo de simulação é desenvolvido para comparar os seguintes estimadores *plug-in*: o estimador de Altman e Léger, os estimadores a zero e a duas etapas, e o estimador das distribuições de referência. O estimador *plug-in* bietápico revela-se o melhor dos estimadores considerados.

Palavras–chave: Função de distribuição; Estimador do núcleo; Selecção da janela.

Abstract: In this paper we propose multistage versions of the plug-in bandwidth selector introduced by Altman and Léger [*J. Statist. Plann. Inference* 46, 1995, 195–214] for kernel distribution function estimators. A simulation study is undertaken to compare the following plug-in kernel estimators: the Altman and Léger estimator, the zero and two stages estimators, and the Normal reference rule estimator. The two stages estimator has a better performance in comparison with the others.

Keywords: Distribution function; Kernel estimator; Plug-in bandwidth selection.

MSC2000: 62G05.

1 Introdução

Sendo $X_1, \ldots, X_n$ variáveis aleatórias reais, independentes e absolutamente contínuas com densidade comum f, o estimador do núcleo da função de distribuição F introduzido por Tiago de Oliveira [22], Nadaraya [13] e Watson e Leadbetter [24], é definido, para $x \in \mathbb{R}$, por

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

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

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

[1]Parcialmente financiado por CMUC/FCT.

com K uma função integrável tal que $\int_{\mathbb{R}} K(u)du = 1$ (núcleo), e (h_n) uma sucessão de números reais estritamente positivos convergindo para zero quando $n \to +\infty$.

A qualidade global da estimação produzida por $\widehat{F}_n$ é habitualmente medida pelo erro quadrático integrado ponderado

$$ISE(h_n, \omega) = \int_{\mathbb{R}} \{\widehat{F}_n(x) - F(x)\}^2 \omega(x) dx, \tag{1}$$

ou pelo erro quadrático médio integrado ponderado

$$MISE(h_n, \omega) = \mathrm{E}\left(\int_{\mathbb{R}} \{\widehat{F}_n(x) - F(x)\}^2 \omega(x) dx \right), \tag{2}$$

onde ω é uma função de peso. O comportamento assintótico destas duas medidas foi estudado por Swanepoel [19], Jones [9] e Shirahata e Chu [17]. Para um núcleo fixo, a janela h_n é habitualmente escolhida em função das observações $X_1, \ldots, X_n$. Tal procedimento conduz a uma classe mais vasta de estimadores de F, ditos estimadores automáticos do núcleo, onde $h_n = \widehat{h}_n(X_1, \ldots, X_n)$ é uma sucessão de funções mensuráveis. Neste contexto, o comportamento assintótico de (1) foi estudado por Tenreiro [20]. A selecção da janela pelo método da validação cruzada é efectuada por Sarda [16] para funções de peso de suporte compacto e por Bowman *et al.* [3] no caso $\omega = 1$. A metodologia *plug-in* é introduzida por Altman e Léger [1] no caso $\omega = f$, e Polansky e Baker [14] implementam um estimador *plug-in* multietápico no caso $\omega = 1$.

Neste artigo introduzimos versões multietápicas do estimador *plug-in* de Altman e Léger [1]. A dificuldade adicional deste caso relativamente ao caso $\omega = 1$, já considerado por Polansky e Baker [14], reside no facto da janela óptima depender dum parâmetro adicional para o qual é necessário introduzir estimadores multietápicos. Com efeito, se F possui derivada de segunda ordem contínua e limitada, e K é um núcleo com $\int y^2 |K(y)| dy < +\infty$, $\int yK(y)dy = 0$, $\int y^2 K(y)dy \neq 0$ e $\int yK(y)\bar{K}(y)dy > 0$, sabemos que (cf. Swanepoel [19])

$$h_{MISE}(F) = \mathrm{argmin}_{h>0} MISE(h, f) = h_{AMISE}(1 + o(1)),$$

onde

$$h_{AMISE} = \left(C_K \int F^{(1)}(x) dF(x) \Big/ \int F^{(2)}(x)^2 dF(x) \right)^{1/3} n^{-1/3}, \tag{3}$$

com

$$C_K = 2 \int yK(y)\bar{K}(y)dy \Big/ \left(\int y^2 K(y)dy \right)^2.$$

Com o objectivo de definir estimadores multietápicos para os parâmetros $\int F^{(1)}(x)dF(x)$ e $\int F^{(2)}(x)^2 dF(x)$, interessamo-nos no §2 pela estimação dos parâmetros

$$\theta_r = \int F^{(r)}(x) dF(x),$$

para $r = 1, 3, 5, \ldots$, e de

$$\theta_{r,r+s} = \int F^{(r)}(x) F^{(r+s)}(x) dF(x),$$

para $r = 1, 2, \ldots$ e $s = 0, 1, 2, \ldots$. A estimação de θ_r foi considerada por Hall e Marron [7, 8], Bickel e Ritov [2] e Jones e Sheather [10], enquanto a de $\theta_{r,r+s}$ foi considerada por Altman e Léger [1] para $r = 2$ e $s = 0$. O resultado que apresentamos sobre a estimação de θ_r é no essencial devido a Jones e Sheather [10]. No entanto, tal resultado é obtido aqui para uma classe mais vasta de estimadores cujos núcleos gozam de certas propriedades de optimalidade na estimação das derivadas duma densidade de probabilidade (cf. Gasser *et al.* [5] e Granovsky *et al.* [6]). O resultado que obtemos sobre a estimação de $\theta_{r,r+s}$ é o principal resultado deste artigo, generalizando e corrigindo o obtido por Altman e Léger [1].

No §3 propomos um estudo de simulação envolvendo o estimador *plug-in* de Altman e Léger, dois estimadores *plug-in* a zero e a duas etapas, e o estimador das distribuições de referência. O estimador *plug-in* bietápico revela-se o melhor dos quatro estimadores.

As demonstrações dos resultados enunciados são remetidas para o §4.

2 Estimação multietápica de θ_r e $\theta_{r,r+s}$

Para $r \in \{0, 1, 2, \ldots\}$ e $\ell \in \{2, 4, 6, \ldots\}$ denotaremos por $\mathcal{D}_b(r)$ o conjunto das funções de distribuição com derivadas limitadas até à ordem r, e por $\mathcal{K}_r(\ell)$ o conjunto dos núcleos K limitados de ordem $(r, r+\ell)$, satisfazendo $\int |y|^{r+\ell+1} |K(y)|\, dy < +\infty$. Assim,

$$\mu_j(K) := \int y^j K(y)\, dy = \begin{cases} (-1)^r r!\, \delta_{j,r}, & \text{se } j = 0, 1, \ldots, r+\ell-1 \\ \beta_{r+\ell} \neq 0, & \text{se } j = r+\ell, \end{cases}$$

onde $\delta_{j,r}$ é o delta de Kronecker (cf. Gasser *et al.* [5]). Descrevemos a seguir uma forma simples de construir núcleos em $\mathcal{K}_r(\ell)$. Consideremos uma função real de variável real K_0 com $\int |y|^{2r+2\ell-1} |K_0(y)| dy < +\infty$ e $\sup_{y \in \mathbb{R}} |y|^{r+\ell-2} |K_0(y)| < +\infty$, e definamos as matrizes N_r e $M_r(u)$ de tipo $(r+\ell-1) \times (r+\ell-1)$, onde N_r tem elemento genérico (i, j) dado por $\mu_{i+j-2}(K_0)$, e $M_r(u)$ é definida como N_r mas com a coluna $r+1$ substituída por $(1, u, \ldots, u^{r+\ell-2})^T$. Então

$$K_{r,r+\ell}(u) = (-1)^r r! \left\{ \det(M_r(u)) / \det(N_r) \right\} K_0(u), \tag{4}$$

é um núcleo em $\mathcal{K}_r(\ell)$ (cf. Lejeune e Sarda [11] e Ruppert e Wand [15]).

2.1 Estimação de θ_r

Para $r = 1, 3, 5, \ldots$, denotemos por $\widehat{\theta}_r$ o estimador definido por (cf. Hall e Marron [7, 8] e Jones e Sheather [10])

$$\widehat{\theta}_r = \frac{C_{r,\ell}}{nh_n^r} + \frac{1}{n^2 h_n^r} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} U\left(\frac{X_i - X_j}{h_n}\right), \tag{5}$$

onde $U \in \mathcal{K}_{r-1}(\ell)$, para $\ell \in \{2, 4, 6, \dots\}$ fixo, e $C_{r,\ell}$ é uma constante real não-nula que fixaremos mais tarde.

No resultado seguinte apresentamos um desenvolvimento assintótico para o viés de $\widehat{\theta}_r$ e determinamos a ordem de convergência da sua variância. A sua demonstração é remetida para o §4.

Teorema 1. *Sejam $U \in \mathcal{K}_{r-1}(\ell)$ e $F \in \mathcal{D}_b(r+\ell+1)$. Então:*

a) $\mathrm{E}(\widehat{\theta}_r) - \theta_r = \dfrac{C_{r,\ell}}{nh_n^r} + h_n^\ell \dfrac{\mu_{r+\ell-1}(U)}{(r+\ell-1)!}\theta_{r+\ell} + O\left(\dfrac{1}{n} + h_n^{\ell+1}\right)$;

b) $\mathrm{Var}(\widehat{\theta}_r) = O\left(\dfrac{1}{n} + \dfrac{2}{n^2 h_n^{2r-1}}\right)$.

Admitiremos no que se segue que $h_n \to 0$ e $nh_n^r \to +\infty$, o que implica a convergência em média quadrática de $\widehat{\theta}_r$ para θ_r, quando $n \to +\infty$. A janela óptima, h_r^{opt}, no sentido da minimização assintótica do erro quadrático médio $MSE(\widehat{\theta}_r) = \mathrm{Var}(\widehat{\theta}_r) + (\mathrm{E}(\widehat{\theta}_r) - \theta_r)^2$, depende do sinal de $C_{r,\ell}\,\mu_{r+\ell-1}(U)\theta_{r+\ell}$.

Corolário 1. *Nas condições anteriores, admitamos que $\theta_{r+\ell} \neq 0$.*

a) Se $C_{r,\ell}\,\mu_{r+\ell-1}(U)\theta_{r+\ell} < 0$, então

$$h_r^{opt} = \left(\frac{-C_{r,\ell}(r+\ell-1)!}{\mu_{r+\ell-1}(U)\,\theta_{r+\ell}}\right)^{1/(r+\ell)} n^{-1/(r+\ell)},$$

e

$$MSE(\widehat{\theta}_r) = O\left(n^{-1} + n^{-(2\ell+1)/(r+\ell)}\right).$$

b) Se $C_{r,\ell}\,\mu_{r+\ell-1}(U)\theta_{r+\ell} > 0$, então

$$h_r^{opt} = \left(\frac{C_{r,\ell} r(r+\ell-1)!}{\ell\mu_{r+\ell-1}(U)\,\theta_{r+\ell}}\right)^{1/(r+\ell)} n^{-1/(r+\ell)},$$

e

$$MSE(\widehat{\theta}_r) = O\left(n^{-1} + n^{-2\ell/(r+\ell)}\right).$$

Atendendo ao resultado anterior, e uma vez que, sob condições gerais sobre F, $\theta_{4k-1} < 0$ e $\theta_{4k-3} > 0$, para todo o $k \in \mathbb{N}$, vamos, no que se segue, tomar $C_{r,\ell}$ tal que $C_{r,\ell}\,\mu_{r+\ell-1}(U) > 0$, se $r+\ell = 4k-1$, para algum $k \in \mathbb{N}$, e $C_{r,\ell}\,\mu_{r+\ell-1}(U) < 0$, se $r+\ell = 4k-3$, para algum $k \in \mathbb{N}$.

Para $\ell \in \{2, 4, 6, \dots\}$, fixo, os resultados anteriores permitem-nos apresentar o seguinte esquema de estimação de θ_r a k-etapas, com $k \in \mathbb{N}_0$ (ver Wand e Jones [23], pg. 67–70, para um estimador análogo com $k = \ell = 2$, e Tenreiro [21], sobre o comportamento assintótico dum estimador a k-etapas análogo ao que aqui consideramos):

Etapa 0: Estimar $\theta_{r+k\ell}$ a partir de $\widehat{\theta}_{r+k\ell}$ com janela $h_{r+k\ell}^{opt}$ onde no cálculo de $\theta_{r+(k+1)\ell}$ a distribuição desconhecida F é substituída pela distribuição de referência $F_{N(0,\sigma^2)}$, isto é, $\theta_{r+(k+1)\ell}$ é substituído por

$$R_{r+(k+1)\ell}(\widehat{\sigma}) = \int F_{N(0,\widehat{\sigma}^2)}^{(r+(k+1)\ell)}(x) dF_{N(0,\widehat{\sigma}^2)}(x),$$

com $F_{N(0,\sigma^2)}$ a função de distribuição da distribuição normal de média 0 e variância σ^2, e $\widehat{\sigma}$ é um estimador da escala da distribuição desconhecida F.

Etapa b ($1 \leq b \leq k$): Estimar $\theta_{r+(k-b)\ell}$ usando $\widehat{\theta}_{r+(k-b)\ell}$ com janela $h_{r+(k-b)\ell}^{opt}$, onde $\theta_{r+(k-b+1)\ell}$ é substituído por $\widehat{\theta}_{r+(k-b+1)\ell}$.

2.2 Estimação de $\theta_{r,r+s}$

Para $r = 1, 2, \ldots$ e $s = 0, 1, 2, \ldots$, consideramos o estimador $\widehat{\theta}_{r,r+s}$ definido por (cf. Altman e Léger [1], para $r = 2$ e $s = 0$)

$$\widehat{\theta}_{r,r+s} = \frac{1}{n(n-1)^2 h_n^{2r+s}} \sum_{i=1}^{n} \sum_{\substack{j \neq i \\ k \neq i}} V\left(\frac{X_i - X_j}{h_n}\right) W\left(\frac{X_i - X_k}{h_n}\right), \qquad (6)$$

onde $V \in \mathcal{K}_{r-1}(\ell)$ e $W \in \mathcal{K}_{r+s-1}(\ell)$, para $\ell \in \{2, 4, 6, \ldots\}$ fixo.

O viés assintótico e a ordem de convergência da variância de $\widehat{\theta}_{r,r+s}$ são dados no resultado seguinte cuja demonstração é apresentada no §4.

Teorema 2. *Sejam* $V \in \mathcal{K}_{r-1}(\ell)$, $W \in \mathcal{K}_{r+s-1}(\ell)$ *e* $F \in \mathcal{D}_b(r+s+\ell+1)$. *Então:*

a) $\mathrm{E}(\widehat{\theta}_{r,r+s}) - \theta_{r,r+s}$

$$= h_n^{\ell} \kappa_{r,r+s}(V, W) + \frac{\mu_0(VW)}{n h_n^{2r+s-1}} \theta_1 + O\left(\frac{1}{n h_n^{2r+s-2}} + h_n^{\ell+1}\right),$$

onde

$$\kappa_{r,r+s}(V, W) = \frac{(-1)^{r-1} \mu_{r+\ell-1}(V)}{(r+\ell-1)!} \theta_{r+\ell,r+s} + \frac{(-1)^{r+s-1} \mu_{r+s+\ell-1}(W)}{(r+s+\ell-1)!} \theta_{r,r+s+\ell};$$

b) $\mathrm{Var}(\widehat{\theta}_{r,r+s}) = O\left(\frac{1}{n} + \frac{1}{n^2 h_n^{4r+2s-3}} + \frac{1}{n^3 h_n^{4r+2s-2}} + \frac{1}{n^4 h_n^{4r+2s-1}}\right).$

Decorre do resultado anterior que a convergência em média quadrática de $\widehat{\theta}_{r,r+s}$ para $\theta_{r,r+s}$, quando $n \to +\infty$, ocorre sempre que $h_n \to 0$ e $n h_n^{2r+s-1} \to$

$+\infty$, o que assumiremos. A janela óptima, $h^{opt}_{r,r+s}$, no sentido da minimização assintótica do erro quadrático médio $MSE(\widehat{\theta}_{r,r+s})$, depende agora não só de $\mu_0(VW)$ ser ou não zero, mas também do sinal de $\kappa_{r,r+s}(V,W)\,\mu_0(VW)$. No resultado seguinte analisamos apenas o caso $\mu_0(UV) \neq 0$.

Corolário 2. *Nas condições anteriores, admitamos que* $\kappa_{r,r+s}(V,W) \neq 0$.
a) Se $\kappa_{r,r+s}(V,W)\,\mu_0(VW) < 0$, *então*

$$h^{opt}_{r,r+s} = \left(\frac{-\mu_0(VW)\,\theta_1}{\kappa_{r,r+s}(V,W)}\right)^{1/(2r+s+\ell-1)} n^{-1/(2r+s+\ell-1)},$$

e

$$MSE(\widehat{\theta}_{r,r+s}) = O\Big(n^{-(2\ell+1)/(2r+s+\ell-1)}\Big).$$

b) Se $\kappa_{r,r+s}(V,W)\,\mu_0(VW) > 0$, *então*

$$h^{opt}_{r,r+s} = \left(\frac{(2r+s-1)\mu_0(VW)\,\theta_1}{\ell\,\kappa_{r,r+s}(V,W)}\right)^{1/(2r+s+\ell-1)} n^{-1/(2r+s+\ell-1)},$$

e

$$MSE(\widehat{\theta}_{r,r+s}) = O\Big(n^{-2\ell/(2r+s+\ell-1)}\Big).$$

De forma análoga ao que fizemos atrás, e admitindo que possuímos já um estimador de θ_1, os resultados anteriores permitem-nos apresentar o seguinte esquema de estimação de $\theta_{r,r+s}$ a k-etapas, com $k \in \mathbb{N}_0$, onde $\ell \in \{2,4,6,\dots\}$ está fixo à partida:

Etapa 0: Estimar $\theta_{r+i,r+s+j}$, para $i,j = 0,\ell,2\ell,\dots,k\ell$ e $i+j = k\ell$, utilizando $\widehat{\theta}_{r+i,r+s+j}$ com janela $h^{opt}_{r+i,r+s+j}$ onde $\theta_{r+i,r+s+j}$ é substituído por

$$R_{r+i,r+s+j}(\widehat{\sigma}) = \int F^{(r+i)}_{N(0,\widehat{\sigma}^2)}(x) F^{(r+s+j)}_{N(0,\widehat{\sigma}^2)}(x) dF_{N(0,\widehat{\sigma}^2)}(x).$$

Etapa b $(1 \leq b \leq k)$: Estimar $\theta_{r+i,r+s+j}$, para $i,j = 0,\ell,2\ell,\dots,(k-b)\ell$ e $i+j = (k-b)\ell$, a partir de $\widehat{\theta}_{r+i,r+s+j}$ com janela $h^{opt}_{r+i,r+s+j}$ definida com $\widehat{\theta}_{r+i+\ell,r+s+j}$ e $\widehat{\theta}_{r+i,r+s+j+\ell}$ no lugar de $\theta_{r+i+\ell,r+s+j}$ e $\theta_{r+i,r+s+j+\ell}$, respectivamente.

3 Estudo de simulação

Propomos neste parágrafo um estudo de simulação envolvendo quatro estimadores do núcleo da função de distribuição que descreveremos a seguir, e um conjunto de distribuições de probabilidade constituído por 15 misturas de normais definidas no §3 de Marron e Wand [12]. Os gráficos das suas densidades e funções de distribuição são apresentados na Figura 1.

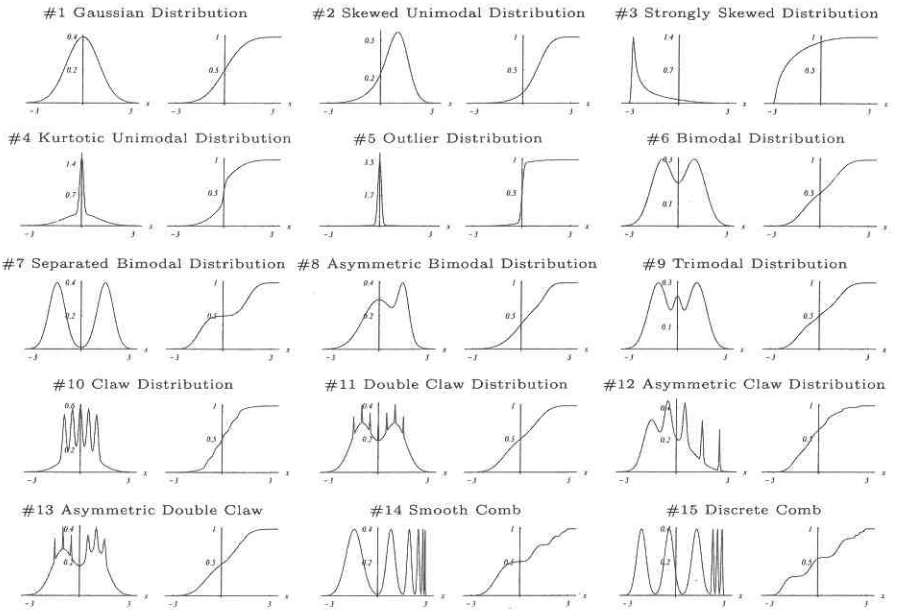

Figura 1: Mistura de normais consideradas no estudo de simulação

Dos resultados de Swanepoel [19] e Jones [9], sabemos que o núcleo uniforme é óptimo no sentido da minimização do erro quadrático médio integrado (2). No entanto, como nesse caso as estimativas produzidas por $\widehat{F}_n$ não são contínuas, não reflectindo a propriedade respectiva de F, tomaremos para K o núcleo normal standard $K(x) = \exp(-x^2/2)/\sqrt{2\pi}$. Relativamente à escolha da janela, e tendo em conta (3) tomaremos

$$\widehat{h}_n = \left(C_K\, \widehat{\theta}_1/\widehat{\theta}_{2,2}\right)^{1/3} n^{-1/3},$$

onde $C_K = 1/\sqrt{\pi}$, e $\widehat{\theta}_1$ e $\widehat{\theta}_{2,2}$ estimadores de θ_1 e $\theta_{2,2}$, respectivamente.

Em cada um dos estimadores $\widehat{\theta}_r$ e $\widehat{\theta}_{r,r+s}$ que consideramos, tomamos $\ell = 2$ e núcleos da forma (4) com $K_0 = K$. Assim $\widehat{\theta}_r$ será definido por (5) com $U = K_{r-1,r+\ell-1}$ e $C_{r,\ell} = K_{r-1,r+\ell-1}(0)$, e $\widehat{\theta}_{r,r+s}$ será definido por (6) com $V = K_{r-1,r+\ell-1}$ e $W = K_{s-1,s+\ell-1}$. Quando optamos por um esquema de estimação multietápica, a etapa 0 de estimação descrita no §2 depende da fixação dum estimador da escala da distribuição desconhecida F. Seguindo a sugestão de Silverman [18], pg. 47, consideramos $\widehat{\sigma} = \min\left(\widehat{S}, \widehat{R}/1.349\right)$, onde $\widehat{S}$ é o desvio-padrão empírico e $\widehat{R}$ a amplitude interquartil empírica.

Estimador *plug-in* de Altman e Léger (AL): Em Altman e Léger [1], $\widehat{\theta}_1$ e $\widehat{\theta}_{2,2}$ são definidos com janelas deterministas iguais a $n^{-0.3}$. Tomaremos aqui

tais janelas iguais a $\widehat{\sigma}\, n^{-1/3}$ e $\widehat{\sigma}\, n^{-1/5}$, respectivamente, que, como vimos, são as ordens de convergência para zero das janelas óptimas, e $\widehat{\sigma}$ é introduzido para correcção do efeito de escala.

Estimador *plug-in* a zero etapas (PI-0): Neste estimador $\widehat{\theta}_1$ e $\widehat{\theta}_{2,2}$ são estimadores a zero etapas de θ_1 e $\theta_{2,2}$, respectivamente.

Estimador *plug-in* a duas etapas (PI-2): Os estimadores $\widehat{\theta}_1$ e $\widehat{\theta}_{2,2}$ são aqui estimadores a duas etapas de θ_1 e $\theta_{2,2}$, respectivamente. Na implementação deste último estimador, substituímos θ_1 pelo seu estimador a duas etapas.

Estimador das distribuições de referência (DRN): O método das distribuições de referência foi introduzido por Deheuvels [4] no contexto da estimação da densidade de probabilidade e consiste na substituição da distribuição desconhecida F que surge na expressão (3) de h_{AMISE}, por uma distribuição de referência que consideramos normal $N(0, \widehat{\sigma}^2)$, onde $\widehat{\sigma}$ é um estimador da escala da distribuição.

Distr.	n	AL	PI-0	PI-2	DRN	Distr.	n	AL	PI-0	PI-2	DRN
#1	20	93.84	89.47	91.56	92.64	#2	20	94.63	86.87	92.59	94.07
	50	96.21	93.35	94.64	96.03		50	94.90	85.06	94.74	95.68
	100	97.33	94.34	96.75	97.88		100	96.76	90.22	96.47	97.20
	200	98.42	97.31	98.02	98.79		200	97.77	94.06	97.49	98.13
	500	98.77	97.72	98.64	99.21		500	98.55	95.59	98.71	98.52
#3	20	88.69	67.84	94.51	91.05	#4	20	92.40	79.94	95.50	94.39
	50	77.60	50.07	89.02	74.62		50	81.71	64.06	89.66	82.57
	100	67.55	36.54	85.83	57.84		100	81.63	59.42	93.18	78.79
	200	72.00	34.04	93.41	53.00		200	84.98	96.59	95.81	97.76
	500	74.15	32.83	95.82	44.86		500	90.46	53.60	98.80	71.21
#5	20	93.22	80.30	93.04	92.51	#6	20	99.79	99.59	96.49	98.17
	50	95.47	82.56	95.99	95.26		50	101.03	98.37	98.54	101.62
	100	96.64	89.18	96.67	96.75		100	99.84	92.63	99.14	100.30
	200	97.40	93.18	97.40	97.43		200	100.34	96.46	99.42	100.60
	500	98.40	96.04	98.44	98.30		500	99.49	93.99	99.52	99.54
#7	20	91.63	77.01	100.48	93.73	#8	20	101.36	98.21	98.02	100.69
	50	89.12	54.73	100.17	83.46		50	102.97	97.59	101.53	102.85
	100	88.71	35.27	99.57	72.34		100	100.86	95.99	100.03	101.68
	200	93.99	29.40	99.70	71.02		200	99.92	95.28	99.75	100.42
	500	96.96	34.82	99.69	70.37		500	99.64	95.62	99.52	99.81
#9	20	106.16	104.54	101.58	104.78	#10	20	120.70	117.01	117.57	118.86
	50	103.62	99.44	101.55	103.91		50	113.38	107.05	111.81	114.04
	100	101.56	93.28	101.22	101.81		100	107.48	103.65	107.08	107.63
	200	100.37	92.06	100.25	100.57		200	103.04	101.08	102.83	103.13
	500	99.70	94.02	99.86	99.59		500	90.17	87.44	91.33	91.08
#11	20	129.71	126.92	124.28	128.15	#12	20	119.96	114.32	117.03	118.15
	50	119.29	115.62	117.01	119.23		50	110.69	103.38	109.67	111.03
	100	112.79	109.18	111.49	112.88		100	108.04	103.17	107.34	108.89
	200	109.44	104.61	108.78	109.81		200	101.67	97.96	101.94	102.98
	500	105.62	100.81	105.55	105.68		500	95.47	90.21	98.64	96.33
#13	20	125.74	122.17	122.07	124.48	#14	20	113.09	105.87	113.40	114.02
	50	116.63	110.93	114.73	116.97		50	95.07	75.77	103.09	95.49
	100	111.55	102.40	110.86	112.15		100	85.18	53.98	100.65	80.36
	200	107.13	99.88	106.72	107.62		200	87.09	46.79	100.79	75.26
	500	103.37	96.45	103.63	103.19		500	88.15	40.38	99.95	66.70
#15	20	119.25	118.13	114.58	116.44						
	50	96.88	92.87	96.48	97.79						
	100	76.48	71.50	82.68	79.85						
	200	65.15	53.92	90.72	67.12						
	500	69.84	36.00	99.24	56.04						

Tabela 1. Eficiência (eff)

Os estimadores anteriores são comparados através da medida de eficiência,

$$\text{eff} = \sqrt{\frac{\mathrm{E}(ISE(h_{AMISE}))^2}{\mathrm{E}(ISE(\widehat{h}_n))^2}} \times 100\%$$

onde h_{AMISE} é dado por (3). Esta medida de eficiência reflecte não só a média (MISE) mas também a variabilidade associada ao erro quadrático integrado $ISE(\widehat{h}_n)$.

Para cada um dos quatros estimadores do núcleo descritos, para cada uma das distribuições de probabilidade consideradas, e para vários valores de n, apresentamos na Tabela 1 estimativas da medida de eficiência anterior baseadas em amostras de tamanho 500. Os resultados obtidos revelam que o estimador PI-2 é o melhor dos estimadores considerados. Os estimadores AL, PI-0 e DRN apesar de revelarem eficiência elevada para a maioria das distribuições consideradas, possuem baixa eficiência para as distribuições #3, #4, #7, #14 e #15.

4 Demonstrações

Demonstração do Teorema 1: a) Consequência da igualdade

$$\mathrm{E}U\left(\frac{X_i - X_j}{h_n}\right) = h_n \int\int U(z) f(x - zh_n) dz\, dF(x),$$

onde, para $U \in \mathcal{K}_{r-1}(\ell)$ e $F \in \mathcal{D}_b(r+\ell+1)$,

$$\begin{aligned}
&\int U(z) f(x - zh_n) dz \qquad (7)\\
&= \frac{(-1)^{r-1}}{(r-1)!} h_n^{r-1} \int U(z) z^{r-1} dz\, F^{(r)}(x)\\
&\quad + \frac{(-1)^{r+\ell-1}}{(r+\ell-1)!} h_n^{r+\ell-1} \int U(z) z^{r+\ell-1} dz\, F^{(r+\ell)}(x)\\
&\quad + \frac{(-1)^{r+\ell}}{(r+\ell-1)!} h_n^{r+\ell} \int U(z) z^{r+\ell} F^{(r+\ell+1)}(x - tzh_n) dz, \text{com } t \in]0,1[.
\end{aligned}$$

b) Temos

$$\mathrm{Var}(\widehat{\theta}_r) = \frac{1}{n^4 h_n^{2r}} \sum_{\sharp\{i_1,j_1\}\cap\{i_2,j_2\}\neq 0} K_n(i_1,j_1,i_2,j_2) + \frac{6-4n}{n(n-1)} \mathrm{E}^2\left(\widehat{\theta}_r - \frac{C_{r,\ell}}{nh_n^r}\right),$$

onde a soma é tomada para todos os índices $i_1, j_1, i_2, j_2 \in \{1, \ldots, n\}$, com $j_1 \neq i_1$ e $j_2 \neq i_2$, e

$$K_n(i_1,j_1,i_2,j_2) = \mathrm{E}U\left(\frac{X_{i_1} - X_{j_1}}{h_n}\right) U\left(\frac{X_{i_2} - X_{j_2}}{h_n}\right).$$

Para obter a alínea b) basta agora usar a) e ter em conta que $K_n(i_1,j_1,i_2,j_2) = h_n^{2r}\mathrm{E}(F^{(r)}(X_1))^2 + O(h_n^{2r+\ell})$, se $\sharp\{i_1,j_1\}\cap\{i_2,j_2\} = 1$, e $K_n(i_1,j_1,i_2,j_2) = O(h_n)$ se $\sharp\{i_1,j_1\}\cap\{i_2,j_2\} = 2$. ■

Demonstração do Teorema 2: a) Temos

$$\mathrm{E}(\widehat{\theta}_{r,r+s}) = \frac{1}{n(n-1)^2h_n^{2r+s}}\sum_{i=1}^{n}\sum_{\substack{j\neq i\\ k\neq i}} K_n(i,j,k),$$

onde

$$K_n(i,j,k) = \mathrm{E}V\left(\frac{X_i - X_j}{h_n}\right)W\left(\frac{X_i - X_k}{h_n}\right).$$

Utilizando (7) obtemos, para $F\in\mathcal{D}_b(r+s+\ell+1)$,

$$\begin{aligned} K_n(i,j,k) &= h_n^2\int\left\{\int V(z)f(x-zh_n)dz\right\}\left\{\int W(z)f(x-zh_n)dz\right\}dF(x)\\ &= h_n^{2r+s}\theta_{r,r+s} + h_n^{2r+s+\ell}\kappa_{r,r+s}(V,W) + O\big(h_n^{2r+s+\ell+1}\big), \end{aligned}$$

se $j\neq k$. Por outro lado, para $j=k$, obtemos $K_n(i,j,k) = h_n\mu_0(VW)\,\theta_1 + O(h_n^2)$.

A alínea a) é consequência imediata das igualdades anteriores.

b) Tendo em conta os desenvolvimentos obtidos acima para $K_n(i,j,k)$, podemos escrever

$$\begin{aligned} \mathrm{Var}(\widehat{\theta}_{r,r+s}) &= \frac{1}{n^2(n-1)^4h_n^{4r+2s}}\sum_{\sharp\{i_1,j_1,k_1\}\cap\{i_2,j_2,k_2\}\neq 0} K_n(i_1,j_1,k_1,i_2,j_2,k_2)\\ &\quad + O\left(\frac{1}{n} + \frac{1}{n^2h_n^{2r+s-1}} + \frac{1}{n^3h_n^{4r+2s-2}}\right), \end{aligned} \tag{8}$$

onde o somatório anterior é tomado para todos os índices $i_1,j_1,k_1,i_2,j_2,k_2\in\{1,\ldots,n\}$, com $j_1\neq i_1$, $k_1\neq i_1$, $j_2\neq i_2$ e $k_2\neq i_2$, e

$$\begin{aligned} &K_n(i_1,j_1,k_1,i_2,j_2,k_2)\\ &= \mathrm{E}V\left(\frac{X_{i_1}-X_{j_1}}{h_n}\right)W\left(\frac{X_{i_1}-X_{k_1}}{h_n}\right)V\left(\frac{X_{i_2}-X_{j_2}}{h_n}\right)W\left(\frac{X_{i_2}-X_{k_2}}{h_n}\right). \end{aligned}$$

Estudemos agora o termo $K_n = K_n(i_1,j_1,k_1,i_2,j_2,k_2)$ em cada um dos casos seguintes:

— $\sharp\{i_1,j_1,k_1\}\cap\{i_2,j_2,k_2\} = 1$: os termos mais significativos ocorrem quando i) $i_1 = i_2, j_1\neq k_1, j_2\neq k_2$, ii) $i_1 = i_2, j_1 = k_1, j_2\neq k_2$ e iii) $i_1 = i_2, j_1 = k_1, j_2 = k_2$, obtendo-se, respectivamente, i) $K_n = O(h_n^{4r+2s})$, ii) $K_n = O(h_n^{2r+s+1})$ e iii) $K_n = O(h_n^2)$. Assim,

$$\frac{1}{n^2(n-1)^4h_n^{4r+2s}}\sum_{\sharp\{i_1,j_1,k_1\}\cap\{i_2,j_2,k_2\}=1} K_n = O\left(\frac{1}{n} + \frac{1}{n^2h_n^{2r+s-1}} + \frac{1}{n^3h_n^{4r+2s-2}}\right)$$

— $\sharp\{i_1, j_1, k_1\} \cap \{i_2, j_2, k_2\} = 2$: os termos mais significativos ocorrem quando i) $i_1 = j_2, j_1 = k_2, j_1 \neq k_1$, ii) $i_1 = k_2, j_1 = j_2$, iii) $i_1 \neq i_2, j_1 = k_1 = j_2 = k_2$ e iv) $i_1 = i_2, j_1 = k_1 = k_2, j_2 \neq k_2$, obtendo-se, respectivamente, i) $K_n = O(h_n^{r+s+2})$, ii) $K_n = O(h_n^3)$, iii) $K_n = O(h_n^2)$ e iv) $K_n = O(h_n^{r+1})$. Assim,

$$\frac{1}{n^2(n-1)^4 h_n^{4r+2s}} \sum_{\sharp\{i_1,j_1,k_1\}\cap\{i_2,j_2,k_2\}=2} K_n = O\left(\frac{1}{n^2 h_n^{4r+2s-3}} + \frac{1}{n^3 h_n^{3r+2s-1}}\right).$$

— $\sharp\{i_1, j_1, k_1\} \cap \{i_2, j_2, k_2\} = 3$: os termos mais significativos ocorrem quando i) $i_1 = i_2, j_1 = j_2 = k_1 = k_2$ e ii) $i_1 = i_2, j_1 = j_2, k_1 = k_2, j_1 \neq k_1$ obtendo-se i) $K_n = O(h_n)$ e ii) $K_n = O(h_n^2)$, respectivamente. Assim,

$$\frac{1}{n^2(n-1)^4 h_n^{4r+2s}} \sum_{\sharp\{i_1,j_1,k_1\}\cap\{i_2,j_2,k_2\}=3} K_n = O\left(\frac{1}{n^3 h_n^{4r+2s-2}} + \frac{1}{n^3 h_n^{4r+2s-1}}\right).$$

As igualdades anteriores, conjuntamente com o desenvolvimento (8), permitem obter o resultado desejado. ■

Bibliografia

[1] Altman, N. e Léger, C. (1995). Bandwidth selection for kernel distribution function estimation. *J. Statist. Plann. Inference*, Vol. 46, p. 195–214.

[2] Bickel, P. e Ritov, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā*, Vol. 50, Ser. A, p. 381–393.

[3] Bowman, A., Hall, P. e Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika*, Vol. 85, p. 799–808.

[4] Deheuvels, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés. *Revue Statist. Appliquée*, Vol. 28, p. 25–55.

[5] Gasser, Th., Müller, H.-G. e Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Scand. J. Statist.*, Vol. 11, p. 197–211.

[6] Granovsky, B.L., Müller, H.G. e Pfeifer, C. (1995). Some remarks on optimal kernel functions. *Statist. Decisions*, Vol. 13, p. 101–116.

[7] Hall, P. e Marron, J.S. (1987). Estimation of integrated squared density derivatives. *Statist. Probab. Letters*, Vol. 6, p. 109–115.

[8] Hall, P. e Marron, J.S. (1991). Lower bounds for bandwidth selection in density estimation. *Probab. Th. Rel. Fields*, Vol. 90, p. 149–173.

[9] Jones, M.C. (1990). The performance of kernel density functions in kernel distribution function estimation. *Statist. Probab. Letters*, Vol. 9, p. 129–132.

[10] Jones, M.C. e Sheather, S.J. (1991). Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statist. Probab. Letters*, Vol. 11, p. 511–514.

[11] Lejeune, M. e Sarda, P. (1992). Smooth estimators of distribution and density functions. *Computat. Statist. Data Anal.*, Vol. 9, p. 129–132.

[12] Marron, J.S. e Wand, M.P. (1992). Exact mean integrated squared error. *Ann. Statist.*, Vol. 20, p. 712–736.

[13] Nadaraya, E.A. (1964). Some new estimates for distribution functions. *Theory Probab. Appl.*, Vol. 9, p. 497–500.

[14] Polansky, A.M. e Baker, E.R. (2000). Multistage plug-in bandwidth selection for kernel distribution function estimates. *J. Statist. Comput. Simul.*, Vol. 65, p. 63–80.

[15] Ruppert, D. e Wand, M.P. (1992). Multivariate locally weighted least squares regression. *Ann. Statist.*, Vol. 22, p. 1346–1370.

[16] Sarda, P. (1993). Smoothing parameter selection for smooth distribution functions. *J. Statist. Plann. Inference*, Vol. 35, p. 65–75.

[17] Shirahata, S. e Chu, I-S. (1992). Integrated squared error of kernel-type estimator of distribution function. *Ann. Inst. Statist. Math.*, Vol. 44, p. 579–591.

[18] Silverman, J. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman & Hall.

[19] Swanepoel, J. (1988). Mean integrated squared error properties and optimal kernels when estimating a distribution function. *Commun. Statist. –Theory Meth.*, Vol. 17, p. 3785–3799.

[20] Tenreiro, C. (2002). On the asymptotic behaviour of the ISE for automatic kernel distribution estimators. *Pré-publicação 02–06, DMUC.*

[21] Tenreiro, C. (2002). On the asymptotic normality of multistage integrated density derivatives estimators. *Pré-publicação 02–18, DMUC.*

[22] Tiago de Oliveira, J. (1963). Estatística de densidades, resultados assintóticos. *Rev. Fac. Ciências Lisboa*, Vol. 9, 2ª Sér. A, p. 111–206.

[23] Wand, M.P. e Jones, M.C. (1995). *Kernel Smoothing.* Chapman & Hall.

[24] Watson, G.S. e Leadbetter, M.R. (1964). Hazard Analysis II. *Sankhyā*, Vol. 26, Ser. A, p. 101–116.