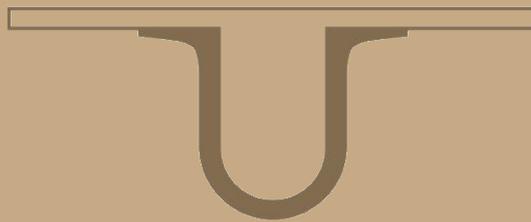




UNIVERSIDADE D
COIMBRA



Clara Isabel Alves Campos Lobo

**JANELAS *PLUG-IN* PARA O ESTIMADOR DO
NÚCLEO DA DENSIDADE BASEADAS NAS
FUNÇÕES DE HERMITE**

Dissertação no âmbito do Mestrado em Matemática, Ramo Estatística, Otimização e Matemática Financeira orientada pelo Professor Doutor Carlos Tenreiro e apresentada ao Departamento de Matemática da Faculdade de Ciências e Tecnologia.

Julho de 2019

Janelas plug-in para o estimador do núcleo da densidade baseadas nas funções de Hermite

Clara Isabel Alves Campos Lobo



UNIVERSIDADE DE
COIMBRA



Mestrado em Matemática

Master in Mathematics

Dissertação de Mestrado | MSc Dissertation

Julho 2019

Agradecimentos

A realização desta tese de mestrado não teria sido possível sem ajuda de algumas pessoas importantes para mim, às quais agradeço.

Ao Professor Doutor Carlos Tenreiro, pelo tema proposto, pela motivação dada, pela orientação, pela ajuda prestada, pela sua disponibilidade para solucionar dúvidas e discussões de resultados.

Às professoras Ana Paula Mouro e Graça Lopes, docentes na Escola Secundária Jaime Cortesão, pela ajuda prestada e por todo o apoio e incentivo dado.

Às minhas amigas, Inês Pedro, Rita Dias e Catarina Dinis, pela amizade, força e apoio dado.

Por fim, mas não menos importante, à minha família, em especial à Mãe, pelo apoio, carinho e pela paciência que teve comigo quando, ao longo do meu percurso académico, as coisas não corriam da melhor forma.

Resumo

A estimação da função densidade é um assunto muito estudado na estatística. Nesta dissertação, iremos recorrer ao estimador do núcleo, que depende do núcleo, K , e da janela h , para a estimação de f , porém, uma vez que a escolha da janela é mais relevante do que a escolha K , iremos desenvolver esta escolha ao longo da dissertação. Sob certas condições, provaremos a existência de uma janela ótima, $h_{0,n}$, que consiste no minimizante do erro quadrático médio integrado, MISE, do estimador do núcleo da densidade e analisaremos o seu comportamento assintótico. Com o intuito de obtermos expressões assintóticas para $h_{0,n}$, iremos obter desenvolvimentos assintóticos para o MISE e para as suas derivadas.

Pelas expressões assintóticas para $h_{0,n}$, verificamos que temos parâmetros dependentes de f , por isso desconhecidos. Assim, através das funções de Hermite vamos conseguir estimar estes parâmetros. Tendo presentes os estimadores obtidos analisaremos as suas convergências e, por fim, como estes estimadores dependem de um novo parâmetro m vamos efetuar a escolha automática deste valor e, em seguida, iremos procederemos ao estudo comparativo de simulação entre as janelas obtidas por nós através da estimação por funções de Hermite e janelas obtidas pelos métodos *plug-in* direto a duas etapas e da validação cruzada tendo f como uma mistura de densidades normais.

Conteúdo

Lista de Figuras	ix
1 Introdução	1
2 Existência e comportamento assintótico da janela ótima	3
2.1 Desenvolvimento exato do MISE	3
2.2 Existência da janela ótima	5
2.3 Comportamento assintótico da janela ótima	6
3 Expressões assintóticas para a janela ótima	11
3.1 Expressões assintóticas para o MISE	11
3.2 Desenvolvimento assintótico de $h_{0,n}$	15
3.3 Comparação de $h_{1,n}$ e $h_{2,n}$	19
4 Seleção da janela usando funções de Hermite	23
4.1 Estimadores de $R(f^{(p)})$ baseados nas funções de Hermite	23
4.2 Convergência em probabilidade de $\hat{R}_{1,p,m}$ e $\hat{R}_{2,p,m}$	26
4.3 Ordem de convergência	27
4.4 Comportamento assintótico das janelas $\hat{h}_{1,j,n}$ e $\hat{h}_{2,j,n}$	33
5 Janelas <i>plug-in</i> baseadas nas funções de Hermite	35
5.1 Escolha automática de m	35
5.2 Análise comparativa	38
Anexo A Alguns resultados auxiliares	45
Anexo B Códigos em R	47
Bibliografia	57

Lista de Figuras

3.1	Mistura de densidades normais #1, #2, #3, #6, #7, #8, #9 e #12 de (Marron e Wand, 1992)	20
3.2	Comparação de $h_{1,n}$ e $h_{2,n}$ para as misturas de densidades normais #1, #2, #3, #6, #7, #8, #9 e #12	21
5.1	Distribuição empírica de $ISE(f; n, \hat{h}_{1,1,n})$ em função de m , para $n=100$, com 500 repetições.	37
5.2	Distribuição $ISE(f; n, h)$ com $h = h_{1,j,n}$ e $h = h_{2,j,n}$, com $j = 1$ e $j = 2$ para $n = 100$. Com 500 repetições.	39
5.3	Norma L_2 dos ISE's das distribuições associadas as distribuições da figura 3.1	41
5.4	Norma L_2 dos ISE's das distribuições associadas as restantes distribuições presentes em (Marron e Wand, 1992)	42

Capítulo 1

Introdução

Ao longo das últimas décadas, a estimação de funções densidade f têm sido um tema bastante estudado para o qual têm sido propostos vários métodos para a resolução deste problema. No entanto, foi no ano 1956 que surgiu o atual estimador do núcleo introduzido num artigo de [Rosenblatt \(1956, p. 23\)](#) e as suas propriedades introduzidas por [Parzen \(1962\)](#). Estes dois autores foram os que deram início ao estudo da estimação não paramétrica pelo método do núcleo. De entre os estimadores não-paramétricos conhecidos, o estimador do núcleo é, sem dúvida o mais popular. Sendo X_1, X_2, \dots, X_n variáveis aleatórias reais, independentes e identicamente distribuídas com densidade desconhecida f , o estimador do núcleo de f é definido, para $x \in \mathbb{R}$, por

$$f_{n,h}(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i),$$

onde $h = h_n$ é uma sucessão de números reais estritamente positivos convergindo para zero quando n tende para infinito, K é uma função integrável (no sentido de Lebesgue) com $\int K(x)dx = 1$, e $K_h(x) = K(x/h)/h$. Para usarmos o estimador do núcleo temos que escolher o núcleo e a janela do estimador, dados respetivamente por K e h . Para K , a escolha mais comum é considerar núcleos positivos, sendo mais popular considerar K gaussiano. No entanto, a escolha que efetuaremos para K não é tão importante como a escolha de h . Assim, uma vez que a escolha da janela h é bastante relevante ([Wand and Jones, 1995, secção 2.7](#) e [Bosq and Lecoutre, 1978, cap.4, secção VII](#)), esta dissertação irá basear-se na escolha deste parâmetro que é crucial para o desempenho do estimador do núcleo.

Para efetuarmos essa escolha, no capítulo 2, baseado em [Chacón, Montanero, Nogales e Pérez \(2007\)](#), vamos começar por provar a existência e estudar o comportamento assintótico da janela ótima, $h_{0,n}$, no sentido da minimização do erro quadrático médio integrado (MISE, do inglês, mean integrated squared error), isto é,

$$h_{0,n} = \arg \min_{h>0} M_{f,n}(h)$$

onde

$$M_{f,n}(h) = MISE(f; n, h) = E \int (f_{n,h}(x) - f(x))^2 dx = \int E(f_{n,h}(x) - f(x))^2 dx$$

que representa uma medida global de qualidade e onde o integral, no sentido de Lebesgue, é estendido a \mathbb{R} . Este integral é finito desde que f e K sejam de quadrado integrável.

Esta escolha para a janela, pode ser feita através de vários métodos (ver [Wand and Jones, 1995](#), p. 58-89), um desses métodos é o método do *plug-in* direto que, sob certas condições para K e f , se baseia nas expressões assintóticas de $h_{0,n}$, dadas por

$$h_{1,n} = R(K)^{\frac{1}{5}} \mu_2(K)^{-\frac{2}{5}} R(f'')^{-\frac{1}{5}} n^{-\frac{1}{5}}$$

e

$$h_{2,n} = h_{1,n} + \frac{1}{20} R(K)^{\frac{3}{5}} \mu_2(K)^{-\frac{22}{5}} \mu_4(K) R(f'')^{-\frac{8}{5}} R(f''') n^{-\frac{3}{5}},$$

com $R(K) = \int K(u)^2 du$, $\mu_2(K) = \int u^2 K(u) du$ e $\mu_4(K) = \int u^4 K(u) du$. A obtenção destas expressões, feita através do $M_{f,n}(h)$ e das suas derivadas de primeira e segunda ordem, vai ser desenvolvida no capítulo 3 desta tese.

Atendendo a que nas expressões de $h_{1,n}$ e $h_{2,n}$ as únicas quantidades desconhecidas são os funcionais quadráticos $R(f^{(p)})$, com $p = 2$ e $p = 3$, o nosso problema da estimação de $h_{1,n}$ e $h_{2,n}$ fica reduzido à estimação destes funcionais. Esta estimação tem sido estudada por vários autores como, por exemplo, [Hall e Marron \(1987\)](#) e [Jones e Sheather \(1991\)](#) que propõem uma classe de estimadores do núcleo que são muito utilizados na prática, levando a métodos de seleção de janelas introduzidos, por exemplo, em [Hall, Sheather, Jones e Marron \(1991\)](#) e [Jones e Sheather \(1991\)](#). Tendo por base os artigos de [Walter \(1977\)](#) e [Greblicki e Pawlak \(1984\)](#), descreveremos, no capítulo 4 desta dissertação, a estimação das funcionais $R(f^{(p)})$ usando estimadores por projeção baseados no sistema ortogonal de Hermite. Os estimadores de Hermite que propomos para estimar a funcional $R(f^{(p)})$ dependem de um novo parâmetro de suavização m , que deve ser escolhido com base na amostra observada X_1, \dots, X_n .

No capítulo 5 desta tese, tomando como referência [Tenreiro \(2011\)](#), fazemos a escolha automática do parâmetro m e comparamos as janelas *plug-in* obtidas a partir de $h_{1,n}$ e $h_{2,n}$ com as janelas clássicas obtidas pelos métodos *plug-in* direto a duas etapas e da validação cruzada. O estudo de simulação que apresentamos para efetuar essa comparação envolve as misturas de densidades normais consideradas em [Marron e Wand \(1992\)](#).

Nesta dissertação, todos os gráficos e simulações foram efetuados usando o software R .

Capítulo 2

Existência e comportamento assintótico da janela ótima

Neste capítulo, baseado em [Chacón, Montanero, Nogales e Pérez \(2007\)](#), provamos que, sob certas condições em K e f , existe, para cada $n \in \mathbb{N}$, uma janela ótima, $h_{0,n}(f)$, no sentido da minimização do erro quadrático médio integrado, e estudamos o seu comportamento assintótico.

2.1 Desenvolvimento exato do MISE

Atendendo à decomposição usual do erro quadrático médio de $f_{n,h}$, no ponto x , como soma da variância e do quadrado do viés de $f_{n,h}$, o erro quadrático médio integrado pode ser escrito na forma

$$M_{f,n}(h) = V_{f,n}(h) + B_f(h), \quad (2.1)$$

onde a variância integrada, $V_{f,n}(h)$, e o viés quadrático integrado, $B_f(h)$, são dados, respetivamente, por

$$V_{f,n}(h) = \int \text{Var}(f_{n,h}(x)) dx \quad (2.2)$$

e

$$B_f(h) = \int (E(f_{n,h}(x)) - f(x))^2 dx. \quad (2.3)$$

No resultado seguinte apresentamos desenvolvimentos exatos para a variância integrada $V_{f,n}(h)$ e para o viés quadrático integrado $B_f(h)$, expressões essas que nos levarão a um desenvolvimento exato para MISE. Dada uma função real de variável real ψ , denotamos $\bar{\psi}(u) = \psi(-u)$. Vamos ainda representar por L_p , para $p \in \mathbb{N}$, o espaço das funções reais de variável real de potência p integrável e por L_∞ o espaço das funções reais de variável real limitadas.

Teorema 2.1. *Para $K \in L_2$ e $f \in L_2$, temos*

$$V_{f,n}(h) = \frac{R(K)}{nh} - \frac{1}{n} R_{K*\bar{K},h}(f) \quad (2.4)$$

e

$$B_f(h) = R_{K*\bar{K},h}(f) - 2R_{K,h}(f) + R(f), \quad (2.5)$$

onde $*$ representa o produto de convolução e

$$R_{K,h}(f) = \int K(u)f*\bar{f}(hu)du.$$

Demonstração. O estimador do núcleo tem por média

$$E(f_{n,h}(x)) = E(K_h(x - X_1)) = \int K_h(x - y)f(y)dy = K_h * f(x) \quad (2.6)$$

e por variância

$$V(f_{n,h}(x)) = \frac{1}{n}Var(K_h(x - X_1)) = \frac{1}{n} \left(\int K_h(x - y)^2 f(y)dy - K_h * f(x)^2 \right).$$

Partindo da expressão (2.2) temos

$$V_{f,n}(h) = \int \frac{1}{n} \left(\int K_h(x - y)^2 f(y)dy - K_h * f(x)^2 \right) dx = \frac{R(K)}{nh} - \frac{1}{n}R(K_h * f).$$

Para o viés, partindo da expressão (2.3) e atendendo a (2.6) temos

$$\begin{aligned} B_f(h) &= \int (E(f_{n,h}(x)))^2 dx - 2 \int E(f_{n,h}(x))f(x)dx + \int f(x)^2 dx \\ &= \int K_h * f(x)^2 dx - 2 \int (K_h * f)(x)f(x)dx + \int f(x)^2 dx \\ &= R(K_h * f) - 2 \int (K_h * f)(x)f(x)dx + R(f). \end{aligned}$$

Para concluir basta demonstrar as seguintes igualdades

$$\int K_h * f(x)f(x)dx = R_{K,h}(f) \quad e \quad R(K_h * f) = R_{K*\bar{K},h}(f).$$

Relativamente à primeira, temos

$$\begin{aligned} \int K_h * f(x)f(x)dx &= \iint K_h(x - y)f(x)f(y)dx dy = \iint K(u)f(x - uh)f(x)dx du \\ &= \int K(u)f*\bar{f}(hu)du = R_{K,h}(f). \end{aligned}$$

No que concerne à segunda, temos

$$\begin{aligned} R(K_h * f) &= \int K_h * f(x)^2 dx = \iiint K_h(x - y)K_h(x - z)f(y)f(z)dz dy dx \\ &= \iiint K_h(u + z - y)K_h(u)f(z)f(y)dz dy du \\ &= \iint \bar{K}_h * K_h(y - z)f(y)f(z)dz dy = R_{K*\bar{K},h}(f). \end{aligned}$$

Então, obtemos o pretendido. ■

Tendo em conta que (2.1) e o Teorema 2.1, temos o seguinte resultado.

Teorema 2.2. *Para $K \in L_2$ e $f \in L_2$ temos*

$$M_{f,n}(h) = \frac{R(K)}{nh} + \left(1 - \frac{1}{n}\right) R_{K*\bar{K},h}(f) - 2R_{K,h}(f) + R(f). \quad (2.7)$$

2.2 Existência da janela ótima

Nesta secção iremos provar que para cada n , existe um valor de h , que denotamos por $h_{0,n} = h_{0,n}(f)$, para o qual o MISE é mínimo, ao qual chamamos janela ótima do estimador do núcleo. Para tal, necessitamos do resultado seguinte.

Lema 2.1. *Sejam $K \in L_2$ e $f \in L_2$.*

(a) *A função $h \mapsto R_{K,h}(f)$, definida em $]0, \infty[$, é contínua e satisfaz*

$$i) \lim_{h \rightarrow 0} R_{K,h}(f) = R(f); \quad ii) \lim_{h \rightarrow \infty} R_{K,h}(f) = 0.$$

(b) *As funções $h \mapsto B_f(h)$, $h \mapsto V_{f,n}(h)$ e $h \mapsto M_{f,n}(h)$ são contínuas em $]0, \infty[$ e satisfazem*

$$\begin{aligned} i) \lim_{h \rightarrow 0} B_f(h) &= 0; & ii) \lim_{h \rightarrow \infty} B_f(h) &= R(f); & iii) \lim_{h \rightarrow 0} V_{f,n}(h) &= \infty; \\ iv) \lim_{h \rightarrow \infty} V_{f,n}(h) &= 0; & v) \lim_{h \rightarrow 0} M_{f,n}(h) &= \infty; & vi) \lim_{h \rightarrow \infty} M_{f,n}(h) &= R(f). \end{aligned}$$

(c) *Se $K \in L_\infty$ é contínuo em zero então*

$$i) \lim_{h \rightarrow \infty} hR_{K,h}(f) = K(0); \quad ii) \lim_{h \rightarrow \infty} h[M_{f,n}(h) - R(f)] = R(K) - 2K(0).$$

Demonstração. (a) Sabemos que se $f \in L_2$ então $f * \bar{f}$ é contínua e satisfaz $(f * \bar{f})(x) \rightarrow 0$, quando $|x| \rightarrow \infty$ (ver Rudin, 1990, p. 4). Assim, $f * \bar{f}$ é limitada e pelo Lema A.1 concluímos que $h \mapsto R_{K,h}(f)$ é contínua em $]0, \infty[$. Além disso, pelo TCD, temos

$$\lim_{h \rightarrow 0} R_{K,h}(f) = \lim_{h \rightarrow 0} \int K(u)(f * \bar{f})(hu) du = \int K(u)(f * \bar{f})(0) du = R(f),$$

e

$$\lim_{h \rightarrow \infty} R_{K,h}(f) = \lim_{h \rightarrow \infty} \int K(u)(f * \bar{f})(hu) du = \int K(u) \lim_{h \rightarrow \infty} (f * \bar{f})(hu) du = 0.$$

(b) Sendo $K * \bar{K}$ um núcleo, pela alínea (a) e (2.5) concluímos que

$$\begin{aligned} \lim_{h \rightarrow 0} B_f(h) &= \lim_{h \rightarrow 0} \left(R_{K*\bar{K},h}(f) - 2R_{K,h}(f) + R(f) \right) = R(f) - 2R(f) + R(f) = 0; \\ \lim_{h \rightarrow \infty} B_f(h) &= \lim_{h \rightarrow \infty} R_{K*\bar{K},h}(f) - 2R_{K,h}(f) + R(f) = R(f); \end{aligned}$$

e de (2.4) temos

$$\begin{aligned}\lim_{h \rightarrow 0} V_{f,n}(h) &= \lim_{h \rightarrow 0} \frac{R(K)}{nh} - \frac{R_{K*\bar{K},h}(f)}{n} = \infty; \\ \lim_{h \rightarrow \infty} V_{f,n}(h) &= \lim_{h \rightarrow \infty} \frac{R(K)}{nh} - \frac{R_{K*\bar{K},h}(f)}{n} = 0.\end{aligned}$$

Os limites $v)$ e $vi)$ são consequência direta dos limites anteriores e do Teorema 2.2.

(c) Do facto de $K * \bar{K}$ ser uma função contínua e por K ser limitado, temos, pelo TCD,

$$\begin{aligned}\lim_{h \rightarrow \infty} hR_{K,h}(f) &= \lim_{h \rightarrow \infty} h \int (K_h * f)(x)f(x)dx = \lim_{h \rightarrow \infty} h \iint K_h(x-y)f(y)f(x)dxdy \\ &= \lim_{h \rightarrow \infty} \iint K\left(\frac{x-y}{h}\right)f(x)f(y)dxdy = \iint K(0)f(x)f(y)dxdy = K(0)\end{aligned}$$

e

$$\begin{aligned}\lim_{h \rightarrow \infty} h[M_{f,n}(h) - R(f)] &= \lim_{h \rightarrow \infty} h \left[\frac{R(K)}{nh} + \left(1 - \frac{1}{n}\right) R(K_h * f) - 2R_{K,h}(f) + R(f) - R(f) \right] \\ &= \frac{R(K)}{n} - 2K(0) + \left(1 - \frac{1}{n}\right) \lim_{h \rightarrow \infty} hR_{K*\bar{K},h}(f) \\ &= \frac{R(K)}{n} - 2K(0) + \left(1 - \frac{1}{n}\right) K * \bar{K}(0) \\ &= \frac{R(K)}{n} - 2K(0) + \left(1 - \frac{1}{n}\right) R(K) \\ &= R(K) - 2K(0).\end{aligned}$$

■

Estamos agora em condições de enunciar o teorema que assegura a existência de $h_{0,n}$.

Teorema 2.3. *Sejam $K \in L_\infty$, contínuo em zero e tal que $R(K) < 2K(0)$ e f em L_2 . Então $M_{f,n}$ tem um minimizante em $]0, \infty[$, isto é,*

$$\exists h_{0,n}(f) > 0 : M_{f,n}(h_{0,n}(f)) \leq M_{f,n}(h), \quad \forall h > 0.$$

Demonstração. Do Lema 2.1 sabemos que $\lim_{h \rightarrow 0} M_{f,n}(h) = \infty$ e $\lim_{h \rightarrow \infty} M_{f,n}(h) = R(f)$. Do mesmo resultado temos que $\lim_{h \rightarrow \infty} h[M_{f,n}(h) - R(f)] = R(K) - 2K(0)$, valor este que, por hipótese, é negativo. Então, para h suficientemente grande, $M_{f,n}(h) < R(f)$, o que nos permite concluir que $M_{f,n}(h)$ tem um mínimo num ponto $h_{0,n}(f) \in]0, \infty[$.

■

2.3 Comportamento assintótico da janela ótima

Atendendo ao Teorema 2.2 e ao Lema 2.1, podemos concluir que $M_{f,n}(h_n) \rightarrow 0$, $n \rightarrow \infty$, quando a janela (h_n) é tal que $h_n \rightarrow 0$ e $nh_n \rightarrow \infty$, $n \rightarrow \infty$. É assim natural questionar se a sucessão $(h_{0,n}(f))$ também satisfaz as condições $h_{0,n} \rightarrow 0$ e $nh_{0,n}(f) \rightarrow \infty$, quando $n \rightarrow \infty$.

Teorema 2.4. *Sejam $K \in L_\infty$ e f em L_2 . Então*

$$\lim_{n \rightarrow \infty} nh_{0,n}(f) = \infty.$$

Demonstração. Começemos por provar $\lim_{n \rightarrow \infty} M_{f,n}(h_{0,n}(f)) = 0$. Suponhamos que $M_{f,n}(h_{0,n}(f)) \not\rightarrow 0$. Nestas condições, existe $\varepsilon > 0$ e uma subsucessão de funções tal que $M_{f,n_k}(h) \geq M_{f,n_k}(h_{0,n_k}(f)) > \varepsilon$ para todo $k \in \mathbb{N}$ e $h > 0$. Por outro lado, do Teorema 2.2 temos

$$\begin{aligned} \lim_{k \rightarrow \infty} M_{f,n_k}(h) &= \lim_{k \rightarrow \infty} \left(\frac{R(K)}{n_k h} + \left(1 - \frac{1}{n_k}\right) R_{K^* \bar{K}, h}(f) + R(f) - 2R_{k,h}(f) \right) \\ &= R_{K^* \bar{K}, h}(f) + R(f) - 2R_{K,h}(f) = B_f(h). \end{aligned} \quad (2.8)$$

Sabendo que $M_{f,n_k}(h) > \varepsilon$, de (2.8) concluímos que $B_f(h) > \varepsilon$, sendo $h > 0$, o que contradiz o Lema 2.1,(b) i).

Logo,

$$\lim_{n \rightarrow \infty} M_{f,n}(h_{0,n}(f)) = 0. \quad (2.9)$$

Agora, suponhamos que $\lim_{n \rightarrow \infty} nh_{0,n}(f) \neq \infty$, então existe uma constante $c > 0$ e uma subsucessão $(n_k h_{0,n_k}(f))_{k \in \mathbb{N}}$ tal que $n_k h_{0,n_k}(f) \leq c, \forall k \in \mathbb{N}$. Portanto, $\lim_{k \rightarrow \infty} h_{0,n_k}(f) = 0$.

Do Teorema 2.2 e de (2.9) concluímos

$$\lim_{k \rightarrow \infty} \left(\frac{R(K)}{n_k h_{0,n_k}} + \left(1 - \frac{1}{n_k}\right) R_{K^* \bar{K}, h_{0,n_k}}(f) - 2R_{K, h_{0,n_k}}(f) + R(f) \right) = 0.$$

Aplicando o Lema 2.1, (a), i) temos

$$\lim_{k \rightarrow \infty} \frac{R(K)}{n_k h_{0,n_k}} = 0,$$

o que implica que $n_k h_{0,n_k} \rightarrow \infty, k \rightarrow \infty$, contradizendo o facto de $n_k h_{0,n_k} \leq c$. Fica assim demonstrado que $\lim_{n \rightarrow \infty} nh_{0,n} = \infty$. ■

No que se segue vamos mostrar que a sucessão $h_{0,n}(f)$ converge para zero, quando $n \rightarrow \infty$, donde a transformada de Fourier de K , definida por

$$\varphi_K(t) = \int e^{itx} K(x) dx,$$

não seja identicamente igual a 1 numa vizinhança da origem. Tal é o caso dos núcleos de 2.^a ordem, isto é, núcleos para os quais $\int u^2 |K(u)| du < \infty, \mu_1(K) = \int u K(u) du = 0$ e $\mu_2(K) \neq 0$, onde $\mu_p(K) = \int u^p K(u) du$ é o momento de ordem $p \in \mathbb{N}$ de K , sempre que $\int |u|^p |K(u)| du < \infty$.

Proposição 2.1. *Se K é um núcleo de 2.^a ordem então φ_K não identicamente igual a 1 numa vizinhança da origem.*

Demonstração. Partindo da integrabilidade de K , $xK(x)$ e $x^2K(x)$ e tendo em conta o Lema A.2, temos que

$$\varphi'_K(t) = \int ixK(x)e^{itx} dx \quad e \quad \varphi''_K(t) = - \int x^2K(x)e^{itx} dx.$$

Então, se as hipótese não fossem válidas, tínhamos $0 = \varphi''(0) = - \int x^2K(x)dx = -\mu_2(K)$, o que contradiz o facto de $\mu_2(K) \neq 0$. ■

Lema 2.2. *Se $K \in L_\infty$, φ_K não é identicamente igual a 1 numa vizinhança da origem e $f \in L_2$, temos que*

$$B_f(h) > 0, \forall h > 0.$$

Demonstração. Tendo em conta as propriedades das transformadas de Fourier (ver Butzer and Nessel, 1971, p. 168), para $h > 0$ temos

$$\begin{aligned} B_f(h) &= \int [E(f_{n,h}(x)) - f(x)]^2 dx = \int [K_h * f(x) - f(x)]^2 dx \\ &= \frac{1}{2\pi} \int |\varphi_{K_h * f}(t) - \varphi_f(t)|^2 dt = \frac{1}{2\pi} \int |\varphi_{K_h}(t) \varphi_f(t) - \varphi_f(t)|^2 dt \\ &= \frac{1}{2\pi} \int |\varphi_{K_h}(t) - 1|^2 |\varphi_f(t)|^2 dt. \end{aligned}$$

Se $B_f(h) = 0$, para algum $h > 0$, então $|\varphi_{K_h}(t) - 1|^2 |\varphi_f(t)|^2 = 0$, para todo o $t \in \mathbb{R}$. Em particular, como $|\varphi_f(t)| > 0$ numa vizinhança da origem teríamos $\varphi_{K_h}(t) = 1$ nessa vizinhança, o que contradiz a hipótese. Assim, $B_f(h) > 0$. ■

Teorema 2.5. *Sejam $K \in L_\infty$ e contínuo em zero, com $R(K) < 2K(0)$, a transformada de Fourier de K não identicamente igual a 1 numa vizinhança da origem e $f \in L_2$. Então*

$$\lim_{n \rightarrow \infty} h_{0,n}(f) = 0.$$

Demonstração. Usando o Teorema 2.2 e (2.9) temos

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} M_{f,n}(h_{0,n}(f)) \\ &= \lim_{n \rightarrow \infty} (V_{f,n}(h_{0,n}(f)) + B_f(h_{0,n}(f))) \\ &= \lim_{n \rightarrow \infty} \left(\frac{R(K)}{nh_{0,n}} - \frac{R_{K * \bar{K}, h_{0,n}}(f)}{n} + B_f(h_{0,n}(f)) \right) \\ &= \lim_{n \rightarrow \infty} B_f(h_{0,n}(f)). \end{aligned}$$

Suponhamos que $\lim_{n \rightarrow \infty} h_{0,n}(f) \neq 0$. Nestas condições, existe uma subsequência $(h_{0,n_k}(f))$ de $h_{0,n}(f)$ tal que $\lim_{n_k \rightarrow \infty} h_{0,n_k}(f) = \lambda$, onde $\lambda \in]0, \infty[$.

Pelo Lema 2.1,

$$\lim_{k \rightarrow \infty} B_f(h_{0,n_k}) = \begin{cases} R(f) & \lambda = \infty \\ B_f(\lambda) & \lambda < \infty. \end{cases}$$

Sabemos que $R(f) > 0$ e, pelo Lema 2.2, $B_f(\lambda) > 0$ o que contradiz $\lim_{n \rightarrow \infty} B_f(h_{0,n}(f)) = 0$. ■

Capítulo 3

Expressões assintóticas para a janela ótima

Neste capítulo, apresentaremos desenvolvimentos assintóticos para $M_{f,n}(h)$ e suas derivadas de primeira e segunda ordem, de modo a que consigamos obter expressões assintóticas para a janela ótima, $h_{0,n}$, cuja existência estabelecemos no capítulo anterior. No final deste capítulo faremos um estudo comparativo das expansões obtidas para $h_{0,n}(f) = h_{0,n}$ para diferentes funções densidades f .

3.1 Expressões assintóticas para o MISE

Para obtermos os desenvolvimentos assintóticos para o MISE, comecemos por reescrever $V_{f,n}(h)$ e $B_f(h)$, definidas por (2.4) e (2.5), na forma

$$\begin{aligned} V_{f,n}(h) &= \frac{R(K)}{nh} - \frac{1}{n} R_{K*\bar{K},h}(f) \\ &= \frac{R(K)}{nh} - \frac{1}{n} \int K*\bar{K}(u) f*\bar{f}(hu) du \\ &= \frac{R(K)}{nh} - \frac{1}{n} \iint K(u-v)\bar{K}(v) f*\bar{f}(hu) dudv \\ &= \frac{R(K)}{nh} - \frac{1}{n} \iint K(u)K(v) f*\bar{f}(h(u-v)) dudv \end{aligned} \quad (3.1)$$

e

$$\begin{aligned} B_f(h) &= \int (E(f_{n,h}(x)) - f(x))^2 dx \\ &= \int (K_h*f(x) - f(x))^2 dx \\ &= \int \left(\int K_h(x-y) f(y) dy - f(x) \right)^2 dx \\ &= \int \left(\int K(u) (f(x-uh) - f(x)) du \right)^2 dx. \end{aligned} \quad (3.2)$$

Teorema 3.1. *Sejam $K \in L_2$ um núcleo de 2.^a ordem e f com derivadas de 2.^a ordem contínuas em \mathbb{R} com $f, f'' \in L_2$. Se $h \rightarrow 0$ temos*

$$M_{f,n}(h) = \frac{R(K)}{nh} + \frac{h^4}{4} \mu_2(K)^2 R(f'') + O(n^{-1}) + o(h^4).$$

Demonstração. Tendo em conta (3.1) e a desigualdade

$$\left| \iint K(u)K(v)f * \bar{f}(h(u-v))dudv \right| \leq R(f) \left(\int |K(u)|du \right)^2,$$

podemos concluir que

$$V_{f,n}(h) = \frac{R(K)}{nh} + O(n^{-1}).$$

Para o viés, da expressão (3.2) e do desenvolvimento de Taylor de segunda ordem com resto integral temos

$$\begin{aligned} B_f(h) &= \int \left(\int K(u)(-uhf'(x) + h^2 \int u^2(1-t)f''(x-thu)dt)du \right)^2 dx \\ &= \int \left(-hf'(x)\mu_1(K) + h^2 \iint_0^1 (1-t)u^2K(u)f''(x-thu)dtdu \right)^2 dx. \end{aligned}$$

Como $\mu_1(K) = 0$ e f'' é contínua concluímos que

$$B_f(h) = h^4 \iiint \int_0^1 \int_0^1 (1-t)(1-s)u^2v^2K(u)K(v)f''(x-thu)f''(x-shv)dtdu dsdv dx.$$

Efetuada a mudança de variável $x = shv - y$ e sabendo que $f'' * f''$ é contínua e limitada, a aplicação do TCD conduz a

$$\begin{aligned} B_f(h) &= h^4 \iiint \int_0^1 \int_0^1 (1-t)(1-s)u^2v^2K(u)K(v)f'' * \bar{f''}(h(sv-tu))dtdu dsdv \\ &= h^4 A(h), \end{aligned}$$

onde

$$\begin{aligned} A(h) &= \iiint \int_0^1 \int_0^1 (1-t)(1-s)u^2v^2K(u)K(v)f'' * \bar{f''}(h(sv-tu))dtdu dsdv \quad (3.3) \\ &= \mu_2^2(K)R(f'') \int_0^1 \int_0^1 (1-t)(1-s)dtdu dsdv + o(1) \\ &= \frac{1}{4} \mu_2^2(K)R(f'') + o(1). \end{aligned}$$

Atendendo que (2.1) a demonstração fica concluída. ■

A ordem de convergência para zero de $M_{f,n}(h)$, obtida do teorema anterior, pode ser melhorada se condições adicionais forem impostas ao núcleo e a densidade f .

Teorema 3.2. *Sejam $K \in L_2$ um núcleo de 2.^a ordem tal que $\int u^4 |k(u)| du < \infty$ e f com derivadas contínuas, limitadas e integráveis até à 3.^a ordem. Se $h \rightarrow 0$ temos*

$$M_{f,n}(h) = \frac{R(K)}{nh} + \frac{h^4}{4} \mu_2(K)^2 R(f'') + O(n^{-1}) + O(h^6) \quad (3.4)$$

e

$$M'_{f,n}(h) = -\frac{R(K)}{nh^2} + h^3 \mu_2(K)^2 R(f'') + O(n^{-1}h) + O(h^5). \quad (3.5)$$

Demonstração. Atendendo ao Teorema 3.1, basta provar que o termo residual de $A(h)$ em (3.3) é de ordem $O(h^2)$.

Por hipótese, f admite derivadas até à 3.^a ordem. Usando a Proposição A.2, com $g = f''$, temos $f''' * \bar{f}''(0) = 0$. Assim,

$$\begin{aligned} f'' * \bar{f}''(h(sv - tu)) &= f'' * \bar{f}''(0) + (sv - tu) h f''' * \bar{f}''(0) \\ &\quad - (sv - tu)^2 h^2 \int_0^1 (1 - t') f''' * \bar{f}'''(t'(sv - tu)h) dt' \\ &= R(f'') - (sv - tu)^2 h^2 \int_0^1 (1 - t') f''' * \bar{f}'''(t'(sv - tu)h) dt', \end{aligned}$$

de onde deduzimos que $A(h) = \frac{1}{4} \mu_2(K)^2 R(f'') + O(h^2)$.

Para demonstrar (3.5) procedemos ao cálculo de $V'_{f,n}(h)$. Para isso vamos derivar a expressão (3.1):

$$V'_{f,n}(h) = -\frac{R(K)}{nh^2} - \frac{1}{n} \iint K(u)K(v)(u - v) f' * \bar{f}(h(u - v)) dudv.$$

Da fórmula de Taylor,

$$f' * \bar{f}(h(u - v)) = f' * \bar{f}(0) - h(u - v) \int_0^1 f' * \bar{f}'(h(u - v)t) dt$$

e pela Proposição A.2, com $g = f$, temos $f' * \bar{f}(0) = 0$.

Assim,

$$\begin{aligned} V'_{f,n}(h) &= -\frac{R(K)}{nh^2} - \frac{h}{n} \iiint_0^1 K(u)K(v)(u - v)^2 f' * \bar{f}'(h(u - v)t) dt dudv \\ &= -\frac{R(K)}{nh^2} + O(n^{-1}h). \end{aligned}$$

A derivada de $B_f(h)$ é dada por

$$B'_f(h) = 4h^3 A(h) + h^4 A'(h). \quad (3.6)$$

Atendendo a que já temos a expressão de $A(h)$ falta apenas calcular a respetiva derivada, $A'(h)$,

$$A'(h) = \iiint_0^1 \int_0^1 (1 - t)(1 - s) u^2 v^2 K(u)K(v) f''' * \bar{f}''((sv - tu)h) dt duds dv. \quad (3.7)$$

Da fórmula de Taylor,

$$f''' * \bar{f}''((sv - tu)h) = f''' * \bar{f}''(0) - (sv - tu)h \int_0^1 f''' * \bar{f}'''(t'(sv - tu)h) dt'$$

onde, pela Proposição A.2 com $g = f''$, temos $f''' * \bar{f}''(0) = 0$. Então,

$$\begin{aligned} A'(h) &= \iiint_0^1 \int_0^1 (1-t)(1-s)u^2v^2K(u)K(v) \left(-(sv - tu)h \int_0^1 f''' * \bar{f}'''(t'(sv - tu)h) dt' \right) dt ds dudv \\ &= O(h). \end{aligned}$$

Então, por (3.6),

$$B'_f(h) = \frac{4h^3}{4} \mu_2(K)^2 R(f'') + O(h^5) + O(h^5) = h^3 \mu_2(K)^2 R(f'') + O(h^5).$$

Logo, concluímos a demonstração. ■

Teorema 3.3. *Sejam $K \in L_2$ um núcleo de 2.^a ordem tal que $\int u^6 |k(u)| du < \infty$ e f com derivadas contínuas, limitadas e integráveis até à 4.^a ordem. Se $h \rightarrow 0$ temos*

$$M'_{f,n}(h) = -\frac{R(K)}{nh^2} + h^3 \mu_2(K)^2 R(f'') - \frac{h^5}{12} (3\mu_2(K)\mu_4(K) - 2\mu_3(K)^2) R(f''') + O(n^{-1}h) + O(h^7).$$

Demonstração. Do teorema anterior, sabemos que $V'_{f,n}(h) = -\frac{R(K)}{nh^2} + O(n^{-1}h)$. Dado que f admite derivadas até à ordem 4, precisamos de recalculer $B'_f(h)$ e, por consequência, também $A(h)$ e $A'(h)$.

Aplicando a fórmula de Taylor, a Proposição A.2 e a definição de $R(f)$,

$$\begin{aligned} f'' * \bar{f}''(h(sv - tu)) &= R(f'') - \frac{h^2}{2} (sv - tu)^2 R(f''') \\ &\quad + \frac{h^4}{6} (sv - tu)^4 \int_0^1 (1-t')^2 f''' * \bar{f}'''(t'(sv - tu)h) dt' \end{aligned}$$

De (3.3) e da expressão anterior,

$$\begin{aligned} A(h) &= \iiint_0^1 \int_0^1 (1-t)(1-s)u^2v^2K(u)K(v) \left[R(f'') - \frac{h^2}{2} (sv - tu)^2 R(f''') \right] dt ds dudv + O(h^4) \\ &= \frac{1}{4} R(f'') \mu_2(K)^2 - \frac{h^2}{48} \mu_2(K)\mu_4(K) R(f''') + \frac{h^2}{36} \mu_3(K)^2 R(f''') - \frac{h^2}{48} \mu_2(K)\mu_4(K) R(f''') + O(h^4) \\ &= \frac{1}{4} R(f'') \mu_2(K)^2 - \frac{h^2}{72} (3\mu_2(K)\mu_4(K) + 2\mu_3(K)^2) R(f''') + O(h^4), \end{aligned}$$

e, por um raciocínio análogo, de (3.7) temos

$$A'(h) = \iiint_0^1 \int_0^1 (1-t)(1-s)u^2v^2K(u)K(v)(sv - tu) f''' * \bar{f}''(h(sv - tu)) dt ds dudv$$

$$\begin{aligned}
&= - \iiint\limits_0^1 \int_0^1 (1-t)(1-s)u^2v^2K(u)K(v)(sv-tu)^2hR(f''') + O(h^3) \\
&= -\frac{h}{24}R(f''')\mu_2(K)\mu_4(K) + \frac{2h}{36}R(f''')\mu_3(K)^2 - \frac{h}{24}R(f''')\mu_2(K)\mu_4(K) + O(h^3) \\
&= -\frac{h}{36}\left(3\mu_2(K)\mu_4(K) - 2\mu_3(K)^2\right)R(f''') + O(h^3).
\end{aligned}$$

Então,

$$\begin{aligned}
B'_f(h) &= 4h^3A(h) + h^4A'(h) \\
&= h^3\mu_2(K)^2R(f'') - \frac{h^5}{12}\left(3\mu_2(K)\mu_4(K) + 2\mu_3(K)^2\right)R(f''') + O(h^7).
\end{aligned}$$

Logo, obtemos o pretendido. ■

Teorema 3.4. *Sejam $K \in L_2$ um núcleo de 2.^a ordem tal que $\int u^4|k(u)|du < \infty$ e f com derivadas contínuas, limitadas e integráveis até à 3.^a ordem. Se $h \rightarrow 0$ temos*

$$M''_{f,n}(h) = \frac{2R(K)}{nh^3} + 3h^2\mu_2(K)^2R(f'') + O(n^{-1}) + O(h^4).$$

Demonstração. Derivando a expressão de $V'_{f,n}(h)$ obtida para (3.5) obtemos,

$$V''_{f,n}(h) = \frac{R(K)}{nh^3} + \frac{1}{n} \iint K(u)K(v)(u-v)^2 f' * \bar{f}'(h(u-v)) dudv = \frac{R(K)}{nh^3} + O(n^{-1}).$$

Para a segunda derivada do viés, derivamos a expressão (3.6) e, assim, obtemos

$$B''_f(h) = 12h^2A(h) + 8h^3A'(h) + 4h^3A'(h) + h^4A(h).$$

Do Teorema 3.2, sabemos que $A'(h) = O(h)$ e que $A(h) = \frac{1}{4}\mu_2(K)^2R(f'') + O(h^2)$, então falta calcular $A''(h)$. Usando (3.7) e o Lema A.2, obtemos

$$\begin{aligned}
A''(h) &= - \iiint\limits_0^1 \int_0^1 (1-t)(1-s)u^2v^2K(u)K(v) \left(-(sv-tu)h \int_0^1 f''' * \bar{f}'''(t'(sv-tu)h) dt' \right) dt dudsdv \\
&= O(1).
\end{aligned}$$

Logo,

$$B''_f(h) = 3h^2\mu_2(K)^2R(f'') + O(h^4),$$

o que permite concluir a demonstração. ■

3.2 Desenvolvimento assintótico de $h_{0,n}$

Nesta secção iremos obter aproximações assintóticas para $h_{0,n}$. Para isso, começemos por estabelecer a ordem de convergência da janela $h_{0,n}$. Na demonstração deste resultado seguimos [Chacón \(2004, p. 41-46\)](#).

Teorema 3.5. *Sejam $K \in L_2$ um núcleo simétrico de ordem 2 tal que $\int u^4 |K(u)| du < \infty$ e f com derivadas contínuas, limitadas e integráveis até à ordem 4. Então,*

$$0 < \liminf n^{\frac{1}{5}} h_{0,n} \leq \limsup n^{\frac{1}{5}} h_{0,n} < \infty$$

e

$$\lim_{n \rightarrow \infty} n^{\frac{1}{5}} h_{0,n} = R(K)^{\frac{1}{5}} \mu_2^2(K)^{-\frac{1}{5}} R(f'')^{-\frac{1}{5}}. \quad (3.8)$$

Demonstração. Do Teorema 3.2, para $C > 0$,

$$M_{f,n}(Cn^{-\frac{1}{5}}) = \left[C^{-1}R(K) + \frac{C^4}{4} \mu_2^2(K)R(f'') \right] n^{-\frac{4}{5}} + O(n^{-1}).$$

Então,

$$\lim_{n \rightarrow \infty} n^{\frac{4}{5}} M_{f,n}(Cn^{-\frac{1}{5}}) = \frac{C^4}{4} \mu_2^2(K)R(f'') + C^{-1}R(K).$$

O valor de C que torna o limite mínimo é $C_0 = R(K)^{\frac{1}{5}} \mu_2^2(K)^{-\frac{1}{5}} R(f'')^{-\frac{1}{5}}$. Mais, pela definição de $h_{0,n}$, temos

$$n^{\frac{4}{5}} M_{f,n}(C_0 n^{-\frac{1}{5}}) \geq n^{\frac{4}{5}} M_{f,n}(h_{0,n}), \quad \forall n \in \mathbb{N}.$$

Tomando o limite vem

$$\frac{C_0^4}{4} \mu_2^2(K)R(f'') + C_0^{-1}R(K) \geq \limsup n^{\frac{4}{5}} M_{f,n}(h_{0,n}). \quad (3.9)$$

Nestas condições, $\limsup n^{\frac{4}{5}} M_{f,n}(h_{0,n}) < \infty$ e, conseqüentemente, $\limsup n^{\frac{1}{5}} h_{0,n}(f) < \infty$.

Caso isto não se verificasse, por $h_{0,n} \rightarrow 0$ e pelo Teorema 3.1, teríamos

$$\begin{aligned} \limsup n^{\frac{4}{5}} M_{f,n}(h_{0,n}) &\geq \limsup n^{\frac{4}{5}} B_f(h_{0,n}) \\ &= \limsup n^{\frac{4}{5}} \left[\frac{h_{0,n}^4}{4} \mu_2^2(K)R(f'') + o(h_{0,n}^4) \right] \\ &\geq \limsup \frac{(n^{\frac{1}{5}} h_{0,n})^4}{4} \mu_2^2(K)R(f'') = \infty, \end{aligned}$$

o que contradiz (3.9).

Da expressão assintótica para a variância, conseguimos provar que

$$\liminf n^{\frac{1}{5}} h_{0,n} > 0,$$

dado que

$$\begin{aligned} n^{\frac{4}{5}} V_{f,n}(h_{0,n}) &= \frac{n^{\frac{4}{5}}}{nh_{0,n}} R(K) + O\left(n^{-1} n^{\frac{4}{5}}\right) \\ &= \frac{1}{n^{\frac{1}{5}} h_{0,n}} R(K) + O\left(n^{-\frac{1}{5}}\right). \end{aligned}$$

Se $\liminf n^{\frac{1}{5}}h_{0,n} = 0$ então $\limsup n^{\frac{4}{5}}V_{f,n}(h_{0,n}) = \infty$ e por maioria de razão $\limsup n^{\frac{4}{5}}M_{f,n}(h_{0,n}) = \infty$, o que contradiz (3.9).

Assim,

$$0 < \liminf n^{\frac{1}{5}}h_{0,n} \leq \limsup n^{\frac{1}{5}}h_{0,n} < \infty.$$

Aplicando, de novo, o Teorema 3.2 para $h_{0,n}$ temos

$$n^{\frac{4}{5}}M_{f,n}(h_{0,n}) = \frac{(n^{\frac{1}{5}}h_{0,n})^4}{4}\mu_2^2(K)R(f'') + \frac{R(K)}{n^{\frac{1}{5}}h_{0,n}} + o(1).$$

Resta-nos provar que a sucessão $(n^{\frac{1}{5}}h_{0,n})$ converge para C_0 . Uma vez que a sucessão é limitada, basta provar que C_0 é o único valor de aderência de $(n^{\frac{1}{5}}h_{0,n})$, (ver Lima, 1995, p. 93–98). Seja então l o limite de alguma subsucessão $(n_k^{\frac{1}{5}}h_{0,n_k}(f))$ de $(n^{\frac{1}{5}}h_{0,n})$. Usando o Teorema 3.2 concluímos que

$$\lim_{k \rightarrow \infty} n_k^{\frac{4}{5}}M_{f,n_k}(h_{0,n_k}) = \frac{l^4}{4}\mu_2^2(K)R(f'') + \frac{R(K)}{l}.$$

De (3.9) deduzimos que

$$\frac{l^4}{4}\mu_2^2(K)R(f'') + l^{-1}R(K) \leq \frac{C_0^4}{4}R(f'')\mu_2^2(K) + C_0^{-1}R(K).$$

de onde concluímos que $l = C_0$. ■

Do Teorema 3.5 concluímos que $h_{0,n}(f) = O(n^{-\frac{1}{5}})$. Estamos então em condições de apresentar as expressões assintóticas para $h_{0,n}$.

Teorema 3.6. *Sejam $K \in L_2$ um núcleo de 2.^a ordem tal que $\int u^4|k(u)|du < \infty$ e f com derivadas contínuas, limitadas e integráveis até à ordem 3. Então,*

$$h_{0,n} = h_{1,n} + O(n^{-\frac{3}{5}})$$

em que

$$h_{1,n} = R(K)^{\frac{1}{5}}\mu_2(K)^{-\frac{2}{5}}R(f'')^{-\frac{1}{5}}n^{-\frac{1}{5}}. \quad (3.10)$$

Demonstração. Atendendo a que $M'_{f,n}(h_{0,n}) = 0$, pelo desenvolvimento de Taylor de 1.^a ordem numa vizinhança de $h_{1,n}$ podemos escrever

$$M'_{f,n}(h_{1,n}) + M''_{f,n}(\tilde{h})(h_{0,n} - h_{1,n}) = 0$$

e ainda

$$h_{0,n} = h_{1,n} - M'_{f,n}(h_{1,n})M''_{f,n}(\tilde{h})^{-1}, \quad (3.11)$$

com \tilde{h} entre $h_{0,n}$ e $h_{1,n}$.

Atendendo a que $\tilde{h} = O(n^{-\frac{1}{5}})$, pelo Teorema 3.4 temos

$$\begin{aligned} M'_{f,n}(\tilde{h}) &= \frac{R(K)}{n\tilde{h}^3} + 3\tilde{h}^2\mu_2(K)^2R(f'') + O(n^{-1}) + O(\tilde{h}^4) \\ &= n^{-1}O(n^{-\frac{3}{5}}) + O(n^{-\frac{2}{5}}) + O(n^{-1}) + O(n^{-\frac{4}{5}}) \\ &= O(n^{-\frac{2}{5}}). \end{aligned} \quad (3.12)$$

Usando (3.11), basta provar que $M'_{f,n}(h_{1,n}) = O(n^{-1})$. Do Teorema 3.2 temos

$$M'_{f,n}(h_{1,n}) = -\frac{R(K)}{nh_{1,n}^2} + h_{1,n}^3\mu_2(K)^2R(f'') + O(n^{-1}h_{1,n}) + O(h_{1,n}^5) = O(n^{-1}).$$

■

Teorema 3.7. *Sejam $K \in L_2$ um núcleo de 2.^a ordem tal que $\int u^6|K(u)|du < \infty$ e f com derivadas contínuas, limitadas e integráveis até à ordem 4. Então*

$$h_{0,n} = h_{2,n} + O(n^{-1})$$

onde

$$h_{2,n} = h_{1,n} + \frac{1}{60}R(K)^{\frac{3}{5}}\mu_2(K)^{-\frac{16}{5}}(3\mu_2(K)\mu_4(K) - 2\mu_3(K)^2)R(f'')^{-\frac{8}{5}}R(f''')n^{-\frac{3}{5}}.$$

Se K é simétrico, isto é, $K(u) = K(-u)$, para todo o $x \in \mathbb{R}$, então

$$h_{2,n} = h_{1,n} + \frac{1}{20}R(K)^{\frac{3}{5}}\mu_2(K)^{-\frac{22}{5}}\mu_4(K)R(f'')^{-\frac{8}{5}}R(f''')n^{-\frac{3}{5}}. \quad (3.13)$$

Demonstração. Atendendo a que a igualdade (3.11) é válida com $h_{2,n}$ no lugar de $h_{1,n}$, em que \tilde{h} , entre $h_{0,n}$ e $h_{2,n}$, satisfaz (3.12), é suficiente provar que $M'_{f,n}(h_{2,n}) = O(n^{-\frac{6}{5}})$.

Do Teorema 3.3 temos

$$\begin{aligned} h_{2,n}^2 M'_{f,n}(h_{2,n}) &= -\frac{R(K)}{n} + Ah_{2,n}^5 - Bh_{2,n}^7 + O(n^{-1}h_{2,n}^3) + O(h_{2,n}^9) \\ &= -\frac{R(K)}{n} + Ah_{2,n}^5 - Bh_{2,n}^7 + O(n^{-1}n^{-\frac{3}{5}}) + O(n^{-\frac{9}{5}}), \end{aligned}$$

com $A = \mu_2(K)^2R(f'')$ e $B = \frac{1}{12}(3\mu_2(K)\mu_4(K) - 2\mu_3(K)^2)R(f''')$.

Tendo em conta que $h_{2,n} = h_{1,n} + cn^{-\frac{3}{5}}$, em que

$$c = \frac{1}{60}R(K)^{\frac{3}{5}}\mu_2(K)^{-\frac{16}{5}}(3\mu_2(K)\mu_4(K) - 2\mu_3(K)^2)R(f'')^{-\frac{8}{5}}R(f'''),$$

temos

$$\begin{aligned} h_{2,n}^2 M'_{f,n}(h_{2,n}) &= -\frac{R(K)}{n} + (h_{1,n} + cn^{-\frac{3}{5}})^5A - (h_{1,n} + cn^{-\frac{3}{5}})^7B + O(n^{-\frac{9}{5}}) \\ &= -\frac{R(K)}{n} + Ah_{1,n}^5 + 5Ah_{1,n}^4cn^{-\frac{3}{5}} - Bh_{1,n}^7 + O(n^{-\frac{9}{5}}). \end{aligned}$$

Atendendo a que $-\frac{R(K)}{n} + Ah_{1,n}^5 = 0$ e $5Ah_{1,n}^4 cn^{-\frac{3}{5}} - Bh_{1,n}^7 = 0$, temos

$$h_{2,n}^2 M'_{f,n}(h_{2,n}) = O(n^{-\frac{9}{5}}),$$

o que conclui a demonstração. ■

3.3 Comparação de $h_{1,n}$ e $h_{2,n}$

Dos resultados apresentados no parágrafo anterior sabemos que, sob certas condições em K e f , se tem

$$\frac{h_{1,n}}{h_{0,n}} = 1 + O(n^{-\frac{2}{5}}) \quad e \quad \frac{h_{2,n}}{h_{0,n}} = 1 + O(n^{-\frac{4}{5}}).$$

Nesta secção analisamos o comportamento de $h_{1,n}/h_{0,n}$ e $h_{2,n}/h_{0,n}$ para diversos valores de n . Com base nas expressões obtidas em (3.10) e (3.13), precisamos de primeiramente definir K e f . Assim, consideramos para K a densidade normal standard, $K(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$, e, para f iremos tomar oito das quinze misturas de densidades normais consideradas em Marron e Wand (1992), que se apresentam na Figura 3.1. Uma vez que f é uma mistura de normais, o cálculo de $h_{0,n}$ é efetuado com base no Teorema 5.1 de Marron e Wand (1992, p. 728).

A densidade #1 representa um papel importante na estimação de densidades uma vez que é a função com as características distribucionais simples. A densidade #2 não difere muito no seu aspecto da #1 mas foi escolhida pois assemelha-se a densidades de valor extremos. A densidade #3 é bastante diferente da normal e foi escolhida pois assemelha-se a lognormal. Estas três misturas são densidades unimodais. As densidades #6, #7 e #8 são densidades bimodais e a #9 é uma densidade trimodal. Estas densidades apresentam desvios simples mas importantes relativamente a unimodalidade. A densidade #12 é multimodal apresentando características distribucionais mais complexas.

Nestas condições, procedemos a execução dos gráficos de $h_{1,n}/h_{0,n}$ e $h_{2,n}/h_{0,n}$, presentes na Figura 3.2, em função do tamanho da amostra n , que representaremos por $\log_{10}(n)$.

Conseguimos, então, verificar que para todas as distribuições e para todos os valores de n , $h_{0,n}$ é melhor aproximado por $h_{2,n}$ do que por $h_{1,n}$ e que para as densidades com características distribucionais mais complexas, $h_{1,n}$ revela-se uma fraca aproximação para $h_{0,n}$, mesmo quando o tamanho da amostra é grande.

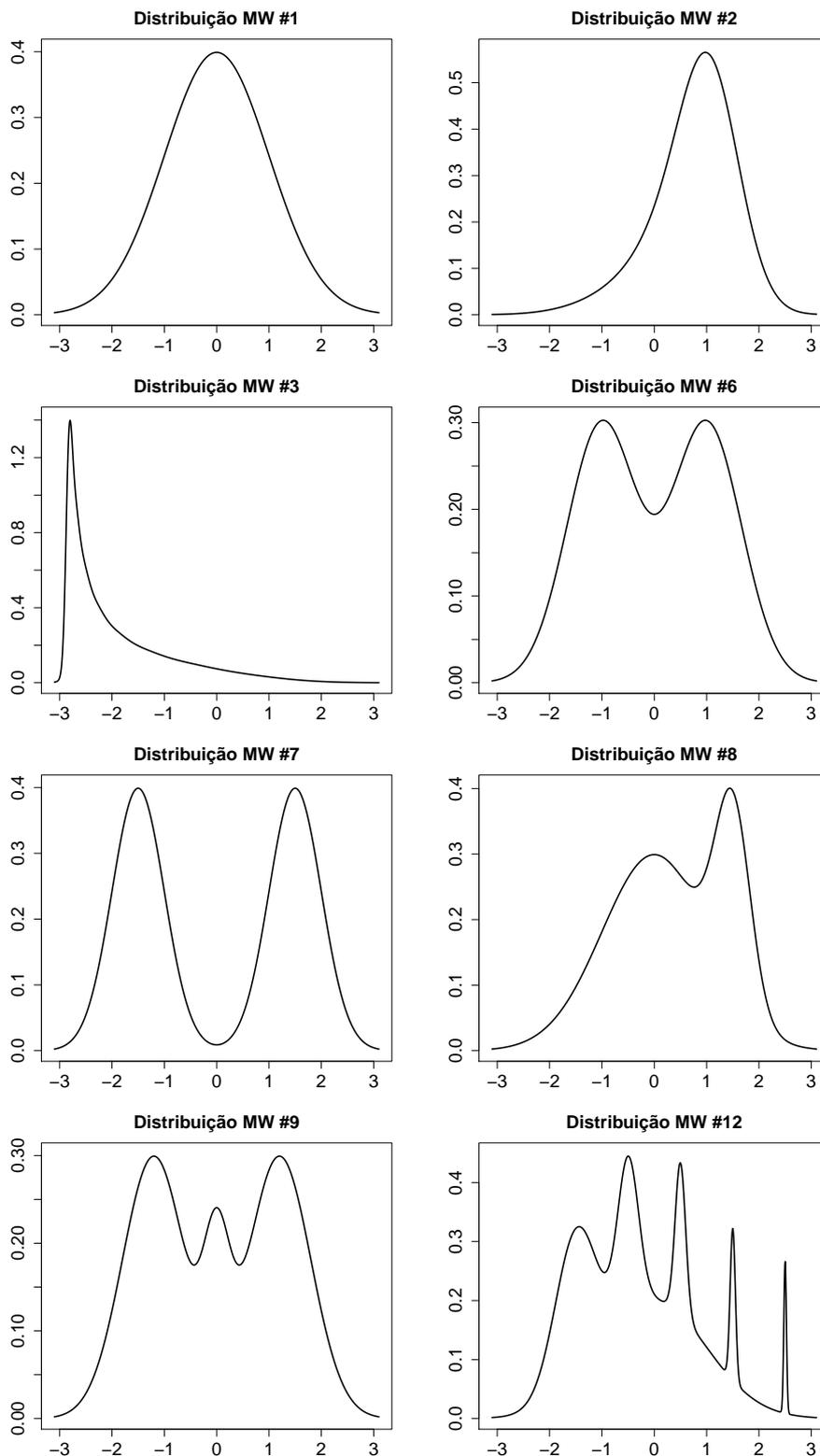


Fig. 3.1 Mistura de densidades normais #1, #2, #3, #6, #7, #8, #9 e #12 de (Marron e Wand, 1992)

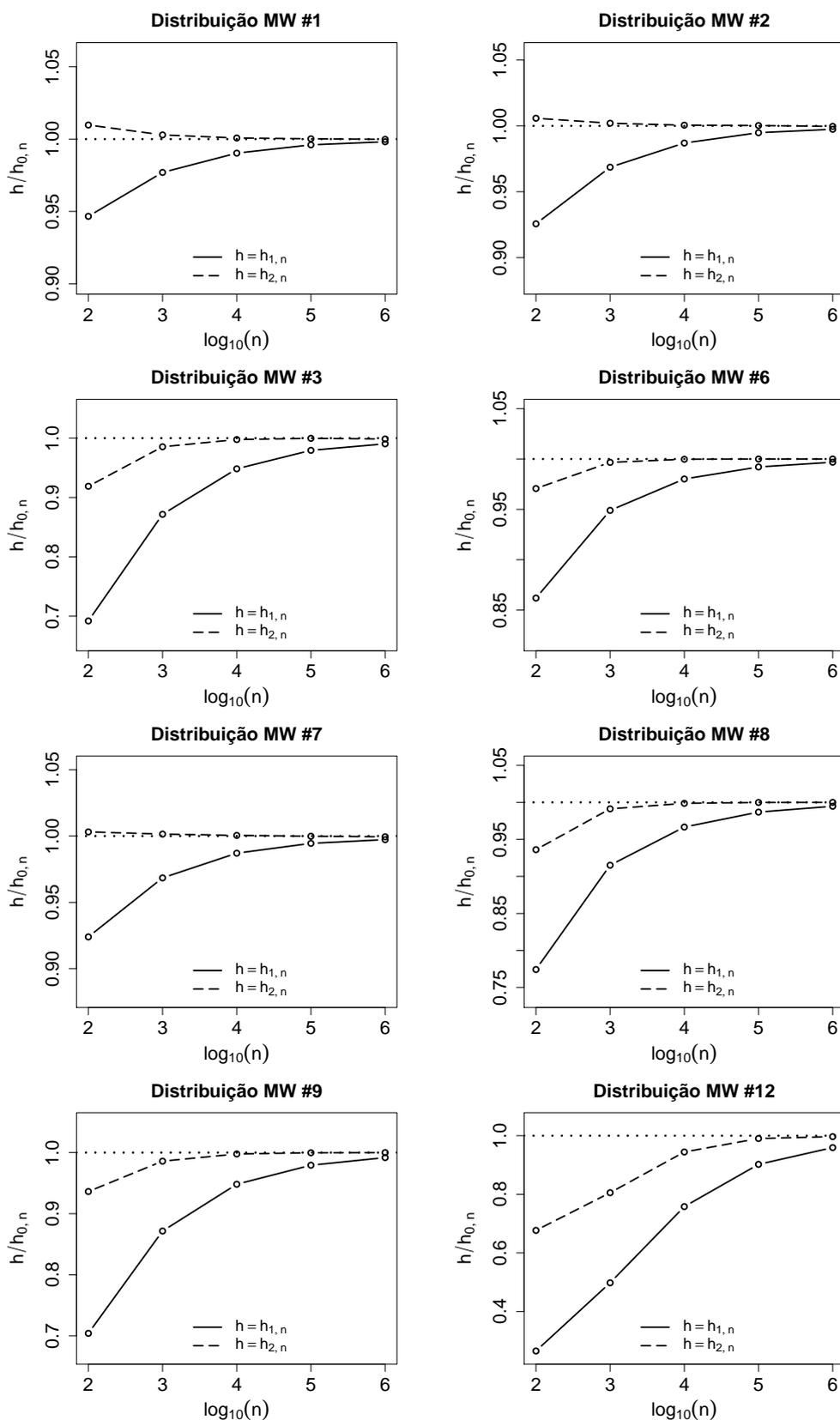


Fig. 3.2 Comparação de $h_{1,n}$ e $h_{2,n}$ para as misturas de densidades normais #1, #2, #3, #6, #7, #8, #9 e #12

Capítulo 4

Seleção da janela usando funções de Hermite

Vimos no capítulo anterior que, sob certas condições em K e f , a janela ótima $h_{0,n}$ podia ser aproximada por $h_{1,n}$ e $h_{2,n}$ dadas por

$$h_{1,n} = C_{1,K} R(f'')^{-\frac{1}{5}} n^{-\frac{1}{5}}$$

e

$$h_{2,n} = h_{1,n} + C_{2,K} R(f'')^{-\frac{8}{5}} R(f''') n^{-\frac{3}{5}}$$

com $C_{1,K} = R(K)^{\frac{1}{5}} \mu_2(K)^{-\frac{2}{5}}$ e $C_{2,K} = \frac{1}{20} R(K)^{\frac{3}{5}} \mu_2(K)^{-\frac{22}{5}} \mu_4(K)$.

As janelas $h_{1,n}$ e $h_{2,n}$ dependem dos funcionais quadráticos $R(f'')$ e $R(f''')$, funcionais estas que desconhecemos visto que desconhecemos f . Para que $h_{1,n}$ e $h_{2,n}$ possam ser usadas na prática para a escolha automática da janela do estimador do núcleo, é necessário estimar $R(f'')$ e $R(f''')$ com base na amostra observada X_1, X_2, \dots, X_n . Neste capítulo vamos apresentar estimadores da funcional $R(f^{(p)})$, para $p \in \mathbb{N}_0$, baseados nas funções de Hermite, e estudamos a convergência de tais estimadores.

4.1 Estimadores de $R(f^{(p)})$ baseados nas funções de Hermite

Seja o sistema ortonormado de Hermite, $\{h_k, k = 0, 1, \dots\}$, dado por

$$h_k(x) = \left(2^k k! \pi^{\frac{1}{2}}\right)^{-\frac{1}{2}} H_k(x) e^{-\frac{x^2}{2}}, \quad (4.1)$$

onde $H_k(x)$ é polinômio de Hermite de grau k dado por

$$H_k(x) = (-1)^k e^{x^2} \frac{d^k}{dx^k} (e^{-x^2}).$$

Sendo $\{h_k, k = 0, 1, \dots\}$ uma base de L_2 , se admitirmos que $f^{(p)}$, para $p \in \mathbb{N}_0$, é de quadrado integrável, onde f é uma densidade comum as variáveis X_1, X_2, \dots, X_n , sabemos que f admite a

representação seguinte em L_2 ,

$$f^{(p)}(x) = \sum_{k=0}^{\infty} a_{p,k} h_k(x), \quad (4.2)$$

em que $a_{p,k}$ são os coeficientes de Hermite de $f^{(p)}$ dados por

$$a_{p,k} = \int f^{(p)}(x) h_k(x) dx.$$

Estes coeficientes podem ser reescritos usando a Proposição A.3 (com $r = s = p$):

Proposição 4.1. *Se $f^{(p)}, f^{(p-1)}, \dots, f$ são limitadas de quadrado integrável então*

$$a_{p,k} = (-1)^p \int h_k^{(p)}(x) f(x) dx = (-1)^p E(h_k^{(p)}(X_1)),$$

com $k = 0, 1, \dots$

Atendendo ao resultado anterior, um estimador natural para $f^{(p)}$ é dado por

$$\hat{f}_{p,m}(x) = \sum_{k=0}^m \hat{a}_{p,k} h_k(x), \quad (4.3)$$

onde $m = m(n)$ depende n e $\hat{a}_{p,k}$ é o estimador cêntrico de $a_{p,k}$ dado por

$$\hat{a}_{p,k} = \frac{(-1)^p}{n} \sum_{i=1}^n h_k^{(p)}(X_i). \quad (4.4)$$

Por outro lado, uma vez que

$$R(f^{(p)}) = \int f^{(p)}(x)^2 dx = \int \left(\sum_{k=0}^{\infty} a_{p,k} h_k(x) \right)^2 dx = \sum_{k=0}^{\infty} a_{p,k}^2, \quad (4.5)$$

um estimador natural para $R(f^{(p)})$ é dado por

$$\hat{R}_{1,p,m} = \sum_{k=0}^m \hat{a}_{p,k}^2. \quad (4.6)$$

Neste estimador, como dito anteriormente, $\hat{a}_{p,k}$ é um estimador cêntrico de $a_{p,k}$. No entanto, $\hat{a}_{p,k}^2$ não é estimador cêntrico de $a_{p,k}^2$. Para obter um estimador cêntrico de $a_{p,k}^2$ comecemos por notar que

$$a_{p,k}^2 = \iint h_k^{(p)}(x) h_k^{(p)}(y) f(x) f(y) dx dy = E[h_k^{(p)}(X_1) h_k^{(p)}(X_2)]$$

e, portanto, um estimador cêntrico de $a_{p,k}^2$ é dado por

$$\hat{a}_{p,k}^2 = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h_k^{(p)}(X_i) h_k^{(p)}(X_j). \quad (4.7)$$

Assim, um estimador alternativo a $\widehat{R}_{1,p,m}$ será

$$\widehat{R}_{2,p,m} = \sum_{k=0}^m \widehat{a}_{p,k}^2. \quad (4.8)$$

Uma vez que

$$\widehat{a}_{p,k}^2 = \frac{1}{n(n-1)} \left[\sum_{i,j=1}^n h_k^{(p)}(X_i) h_k^{(p)}(X_j) - \sum_{i=1}^n h_k^{(p)}(X_i)^2 \right] = \frac{n}{n-1} \left[\widehat{a}_{p,k}^2 - \frac{1}{n^2} \sum_{i=0}^n h_k^{(p)}(X_i)^2 \right], \quad (4.9)$$

concluimos que os estimadores $\widehat{R}_{1,p,m}$ e $\widehat{R}_{2,p,m}$ estão relacionados pela igualdade

$$\widehat{R}_{1,p,m} = \frac{n-1}{n} \widehat{R}_{2,p,m} + R_{n,m}, \quad (4.10)$$

com

$$R_{n,m} = \frac{1}{n^2} \sum_{k=0}^m \sum_{i=1}^n h_k^{(p)}(X_i)^2.$$

Dos artigos [Greblicki e Pawlak \(1984, p. 176\)](#) e [Walter \(1977, p. 1260\)](#), temos que

Lema 4.1. *Para $k = 0, 1, \dots$, as derivadas $h_k^{(p)}$ exprimem-se como combinação linear das funções de Hermite. Mais precisamente,*

$$h_k^{(p)}(x) = \sum_{j=-p}^p \alpha_{kpj} h_{k+j}(x), \quad (4.11)$$

em que $h_{-k} = h_k$, para $k \in \mathbb{N}$,

$$|\alpha_{kpj}| \leq K_p (k+p)^{\frac{p}{2}} \quad (4.12)$$

e K_p uma constante dependente de p .

Teorema 4.1. *Se $f^{(p)}, f^{(p-1)}, \dots, f$ são limitadas e de quadrado integrável então*

$$\widehat{R}_{1,p,m} = \frac{n-1}{n} \widehat{R}_{2,p,m} + O_p \left(\frac{m^{p+\frac{5}{6}}}{n} \right).$$

Demonstração. Atendendo a (4.10) basta provar que $E(R_{n,m}) = O \left(m^{p+\frac{5}{6}}/n \right)$. Partindo de

$$\sup_{x \in \mathbb{R}} |h_k(x)| \leq C(k+1)^{-\frac{1}{12}},$$

([Greblicki e Pawlak, 1984](#)) temos, pelo Lema 4.1, que

$$\begin{aligned} \sup_{x \in \mathbb{R}} |h_k^{(p)}(x)| &= \sup_{x \in \mathbb{R}} \left| \sum_{j=-p}^p \alpha_{kpj} h_{k+j}(x) \right| \leq \sum_{j=-p}^p |\alpha_{kpj}| \max_{x \in \mathbb{R}} |h_{k+j}(x)| \\ &\leq K_p C(k+p)^{\frac{p}{2}} \sum_{j=-p}^p (|k+j|+1)^{-\frac{1}{12}}. \end{aligned}$$

Atendendo a que existe $D_p > 0$ tal que

$$(|k+j|+1)^{-\frac{1}{12}} \leq D_p(k+p)^{-\frac{1}{12}},$$

para todo o $k = 0, 1, 2, \dots$ e $j = -p, \dots, p$, concluímos que

$$\sup_{x \in \mathbb{R}} |h_k^{(p)}(x)| \leq C_p(k+p)^{\frac{p}{2}-\frac{1}{12}}, \quad (4.13)$$

com $C_p = K_p C D_p(2p+1)$.

Assim temos que

$$E\left(h_k^{(p)}(X)^2\right) = \int h_k^{(p)}(x)^2 f(x) dx \leq C_p^2(k+p)^{p-\frac{1}{6}} \quad (4.14)$$

e, portanto,

$$E(R_{n,m}) = \frac{1}{n^2} \sum_{k=0}^m \sum_{i=1}^n E\left(h_k^{(p)}(X_i)^2\right) \leq \frac{C_p^2}{n} \sum_{k=0}^m (k+p)^{p-1/6} = O\left(\frac{m^{p+\frac{5}{6}}}{n}\right).$$

■

4.2 Convergência em probabilidade de $\widehat{R}_{1,p,m}$ e $\widehat{R}_{2,p,m}$

Com base nos resultados anteriores, verificamos que apenas precisamos da convergência de um dos estimadores para termos a convergência de ambos. Para obter a convergência de um deles vamos utilizar um resultado de [Greblicki e Pawlak \(1984\)](#) que estabelece a convergência em média quadrática integrada de $\widehat{f}_{p,m}$ para $f^{(p)}$.

Teorema 4.2. *Sejam $f^{(p)}, f^{(p-1)}, \dots, f$ limitadas e de quadrado integrável e $m \rightarrow \infty, n^{-1} m^{p+\frac{5}{6}} \rightarrow 0$ então*

$$\widehat{R}_{j,p,m} \xrightarrow{P} R(f^{(p)})$$

com $j = 1, 2$.

Demonstração. Atendendo a que

$$\widehat{R}_{1,p,m} = \sum_{k=0}^m \widehat{a}_{p,k}^2 = \sum_{k=0}^m (\widehat{a}_{p,k} - a_{p,k})^2 + 2 \sum_{k=0}^m (\widehat{a}_{p,k} - a_{p,k}) a_{p,k} + \sum_{k=0}^m a_{p,k}^2,$$

onde

$$\sum_{k=0}^m a_{p,k}^2 R(f^{(p)}), \quad n \rightarrow \infty,$$

basta que provemos que

$$\sum_{k=0}^m (\widehat{a}_{p,k} - a_{p,k})^2 \xrightarrow{P} 0.$$

Ora, usando (4.13) temos

$$\begin{aligned} E \left(\sum_{k=0}^m (\widehat{a}_{p,k} - a_{p,k})^2 \right) &= \sum_{k=0}^m \text{Var}(\widehat{a}_{p,k}) \leq \sum_{k=0}^m n^{-1} \text{Var}(h_k^{(p)}(X_1)) \\ &\leq \sum_{k=0}^m n^{-1} E(h_k^{(p)}(X_1))^2 \leq \sum_{k=0}^m n^{-1} C_p^2 (k+p)^{p-\frac{1}{6}} \\ &= O(n^{-1} m^{p+\frac{5}{6}}), \end{aligned}$$

de onde concluímos o pretendido usando o facto de $n^{-1} m^{p+\frac{5}{6}} \rightarrow 0$, $n \rightarrow \infty$.

Finalmente, usando o Teorema 4.1 concluímos que também se tem $\widehat{R}_{2,p,m} \xrightarrow{P} R(f^{(p)})$. ■

4.3 Ordem de convergência

Analisemos, agora, a ordem de convergência de $\widehat{R}_{2,p,m}$ para $R(f^{(p)})$. Para tal vamos começar por estabelecer duas proposições sobre os coeficientes de Hermite de $f^{(p)}$ e da função $(x-D)^{r-p} f^{(p)}$ onde o operador $(x-D)^k$, para $k \in \mathbb{N}$ e $x \in \mathbb{R}$, é definido por

$$(x-D)g(x) = xg(x) - g'(x)$$

e

$$(x-D)^k g(x) = (x-D) \left[(x-D)^{k-1} g(x) \right],$$

onde g é uma função real de variável real com derivadas até à ordem k , com $k \in \mathbb{N}$.

Proposição 4.2. *Se $g, g', h, h' \in L_2$, $xg \in L_2$ ou $xh \in L_2$ e $\lim_{x \rightarrow \pm\infty} g(x)h(x) = 0$ então*

$$\int (x-D)g(x)h(x)dx = \int g(x)(x+D)h(x)dx$$

Demonstração. Pela fórmula de integração por partes (ver Bogachev, 2007, p. 343), para $a < b$ temos

$$\int_a^b (x-D)g(x)h(x)dx = \int_a^b xg(x)h(x) - g'(x)h(x)dx = \int_a^b xg(x)h(x) - [g(x)h(x)]_a^b + \int_a^b g(x)h'(x)dx.$$

O resultado enunciado é a consequência do TCD e do facto de $\lim_{|x| \rightarrow \infty} g(x)h(x) = 0$. ■

Nos resultados seguintes, admitiremos, para $r > p$, que as funções $(x-D)^i f^{(p)}$ para $i = 1, \dots, r-p$ são de quadrado integrável. Tal é verdade se $x^{r-i} f^{(i)} \in L_2$ para $i = p, p+1, \dots, r$.

Proposição 4.3. *Dado $r > p$, se $x^{r-i} f^{(i)} \in L_2$ para $i = p, p+1, \dots, r$ então*

$$b_{p,r,j} = (2j)^{\frac{1}{2}} b_{p,r-1,j-1},$$

onde $b_{p,r,j}$ é o coeficiente de Hermite de ordem j de $(x-D)^{r-p} f^{(p)}$ e $j \geq 1$.

Demonstração. Atendendo às igualdades seguintes, estabelecidas em Szegö (1939, p. 106),

$$xh_j = \left(\frac{j}{2}\right)^{\frac{1}{2}} h_{j-1} + \left(\frac{j+1}{2}\right)^{\frac{1}{2}} h_{j+1} \quad e \quad Dh_j = h'_j = \left(\frac{j}{2}\right)^{\frac{1}{2}} h_{j-1} - \left(\frac{j+1}{2}\right)^{\frac{1}{2}} h_{j+1},$$

deduzimos que

$$(x+D)h_j = xh_j + Dh_j = 2\left(\frac{j}{2}\right)^{\frac{1}{2}} h_{j-1} = (2j)^{\frac{1}{2}} h_{j-1}.$$

Desta igualdade e da Proposição 4.2 obtemos

$$\begin{aligned} b_{p,r,j} &= \int (x-D)^{r-p} f^{(p)}(x) h_j(x) dx = \int (x-D)(x-D)^{r-p-1} f^{(p)}(x) h_j(x) dx \\ &= \int (x-D)^{r-p-1} f^{(p)}(x) (x+D) h_j(x) dx = \int (x-D)^{r-p-1} f^{(p)}(x) (2j)^{\frac{1}{2}} h_{j-1}(x) dx \\ &= (2j)^{\frac{1}{2}} b_{p,r-1,j-1}. \end{aligned}$$

■

Na proposição seguinte relacionamos os coeficientes de Hermite de $f^{(p)}$ e de $(x-D)^{r-p} f^{(p)}$.

Proposição 4.4. Dado $r > p$, se $x^{r-i} f^{(i)} \in L_2$ para $i = p, p+1, \dots, r$, então

$$|a_{p,k}| \leq 2^{-\frac{r-p}{2}} (k+1)^{-\frac{r-p}{2}} |b_{p,r,k+(r-p)}|,$$

para $k = 0, 1, \dots$

Demonstração. Para $j \geq r-p$, aplicando repetidamente a Proposição 4.3 temos

$$\begin{aligned} b_{p,r,j} &= (2j)^{\frac{1}{2}} (2(j-1))^{1/2} \dots (2(j-r+p+1))^{\frac{1}{2}} b_{p,r-(r-p),j-(r-p)} \\ &= (2j)^{\frac{1}{2}} (2(j-1))^{\frac{1}{2}} \dots (2(j-r+p+1))^{\frac{1}{2}} b_{p,p,j-r+p}, \end{aligned} \quad (4.15)$$

com

$$\begin{aligned} b_{p,p,j-r+p} &= \int (x-D)^{p-p} f^{(p)}(x) h_{j-r+p}(x) dx \\ &= \int f^{(p)}(x) h_{j-r+p}(x) dx = a_{p,j-r+p}. \end{aligned}$$

Tendo em conta que (4.15) é válida para $r > p$ e $j \geq r-p$ podemos escrever

$$|a_{p,j-(r-p)}| = 2^{-\frac{r-p}{2}} (j(j-1)(j-2) \dots (j-r+p+1))^{-\frac{1}{2}} |b_{p,r,j}|.$$

Assim, majorando

$$\begin{aligned} j &= j - (r-p) + (r-p) > j - (r+p) + 1, \\ j-1 &= j-1 - (r-p) + (r-p) > j - (r+p) + 1, \\ &\dots, \\ j-r+p+1 &= j - (r-p) + 1 \geq j - (r-p) + 1, \end{aligned}$$

obtemos,

$$|a_{p,j-(r-p)}| \leq 2^{-\frac{r-p}{2}} (j - (r-p) + 1)^{-\frac{r-p}{2}} |b_{p,r,j}|. \quad (4.16)$$

Para concluir basta agora fazer $k = j - (r-p)$. ■

Logo, nestas condições, podemos enunciar o seguinte teorema.

Teorema 4.3. *Dado $r > p$, se $f^{(i)} \in L_2 \cap L_\infty$ para $i = 0, \dots, p$, e $x^{r-i} f^{(i)} \in L_2$ para $i = p, p+1, \dots, r$, então*

$$E \left(\widehat{R}_{2,p,m} - R(f^{(p)}) \right)^2 \leq C_1 n^{-1} m^{-\min(0, r-2p-1)} + C_2 n^{-2} m^{2p+\frac{5}{3}} + C_3 m^{-2(r-p)},$$

para $m \in \mathbb{N}$ com $C_1, C_2, C_3 > 0$.

Demonstração. Como já vimos anteriormente, para o cálculo do erro quadrático médio integrado temos uma parcela de variância e uma de quadrado do viés. Tendo em conta (4.7) temos

$$\widehat{R}_{2,p,m} = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} G_n(X_i, X_j). \quad (4.17)$$

onde

$$G_n(x, y) = \sum_{k=0}^m h_k^{(p)}(x) h_k^{(p)}(y).$$

Nestas condições e tendo em conta que temos uma U-estatística de 2ª ordem temos (ver Lee, 1990, p. 12, Teorema 3),

$$\text{Var}(\widehat{R}_{2,p,m}) = \binom{n}{2}^{-1} \sum_{c=1}^2 \binom{2}{c} \binom{n-2}{2-c} \sigma_c^2 = \frac{2}{n(n-1)} (2(n-2)\sigma_1^2 + \sigma_2^2), \quad (4.18)$$

com $\sigma_1^2 = \text{Var}(g_n(X_1))$ onde $g_n(y) = E(G_n(X_1, y))$, e $\sigma_2^2 = \text{Var}(G_n(X_1, X_2))$.

Calculemos então σ_1^2 e σ_2^2 . Para o cálculo do primeiro, temos que calcular $g_n(X_1)$. Da Proposição 4.1, temos

$$g_n(y) = \sum_{k=0}^m E \left(h_k^{(p)}(X_1) \right) h_k^{(p)}(y) = (-1)^p \sum_{k=0}^m a_{p,k} h_k^{(p)}(y).$$

Então,

$$\begin{aligned} \sigma_1^2 &= \text{Var}(g_n(X_1)) \leq E(g_n(X_1))^2 = \int g_n(x)^2 f(x) dx \leq \|f\|_\infty \int g_n(x)^2 dx \\ &= \|f\|_\infty \int \left((-1)^p \sum_{k=0}^m a_{p,k} h_k^{(p)}(x) \right)^2 dx = \|f\|_\infty \int \left(\sum_{k=0}^m a_{p,k} h_k^{(p)}(x) \right)^2 dx \\ &\leq \|f\|_\infty \left(\sum_{k=0}^m |a_{p,k}| \left[\int h_k^{(p)}(x)^2 dx \right]^{\frac{1}{2}} \right)^2. \end{aligned} \quad (4.19)$$

De (4.11) e de ortonormalidade das funções h_k temos,

$$\begin{aligned} \int h_k^{(p)}(x)^2 dx &= \sum_{i,j=-p}^p \alpha_{kpj} \alpha_{kpi} \int h_{k+j}(x) h_{k+i}(x) dx \\ &= \left[\sum_{j=-p}^p \alpha_{kpj}^2 + \sum_{j=-p}^{p-2k} \alpha_{kpj} \alpha_{kp(-2k-j)} \right]. \end{aligned}$$

Usando a desigualdade de Cauchy-Schwarz e a majoração (4.12) temos,

$$\int h_k^{(p)}(x)^2 dx \leq 2 \sum_{j=-p}^p \alpha_{kpj}^2 \leq 2(2p+1)K_p^2(k+p)^p = N_p(k+p)^p, \quad (4.20)$$

com $N_p = 2(2p+1)K_p^2$.

Utilizando a majoração (4.19) e (4.20) temos

$$\sigma_1^2 \leq \|f\|_\infty \left(\sum_{k=0}^m |a_{p,k}| [N_p(k+p)^p]^{\frac{1}{2}} \right)^2 \leq \|f\|_\infty \left(\sum_{k=0}^m |a_{p,k}| N_p'(k+1)^{p/2} \right)^2,$$

para alguma constante $N_p' > 0$ dependente de p . Usando agora a Proposição 4.4, temos

$$\begin{aligned} \sigma_1^2 &\leq \|f\|_\infty N_p'' \left(\sum_{k=0}^m (k+1)^{-(r-p)/2} |b_{p,r,k+r-p}| (k+1)^{p/2} \right)^2 \\ &\leq \|f\|_\infty N_p'' \left(\sum_{k=0}^m (k+1)^{-\frac{r}{2}+p} |b_{p,r,k+r-p}| \right)^2 \\ &\leq \|f\|_\infty N_p'' \left(\sum_{k=0}^m (k+1)^{-r+2p} \right) \left(\sum_{k=0}^m |b_{p,r,k+r-p}|^2 \right), \end{aligned}$$

com $N_p'' > 0$ constante dependente de p .

Se $r-2p > 1$ obtemos uma série convergente e, nesse caso, $\sigma_1^2 = O(1)$. Se $r-2p \leq 1$ obtemos uma série divergente e, nesse caso, $\sigma_1^2 = O(m^{-r+2p+1})$. Assim, para $r > p$, temos

$$\sigma_1^2 = O\left(m^{-\min(0,r-2p-1)}\right). \quad (4.21)$$

Relativamente a σ_2^2 temos

$$\begin{aligned} \sigma_2^2 &= \text{Var}(G_n(X_1, X_2)) \leq E(G_n(X_1, X_2))^2 \leq E \left(\sum_{k=0}^m h_k^{(p)}(X_1) h_k^{(p)}(X_2) \right)^2 \\ &= \sum_{k,k'=0}^m E \left(h_k^{(p)}(X_1) h_k^{(p)}(X_2) h_{k'}^{(p)}(X_1) h_{k'}^{(p)}(X_2) \right) \\ &= \sum_{k,k'=0}^m \left(E \left(h_k^{(p)}(X_1) h_{k'}^{(p)}(X_1) \right) \right)^2 \end{aligned}$$

$$\leq \sum_{k,k'=0}^m E\left(h_k^{(p)}(X_1)\right)^2 E\left(h_{k'}^{(p)}(X_1)\right)^2 = \left(\sum_{k=0}^m E\left(h_k^{(p)}(X_1)\right)^2\right)^2,$$

e, de (4.14), temos que $E\left(h_k^{(p)}(X_1)\right)^2 \leq C_p^2(k+p)^{p-\frac{1}{6}}$.

Assim,

$$\sigma_2^2 \leq \left(\sum_{k=0}^m C_p^2(k+p)^{p-\frac{1}{6}}\right)^2 \leq C_p^4(m+p)^{2(p+\frac{5}{6})} \leq K_p m^{2p+\frac{5}{3}}. \quad (4.22)$$

De (4.18), (4.21) e (4.22) concluímos que existem constantes $C_1, C_2 > 0$ tais que

$$\text{Var}(\widehat{R}_{2,p,m}) \leq C_1 n^{-1} m^{-\min(0, r-2p-1)} + C_2 n^{-2} m^{2p+\frac{5}{3}}.$$

Vamos agora analisar o viés de $\widehat{R}_{2,p,m}$, que já vimos ser dado por $\text{Vies}(\widehat{R}_{2,p,m}) = \sum_{k=m+1}^{\infty} a_{p,k}^2$.

Partindo da hipótese $x^{r-i} f^{(i)} \in L_2$ com $i = p, p+1, p+2, \dots, r$ e atendendo à Proposição 4.4 temos

$$\sum_{k=m+1}^{\infty} a_{p,k}^2 \leq \sum_{k=m+1}^{\infty} k^{-(r-p)} b_{p,r,k+r-p}^2 \leq m^{-(r-p)} \sum_{k=m+1}^{\infty} b_{p,r,k+r-p}^2 = O\left(m^{-(r-p)}\right),$$

uma vez que $\sum_{k=m+1}^{\infty} b_{p,r,j}^2 < \sum_{j=0}^{\infty} b_{p,r,j}^2 < \infty$ visto que $(x-D)^{r-p} f^{(p)} \in L_2$.

Assim, para o quadrado do viés temos $\text{Vies}(\widehat{R}_{2,p,m})^2 = O\left(m^{-2(r-p)}\right)$, concluindo, assim, a demonstração. ■

Do resultado anterior temos,

Teorema 4.4. Para $r > p$, sejam $f^{(i)} \in L_2 \cap L_{\infty}$, com $i = 0, \dots, p$ e $x^{r-i} f^{(i)} \in L_2$ com $i = p, p+1, \dots, r$. Se

$$\beta_1 n^{\varepsilon} \leq m \leq \beta_2 n^{\varepsilon}, \quad (4.23)$$

com

$$0 < \varepsilon < \frac{1}{p+1},$$

e β_1, β_2 constantes positivas, então

$$\widehat{R}_{j,p,m} - R(f^{(p)}) = \begin{cases} O_p\left(n^{-\min(\varepsilon(r-p), 1-\varepsilon(p+\frac{5}{6}))}\right), & p < r < 2p+1 \\ O_p\left(n^{-\min(\frac{1}{2}, \varepsilon(r-p), 1-\varepsilon(p+\frac{5}{6}))}\right), & r \geq 2p+1 \end{cases},$$

com $j = 1, 2$.

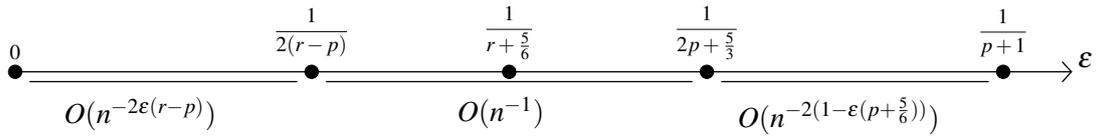
Demonstração. Do Teorema 4.3 verificamos a necessidade de analisar dois casos: o caso em que $p < r < 2p+1$ e o caso que $r \geq 2p+1$.

Para $r \geq 2p + 1$, vamos ter que

$$E\left(\widehat{R}_{2,p,m} - R(f^{(p)})\right)^2 \leq C_1 n^{-1} + C_2 n^{-(2-\varepsilon(2p+\frac{5}{3}))} + C_3 n^{-2\varepsilon(r-p)}$$

em que C_1, C_2, C_3 são constantes positivas.

Assim, para os seguintes intervalos de variação de ε , temos para $E\left(\widehat{R}_{2,p,m} - R(f^{(p)})\right)^2$ as ordens de convergência indicadas,



Assim concluímos que

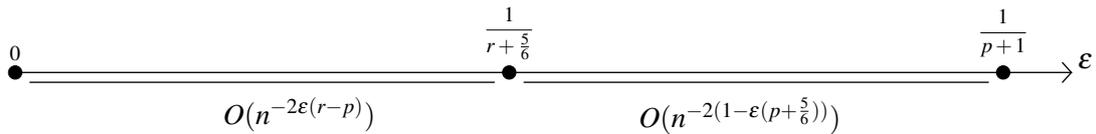
$$\widehat{R}_{2,p,m} - R(f^{(p)}) = O_p\left(n^{-\min(\frac{1}{2}, \varepsilon(r-p), 1-\varepsilon(p+\frac{5}{6}))}\right).$$

Para $p < r < 2p + 1$, vamos ter que

$$E\left(\widehat{R}_{2,p,m} - R(f^{(p)})\right)^2 \leq C_1 n^{-(1+\varepsilon(r-2p-1))} + C_2 n^{-(2-\varepsilon(2p+\frac{5}{3}))} + C_3 n^{-2\varepsilon(r-p)}.$$

em que C_1, C_2, C_3 são constantes positivas.

Assim, para os seguintes intervalos de variação de ε temos, para $E\left(\widehat{R}_{2,p,m} - R(f^{(p)})\right)^2$, as ordens de convergências indicadas,



Logo,

$$\widehat{R}_{2,p,m} - R(f^{(p)}) = O_p\left(n^{-\min(\varepsilon(r-p), 1-\varepsilon(p+1))}\right).$$

Verificado o resultado dado no teorema para $\widehat{R}_{2,p,m}$, analisemos agora a ordem do estimador $\widehat{R}_{1,p,m}$. Do Teorema 4.1,

$$\widehat{R}_{1,p,m} - R(f^{(p)}) = \begin{cases} O_p\left(n^{-\min(\varepsilon(r-p), 1-\varepsilon(p+\frac{5}{6}))}\right) + O(n^{-1+\varepsilon(p+5/6)}), & p < r < 2p + 1 \\ O_p\left(n^{-\min(\frac{1}{2}, \varepsilon(r-p), 1-\varepsilon(p+\frac{5}{6}))}\right) + O(n^{-1+\varepsilon(p+5/6)}), & r \geq 2p + 1 \end{cases}$$

No entanto, facilmente verificamos que os termos dominantes para $\widehat{R}_{2,p,m}$ vão dominar $O(n^{-1+\varepsilon(p+5/6)})$ e assim, obtemos o pretendido. ■

4.4 Comportamento assintótico das janelas $\widehat{h}_{1,j,n}$ e $\widehat{h}_{2,j,n}$

Analisada a convergência e a ordem de $\widehat{R}(f^{(p)})$ para $R(f^{(p)})$, onde $\widehat{R}(f^{(p)})$ denota um dos estimadores $\widehat{R}_{1,p,m}$ ou $\widehat{R}_{2,p,m}$, e considerando estes estimadores, vamos definir as janelas automática, isto é, dependente X_1, X_2, \dots, X_n , baseadas em (3.10) e (3.13) que são dadas por

$$\widehat{h}_{1,j,n} = C_{1,K} \widehat{R}_{j,2,m}^{-\frac{1}{5}} n^{-\frac{1}{5}} \quad e \quad \widehat{h}_{2,j,n} = C_{1,K} \widehat{R}_{j,2,m}^{-\frac{1}{5}} n^{-\frac{1}{5}} + C_{2,K} \widehat{R}_{j,2,m}^{-\frac{8}{5}} \widehat{R}_{j,3,m} n^{-\frac{3}{5}}, \quad (4.24)$$

com $j = 1, 2$, consoante o estimador que utilizaremos.

O próximo resultado vai permitir-nos estudar os comportamentos assintóticos do estimador das janelas.

Teorema 4.5. *Sejam $K \in L_2$ um núcleo simétrico com $\int u^6 |K(u)| du < \infty$ e f com derivadas até à ordem $r \geq 5$, onde $f, \dots, f^{(4)}$ são limitadas e integráveis e $x^{r-i} f^{(i)} \in L_2$ com $i = 2, \dots, r$.*

(a) *Se $m \rightarrow \infty$, $n^{-1} m^{2+\frac{5}{6}} \rightarrow 0$ então*

$$\frac{\widehat{h}_{1,j,n}}{h_{0,n}} \xrightarrow{P} 1.$$

Mais, se m satisfaz (4.23) com

$$\frac{2}{5(r-2)} \leq \varepsilon \leq \frac{1}{5} \times \frac{18}{17}$$

então

$$n^{2/5} \left(\frac{\widehat{h}_{1,j,n}}{h_{0,n}} - 1 \right) \xrightarrow{P} C_{1,K}^{-1} C_{2,K} R(f'')^{-\frac{7}{5}} R(f''').$$

(b) *Se $m \rightarrow \infty$, $n^{-1} m^{3+\frac{5}{6}} \rightarrow 0$ então*

$$\frac{\widehat{h}_{2,j,n}}{h_{0,n}} \xrightarrow{P} 1.$$

Mais, se m satisfaz (4.23) com

$$\frac{1}{2(r-2)} \leq \varepsilon \leq \frac{1}{6} \times \frac{18}{17}$$

então

$$\frac{\widehat{h}_{2,j,n}}{h_{0,n}} - 1 = O_p(n^{-\frac{1}{2}}).$$

Demonstração. Do Teorema 3.7 temos

$$h_{0,n} = C_{1,K} R(f'')^{-\frac{1}{5}} n^{-\frac{1}{5}} + C_{2,K} R(f'')^{-\frac{8}{5}} R(f''') n^{-\frac{3}{5}} + O(n^{-1}) \quad (4.25)$$

com $C_{1,K} = R(K)^{\frac{1}{5}} \mu_2(K)^{-\frac{2}{5}}$ e $C_{2,K} = \frac{1}{20} R(K)^{\frac{3}{5}} \mu_2(K)^{-\frac{22}{5}} \mu_4(K)$.

(a) Da definição de $\widehat{h}_{1,j,n}$ e de (4.25) temos

$$\frac{\widehat{h}_{1,j,n}}{h_{0,n}} - 1 = \frac{C_{1,K}(\widehat{R}(f'')^{-\frac{1}{5}} - R(f'')^{-\frac{1}{5}})}{n^{\frac{1}{5}}h_{0,n}} - \frac{C_{2,K}R(f'')^{-\frac{8}{5}}R(f''')n^{-\frac{2}{5}}}{n^{\frac{1}{5}}h_{0,n}} + O(n^{-\frac{4}{5}}), \quad (4.26)$$

onde, do Teorema 4.4,

$$\widehat{R}(f'') - R(f'') = O_p\left(n^{-\min(\frac{1}{2}, \varepsilon(r-2), 1-3\varepsilon)}\right) = \begin{cases} O_p(n^{-\varepsilon(r-2)}) & 0 < \varepsilon \leq \frac{1}{2(r-2)} \\ O_p(n^{-1/2}) & \frac{1}{2(r-2)} \leq \varepsilon \leq \frac{3}{17} \\ O_p(n^{-1+\frac{17}{6}\varepsilon}) & \frac{3}{17} \leq \varepsilon < \frac{1}{3}. \end{cases}$$

Assim, para $\frac{2}{5(r-2)} \leq \varepsilon \leq \frac{1}{5}$, o primeiro termo de (4.26) é um $o_p(n^{-\frac{2}{5}})$, de onde decorre

$$\frac{\widehat{h}_{1,j,n}}{h_{0,n}} - 1 = \frac{C_{2,K}R(f'')^{-\frac{8}{5}}R(f''')n^{-\frac{2}{5}}}{n^{\frac{1}{5}}h_{0,n}} + o_p(n^{-\frac{2}{5}}).$$

Finalmente, atendendo ao Teorema 3.5 temos

$$n^{\frac{2}{5}} \left(\frac{\widehat{h}_{1,j,n}}{h_{0,n}} - 1 \right) \xrightarrow{P} C_{1,K}^{-1} C_{2,K} R(f'')^{-\frac{7}{5}} R(f''').$$

(b) Da definição de $\widehat{h}_{2,j,n}$ e de (4.25) temos

$$\begin{aligned} \frac{\widehat{h}_{2,j,n}}{h_{0,n}} - 1 &= \frac{C_{1,K}(\widehat{R}(f'')^{-\frac{1}{5}} - R(f'')^{-\frac{1}{5}}) + C_{2,K} \left(\widehat{R}(f'')^{-\frac{8}{5}} \widehat{R}(f''') - R(f'')^{-\frac{8}{5}} R(f''') \right) n^{-\frac{2}{5}}}{n^{\frac{1}{5}}h_{0,n}} + O(n^{-\frac{4}{5}}) \\ &= \frac{C_{1,K} \left(\widehat{R}(f'')^{-\frac{1}{5}} - R(f'')^{-\frac{1}{5}} \right) + C_{2,K} \left(\widehat{R}(f'')^{-\frac{8}{5}} - R(f'')^{-\frac{8}{5}} \right) \widehat{R}(f''')^{-\frac{8}{5}} n^{-\frac{2}{5}}}{n^{\frac{1}{5}}h_{0,n}} \\ &\quad + \frac{C_{2,K} \left(\widehat{R}(f''') - R(f''') \right) R(f'')^{-\frac{8}{5}} n^{-\frac{2}{5}}}{n^{\frac{1}{5}}h_{0,n}} + O(n^{-\frac{4}{5}}). \end{aligned}$$

Para $\frac{1}{2(r-2)} \leq \varepsilon \leq \frac{1}{6} \times \frac{18}{17}$, do Teorema 4.4, temos que $\widehat{R}(f'') - R(f'') = O_p(n^{-\frac{1}{2}})$ e $\widehat{R}(f''') - R(f''') = O_p(n^{-\frac{1}{10}})$, o que nos permite concluir o pretendido. ■

Capítulo 5

Janelas *plug-in* baseadas nas funções de Hermite

Estudámos no capítulo anterior o comportamento assintótico das janelas *plug-in*

$$\widehat{h}_{1,j,n} = C_{1,K} \widehat{R}_{j,2,m}^{-\frac{1}{5}} n^{-\frac{1}{5}}$$

e

$$\widehat{h}_{2,j,n} = C_{1,K} \widehat{R}_{j,2,m}^{-\frac{1}{5}} n^{-\frac{1}{5}} + C_{2,K} \widehat{R}_{j,2,m}^{-\frac{8}{5}} \widehat{R}_{j,3,m} n^{-\frac{3}{5}},$$

para $j = 1, 2$ onde

$$\widehat{R}_{1,p,m} = \sum_{k=0}^m \widehat{a}_{p,k}^2 \quad e \quad \widehat{R}_{2,p,m} = \sum_{k=0}^m \widehat{a}_{p,k}^2.$$

Para serem usadas na prática, as janelas anterior deveriam poder ser calculadas a partir da amostra X_1, \dots, X_n observada. Tal não acontece uma vez que ambas as janelas dependem também do número de parcelas, m , consideradas nos estimadores $\widehat{R}_{1,p,m}$ e $\widehat{R}_{2,p,m}$.

Neste capítulo analisamos a escolha automática de m , isto é, a escolha m com base nas observações, isto é, $m = \widehat{m}(X_1, \dots, X_n)$, bem como o comportamento a distância finita das janelas $\widehat{h}_{1,j,n}$ e $\widehat{h}_{2,j,n}$ associadas.

5.1 Escolha automática de m

O método que consideramos para a escolha automática de m é inspirado em [Hart \(1985\)](#) e [Tenreiro \(2011\)](#). Para $f \in L_2$, o erro quadrático médio integrado do estimador de Hermite $\widehat{f}_{0,m}$ de f , definido por (4.3), pode ser expresso como (ver [Schwartz, 1967](#), p. 1263)

$$E \int (\widehat{f}_{0,m}(x) - f(x))^2 dx = \sum_{k=0}^m V(\widehat{a}_{0,k}) + \sum_{k=m+1}^{\infty} a_{0,k}^2 = \sum_{k=0}^m \frac{1}{n} V(h_k(X_1)) + \sum_{k=m+1}^{\infty} a_{0,k}^2$$

onde

$$\text{Var}(h_k(X_1)) = \frac{1}{n} \int h_k^2(x) f(x) dx - \frac{1}{n} a_{0,k}^2.$$

Assim,

$$E \int (\hat{f}_{0,m}(x) - f(x))^2 dx = H(m) + \sum_{k=0}^{\infty} a_{0,k}^2$$

com

$$H(m) = \frac{1}{n} \sum_{k=0}^m \int h_k^2(x) f(x) dx - \left(\frac{n+1}{n} \right) \sum_{k=0}^m a_{0,k}^2.$$

Como o último termo da expressão anterior não depende de m , a proposta feita por Hart (1985), é a de tomar para m o primeiro inteiro \hat{m}_H satisfazendo

$$\hat{m}_H = \arg \min_{m \in \mathcal{M}_n} \hat{H}(m),$$

onde $\mathcal{M}_n = \{L_n, L_n + 1, \dots, U_n\}$ e, $L_n < U_n$ são sucessões deterministas de números inteiros positivos, e $\hat{H}(m)$ é o estimador cêntrico de $H(m)$ dado por

$$\hat{H}(m) = \frac{1}{n} \sum_{k=0}^m \frac{1}{n} \sum_{i=1}^n h_k(X_i)^2 - \frac{n+1}{n} \sum_{k=0}^m \hat{a}_{0,k}^2$$

onde $\hat{a}_{0,k}^2$ é definido por (4.9).

O valor \hat{m}_H depende de \mathcal{M}_n através das sucessões L_n e U_n que necessitam de ser escolhidas por nós. Para isso, iremos considerá-los iguais a

$$L_n = \lfloor C_1 n^\varepsilon \rfloor \quad e \quad U_n = \lfloor C_2 n^\varepsilon \rfloor,$$

onde $\lfloor x \rfloor$ é a parte inteira de x , e C_1, C_2 e ε são constantes estritamente positivas nas condições do Teorema 4.5.

Tomando $r = 5$ no referido teorema, concluímos que as maiores ordens de convergência para as janelas $\hat{h}_{1,j,n}$ são obtidas tomando $\frac{2}{15} \leq \varepsilon \leq \frac{1}{5} \times \frac{18}{17}$ e para as janelas $\hat{h}_{2,j,n}$ são obtidas tomando $\frac{1}{6} \leq \varepsilon \leq \frac{1}{6} \times \frac{18}{17}$. Nestas condições, somos levadas a tomar $\varepsilon = 1/6$.

Uma vez que as sucessões L_n e U_n dependem fortemente das constantes C_1 e C_2 , para que possamos ganhar alguma sensibilidade sobre a escolha mais apropriada dessas constantes, apresentamos na Figura 5.1 a distribuição empírica do erro quadrático integrado (ISE, do inglês, Integrated Square Error) em função de m para as misturas de normais, sendo elas as distribuições #1, #3, #8 e #12 da Figura 3.1. Os gráficos são relativos à janela $\hat{h}_{1,1,n}$, mas semelhantes conclusões podem ser tiradas a partir das outras janelas. Verificamos que temos distribuições que requerem um valor de m pequeno, caso da #1 e #8, e temos distribuições que precisam de um valor de m superior, caso das distribuições #3 e #12.

Verificamos que temos distribuições que requerem um valor de m pequeno, caso da #1 e #8, e temos distribuições que precisam de um valor de m superior, caso das distribuições #3 e #12. Nestas condições, concluímos as escolhas das constantes têm de ser feitas de modo a permitir que em \mathcal{M}_n tenha valores pequenos e grandes para m . Assim, tomamos $C_1 = 0.2$ e $C_2 = 40$ que nos conduz a $L_n = 0$ e $58 \leq U_n \leq 185$, para $10 \leq n \leq 10^4$.

Observamos assim que para densidades com características distribucionais mais complicadas é necessário tomar U_n grande. No entanto, esta escolha de U_n faz com que para densidades com

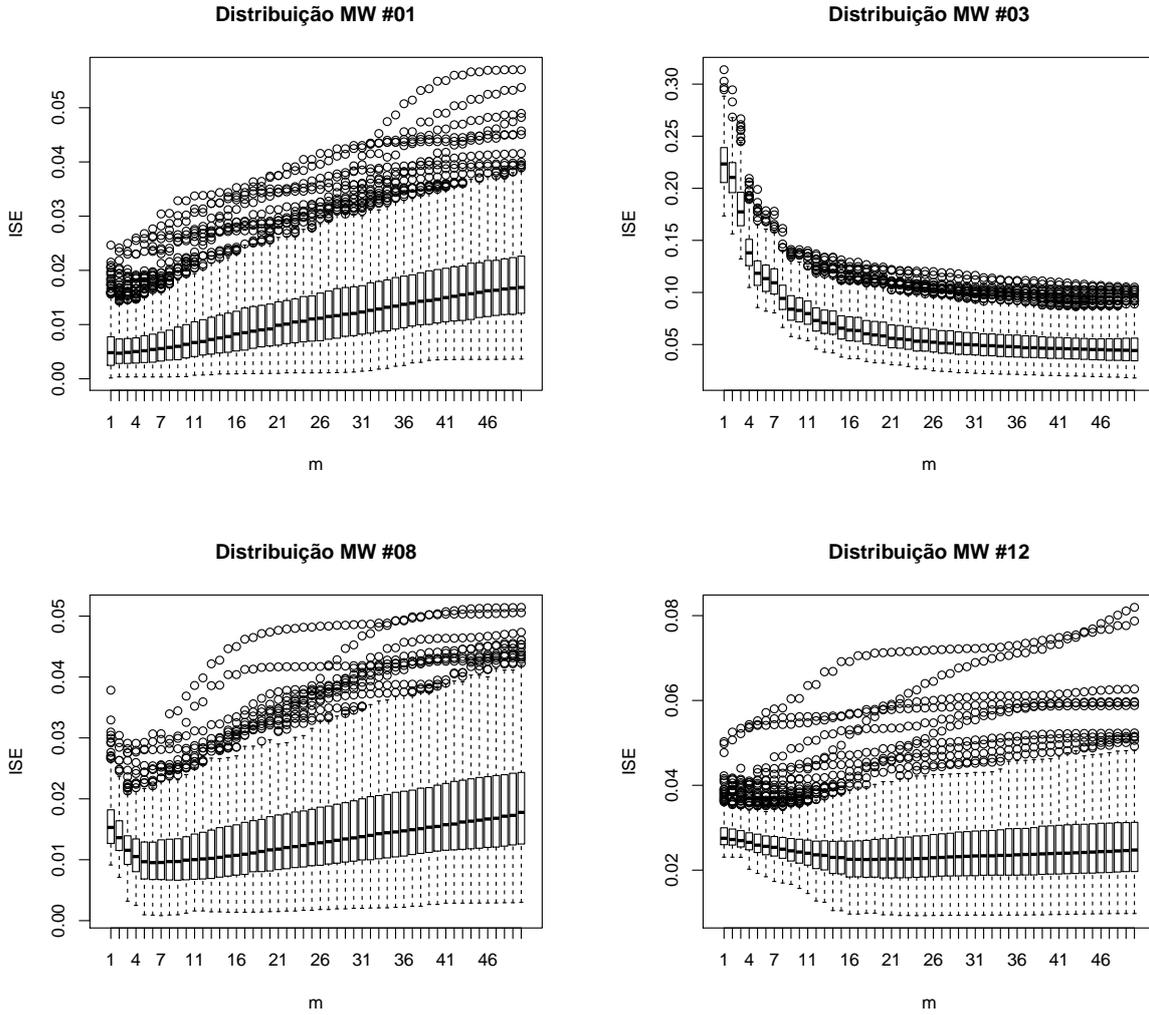


Fig. 5.1 Distribuição empírica de $ISE(f; n, \hat{h}_{1,1,n})$ em função de m , para $n=100$, com 500 repetições.

caraterísticas distribucionais mais simples seja selecionado um valor \hat{m}_H demasiado grande. Uma forma de contornar esta dificuldade, sugerida por Hart (1985), é substituir \hat{H} pela função $\hat{H}_\gamma(m)$, dada por

$$\hat{H}_\gamma(m) = \frac{1}{n} \sum_{k=0}^m \frac{1}{n} \sum_{i=1}^n h_k(X_i)^2 - \gamma \frac{n+1}{n} \sum_{k=0}^m \tilde{a}_{0,k}^2,$$

onde $0 < \gamma < 1$. Assim tomamos para o primeiro inteiro \hat{m}_{H_γ} satisfazendo

$$\hat{m}_{H_\gamma} = \arg \min_{m \in \mathcal{M}_n} \hat{H}_\gamma(m).$$

Como provamos no resultado seguinte, o valor \hat{m}_{H_γ} é efetivamente melhor do que \hat{m}_H , o que nos leva a que \hat{m}_{H_γ} seja preferível a \hat{m}_H para densidades com características distribucionais mais simples.

Teorema 5.1. Para $0 < \gamma_1 < \gamma_2 \leq 1$ temos $P(\hat{m}_{H_{\gamma_1}} \leq \hat{m}_{H_{\gamma_2}}) = 1$

Demonstração. Começemos por escrever que

$$\widehat{H}_\gamma(m) = L(m)(1 - \gamma) + \widehat{H}(m), \quad (5.1)$$

onde $L(m) = \frac{1}{n} \sum_{k=0}^m \frac{1}{n} \sum_{i=1}^n h_k(X_i)^2$. Por definição, $m_{H_{\gamma_1}}$ e $m_{H_{\gamma_2}}$ temos

$$\widehat{H}_{\gamma_1}(m_{H_{\gamma_1}}) \leq \widehat{H}_{\gamma_1}(m_{H_{\gamma_2}}) \text{ e } \widehat{H}_{\gamma_2}(m_{H_{\gamma_2}}) \leq \widehat{H}_{\gamma_2}(m_{H_{\gamma_1}}).$$

Assim,

$$L(m_{H_{\gamma_1}})(1 - \gamma_1) + \gamma_1 \widehat{H}(m_{H_{\gamma_1}}) \leq L(m_{H_{\gamma_2}})(1 - \gamma_1) + \gamma_1 \widehat{H}(m_{H_{\gamma_2}})$$

e

$$L(m_{H_{\gamma_2}})(1 - \gamma_2) + \gamma_2 \widehat{H}(m_{H_{\gamma_2}}) \leq L(m_{H_{\gamma_1}})(1 - \gamma_2) + \gamma_2 \widehat{H}(m_{H_{\gamma_1}}),$$

ou seja,

$$L(m_{H_{\gamma_1}}) \left(\frac{1 - \gamma_1}{\gamma_1} \right) + \widehat{H}(m_{H_{\gamma_1}}) \leq L(m_{H_{\gamma_2}}) \left(\frac{1 - \gamma_1}{\gamma_1} \right) + \widehat{H}(m_{H_{\gamma_2}})$$

e

$$L(m_{H_{\gamma_2}}) \left(\frac{1 - \gamma_2}{\gamma_2} \right) + \widehat{H}(m_{H_{\gamma_2}}) \leq L(m_{H_{\gamma_1}}) \left(\frac{1 - \gamma_2}{\gamma_2} \right) + \widehat{H}(m_{H_{\gamma_1}}).$$

Adicionando as duas últimas desigualdades, obtemos sucessivamente

$$L(m_{H_{\gamma_1}}) \left(\frac{1 - \gamma_1}{\gamma_1} \right) + L(m_{H_{\gamma_2}}) \left(\frac{1 - \gamma_2}{\gamma_2} \right) \leq L(m_{H_{\gamma_2}}) \left(\frac{1 - \gamma_1}{\gamma_1} \right) + L(m_{H_{\gamma_1}}) \left(\frac{1 - \gamma_2}{\gamma_2} \right),$$

$$\frac{1 - \gamma_1}{\gamma_1} (L(m_{H_{\gamma_2}}) - L(m_{H_{\gamma_1}})) + \frac{1 - \gamma_2}{\gamma_2} (L(m_{H_{\gamma_1}}) - L(m_{H_{\gamma_2}})) \geq 0,$$

$$(L(m_{H_{\gamma_2}}) - L(m_{H_{\gamma_1}})) \left(\frac{1 - \gamma_1}{\gamma_1} - \frac{1 - \gamma_2}{\gamma_2} \right) \geq 0,$$

e

$$(L(m_{H_{\gamma_2}}) - L(m_{H_{\gamma_1}}))(\gamma_2 - \gamma_1) \geq 0. \quad (5.2)$$

Uma vez que $\gamma_1 < \gamma_2$ e L é estritamente constante com probabilidade 1, temos que $m_{H_{\gamma_2}} \geq m_{H_{\gamma_1}}$. ■

Depois de algum estudo concluímos que para distribuições mais simples, valores inferiores de γ dão melhores resultados. O mesmo se verifica para as mais complexas mas neste caso para valores superiores de γ . De modo equilibrar a situação, vamos escolher $\gamma = 0.5$ que para distribuições simples faz a diferença e as distribuições mais complexas não são muito afetadas, que podemos verificar na Figura 5.2.

5.2 Análise comparativa

Procederemos, agora, ao estudo comparativo entre as janelas que obtivemos através da estimação por funções de Hermite com as janelas dadas pelos métodos *plug-in* e validação cruzada. Estas duas últimas janelas, foram escolhidas devido ao facto de as janelas obtidas pelo método *plug-in* mostrarem

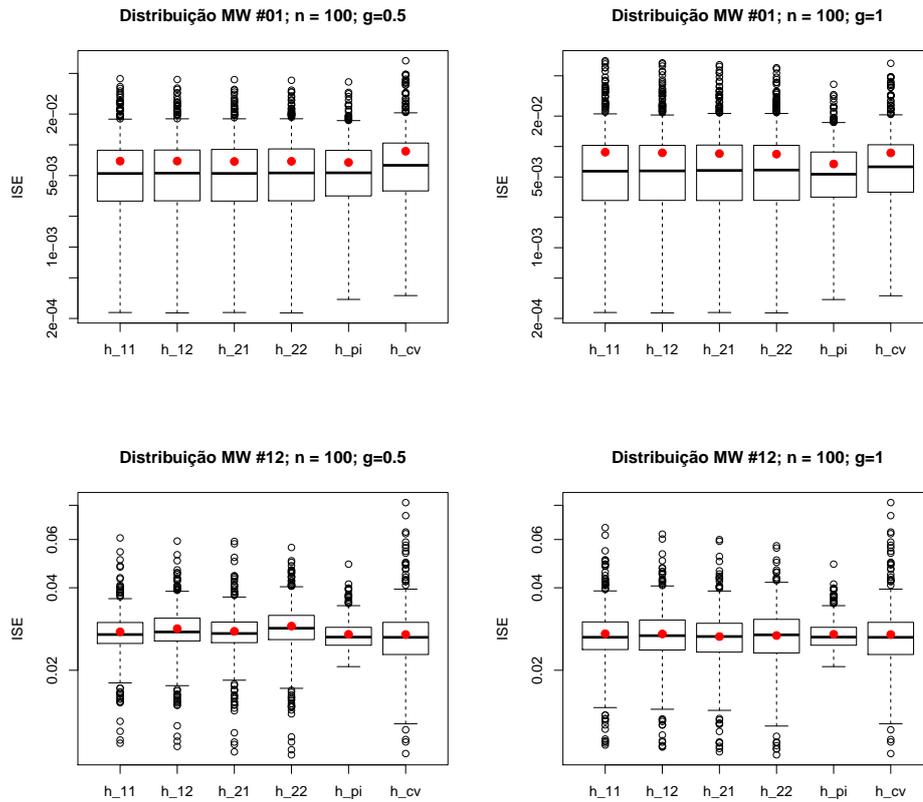


Fig. 5.2 Distribuição $ISE(f; n, h)$ com $h = h_{1,j,n}$ e $h = h_{2,j,n}$, com $j = 1$ e $j = 2$ para $n = 100$. Com 500 repetições.

melhores resultados para distribuições com características distribucionais simples e as janelas da validação cruzada terem melhores resultados para funções com características mais complexas.

Para a implementação dos estimadores, dados (4.6) e (4.8), precisamos de equações recursivas, não só para o sistema ortonormado de Hermite, $\{h_k, k = 0, 1, \dots\}$, bem como para as suas derivadas até à terceira ordem. Deste modo precisamos de equações de recorrência para cada um.

Por (4.1) concluímos que para $k = 0$ e $k = 1$ temos, respetivamente,

$$h_0(x) = \pi^{-\frac{1}{4}} e^{-\frac{x^2}{2}}$$

e

$$h_1(x) = \sqrt{2x} \pi^{-\frac{1}{4}} e^{-\frac{x^2}{2}}.$$

Uma vez que para o polinómio de Hermite de grau k a igualdade seguinte é verificada, (ver Szegő, 1939, pág.106)

$$H_k(x) = 2xH_{k-1}(x) - 2(k-1)H_{k-2}(x) \quad (5.3)$$

temos, para $k \geq 2$, a seguinte equação de recorrência,

$$h_{k+1}(x) = \left(\frac{2}{k+1}\right)^{1/2} x h_k(x) - \left(\frac{k}{k+1}\right)^{1/2} h_{k-1}(x).$$

Para obtermos a equação de recorrência da primeira derivada, comecemos por derivar (4.1) e, para $k = 0$, temos

$$h'_0(x) = -x\pi^{-1/4} e^{-x^2/2}.$$

Considerando (5.3) e a seguinte igualdade também presente em Szegö (1939, pág.106),

$$H'_k(x) = 2kH_{k-1}(x)$$

temos, para $k \geq 1$,

$$h'_k(x) = -xh_k(x) + (2k)^{1/2}h_{k-1}(x). \quad (5.4)$$

Derivando (5.4) obtemos

$$h''_k(x) = -h_k(x) - xh'_k(x) + (2k)^{1/2}h'_{k-1}(x)$$

e

$$h'''_k(x) = -2h'_k(x) - xh''_k(x) + (2k)^{1/2}h''_{k-1}(x),$$

tendo que considerar, respetivamente, para $k = 0$ as seguintes expressões provenientes das derivadas de (4.1),

$$h''_0(x) = \pi^{-1/4}(x^2 - 1)e^{-x^2/2}$$

e

$$h'''_0(x) = \pi^{-1/4}(3x - x^3)e^{-x^2/2}.$$

Nestas condições, temos tudo o que é necessário para estimar as janelas obtidas pelo método desenvolvido. Neste estudo vamos considerar quatro tamanhos possíveis de amostras $n = 50, 100, 150$ e 200 , K será normal standard, repetiremos 500 vezes e, também, consideramos as oito misturas de densidades normais da Figura 3.1.

Para cada janela $\hat{h}_{1,1,n}, \hat{h}_{1,2,n}, \hat{h}_{2,1,n}, \hat{h}_{2,2,n}, h_{pi}$ e h_{cv} , descreveremos o comportamento da seguinte medida estocástica do desempenho da janela $\hat{h} = \hat{h}(X_1, \dots, X_n)$:

$$\text{norma } L_2 \text{ do ISE}(f; n, \hat{h}) = \sqrt{\text{Var}(\text{ISE}(f; n, \hat{h})) + E^2(\text{ISE}(f; n, \hat{h}))}$$

e analisamos este comportamento na Figura 5.3.

Como era de esperar a janela h_{pi} comporta-se bem nas distribuições com características simples e a janela h_{cv} comporta-se melhor para distribuições mais complexas. Pela Figura 5.3, verificamos que, para os tamanhos de amostras que estamos a considerar, as janelas $\hat{h}_{1,j,n}$ e $\hat{h}_{2,j,n}$ têm, no geral, bons resultados quer para distribuições complexas quer para distribuições simples uma vez que se encontram sempre entre as duas janelas de comparação ou até mesmo a baixo do melhor método.

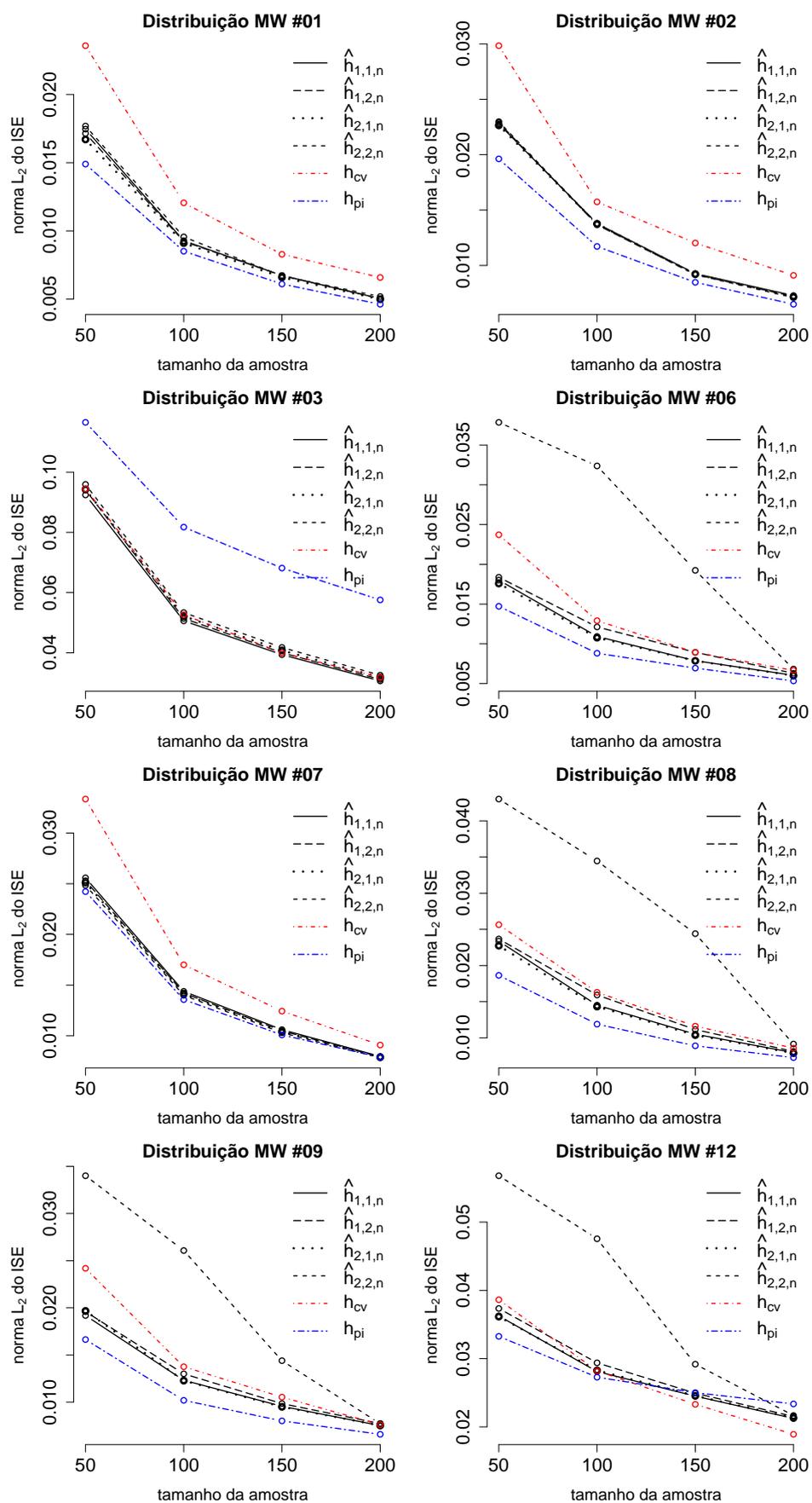


Fig. 5.3 Norma L_2 dos ISE's das distribuições associadas as distribuições da figura 3.1

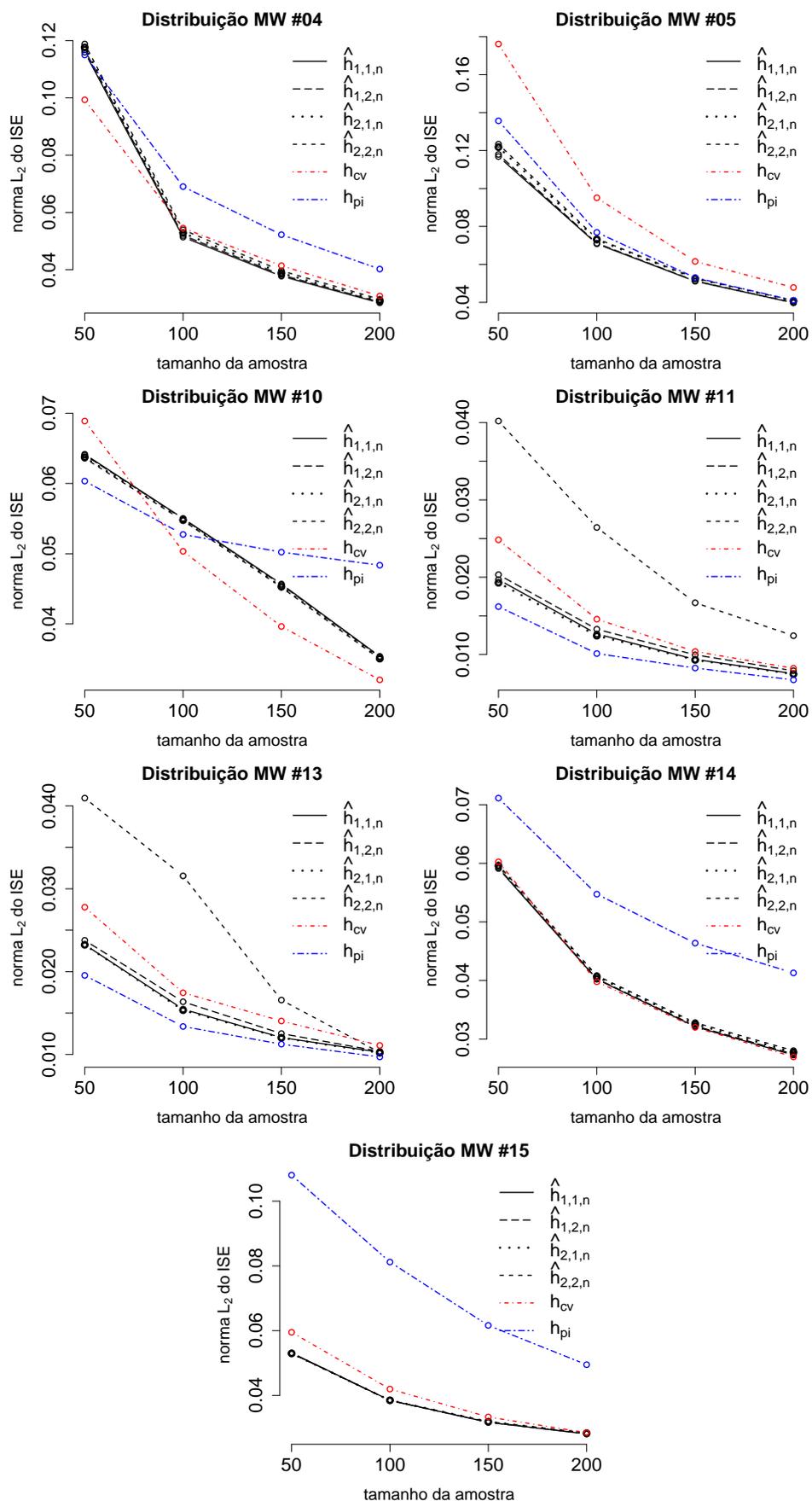


Fig. 5.4 Norma L_2 dos ISE's das distribuições associadas as restantes distribuições presentes em (Marron e Wand, 1992)

Mais, para todas as densidades consideradas as janelas $\hat{h}_{1,1,n}$, $\hat{h}_{1,2,n}$, $\hat{h}_{2,1,n}$ têm comportamentos muito semelhantes. No caso da janela $\hat{h}_{2,2,n}$ apresenta maus resultados para algumas densidades consideradas, como são os casos das densidades #6, #8, #9 e #12. Destacamos a janela $\hat{h}_{1,1,n}$ que, para os tamanhos de amostras considerados, é a que tem melhores resultados das nossas quatro janelas.

Estas conclusões são reforçadas pelos gráficos da Figura 5.4, que representam gráficos análogos a Figura 5.3 mas das restantes misturas de densidades normais presentes em Marron e Wand (1992), onde podemos ver que para as distribuições #14 e #15, para os tamanhos de amostras considerados, as janelas por nós obtidas são de facto a melhor opção e nas outras distribuições, a excepção da janela $h_{2,2,n}$, temos bons resultados com as janelas obtidas pelas funções de Hermite.

Tendo em conta os resultados da simulação, como o método *plug-in* baseado nas funções de Hermite se comporta próximo do melhor dos métodos considerados para cada distribuição, concluímos que deve ser tido em consideração para a estimação das densidades uma vez que não sabemos a complexidade das características distribucionais e ao decidirmos utilizar o método *plug-in* ou o método da validação cruzada podemos estar a obter maus resultados para a estimação da densidade.

Anexo A

Alguns resultados auxiliares

Os dois seguintes resultados são consequência do Teorema da Convergência Dominada de Lebesgue (TCD) que pode ser encontrado em [Cohn \(1980\)](#).

Lema A.1 (continuidade sob o sinal de integral). *Sejam f uma aplicação real definida em $\mathbb{R} \times U$ onde U é um subconjunto aberto de \mathbb{R} , e y_0 fixo em U . Se*

- a) *para cada $x \in \mathbb{R}$, $y \mapsto f(x, y)$ é contínua em y_0 ,*
 - b) *existe g integrável em \mathbb{R} tal que, para todo $y \in U$, $|f(\cdot, y)| \leq g(\cdot)$,*
- então, a função real $y \mapsto \int f(x, y) dx$ é contínua em y_0 .*

Lema A.2 (derivação sob o sinal de integral). *Sejam f uma aplicação real definida em $\mathbb{R}^d \times U$, em que U é um subconjunto aberto de \mathbb{R} , e y_0 fixo em U . Se*

- a) *para cada $y \in U$, a função $x \mapsto f(x, y)$ é integrável,*
 - b) *para cada $x \in \mathbb{R}^d$, $y \mapsto \frac{\partial f}{\partial y}(x, y)$ existe numa vizinhança de y_0 ,*
 - c) *existe g integrável em \mathbb{R}^d tal que, para todo $y \in U$, $\left| \frac{\partial f}{\partial y}(\cdot, y) \right| \leq g(\cdot)$,*
- então, a função real $y \mapsto \int f(x, y) dx$ é diferenciável em y_0 e $\frac{\partial}{\partial y} \left(\int f(x, y) dx \right) \Big|_{y=y_0} = \int \frac{\partial f}{\partial y}(x, y_0) dx$.*

Nos resultados seguintes, que podem ser obtidos usando os lemas anteriores, g e h são funções reais de variável real.

Proposição A.1. *Sejam g com derivadas limitadas até a ordem r e h integrável. Então $g * h$ é diferenciável até a ordem r com derivadas de j -ésima ordem limitadas dadas por*

$$(g * h)^{(j)} = g^{(j)} * h, \text{ para } j = 1, \dots, r$$

*Além disso, se $g^{(j)}$ e h são de quadrado integrável então a derivada $(g * h)^{(j)}$ é contínua em \mathbb{R} .*

Proposição A.2. *Se g tem derivada limitada em \mathbb{R} e $g, g' \in L_2$ então $\lim_{|x| \rightarrow \infty} g(x) = 0$ e $\int g g' = 0$.*

Precisamos ainda de mais dois resultados auxiliares.

Proposição A.3. *Para $s, r \in \mathbb{N}$ com $s \geq r$, assumindo que $f^{(r)}, \dots, f^{(r-s)}$ e $g, g', \dots, g^{(s)}$ são limitadas e de quadrado integrável então*

$$\int f^{(r)}(x) g(x) dx = (-1)^s \int g^{(s)}(x) f^{(r-s)}(x).$$

Anexo B

Códigos em R

```
# Capítulo 3
#Densidades de Marron and Wand

mwweights = function(d){
  switch(d,
    "1"=c(1),
    "2"=c(0.2,0.2,0.6),
    "3"=rep(0.125,times=8),
    "4"=c(2/3,1/3),
    "5"=c(0.1,0.9),
    "6"=c(0.5,0.5),
    "7"=c(0.5,0.5),
    "8"=c(0.75,0.25),
    "9"=c(0.45,0.45,0.1),
    "10"=c(0.5,rep(0.1,times=5)),
    "11"=c(0.49,0.49,rep(1/350,times=7)),
    "12"=c(0.5,8/31,4/31,2/31,1/31,1/62),
    "13"=c(rep(0.46,times=2),rep(1/300,times=3),rep(7/300,times=3)),
    "14"=c(32/63,16/63,8/63,4/63,2/63,1/63),
    "15"=c(rep(2/7,times=3),rep(1/21,times=3)),
    "16"=c(rep(1/10,times=10)))}

mwmeans = function(d){
  switch(d,
    "1"=c(0),
    "2"=c(0,0.5,13/12),
    "3"=c(0,-1,-5/3,-19/9,-65/27,-211/81,-665/243,-2059/729),
    "4"=c(0,0),
    "5"=c(0,0),
    "6"=c(-1,1),
    "7"=c(-3/2,3/2),
    "8"=c(0,3/2),
    "9"=c(-1.2,1.2,0),
    "10"=c(0,-1,seq(-0.5,1,0.5)),
    "11"=c(-1,1,seq(-1.5,1.5,0.5)),
    "12"=c(0,seq(-1.5,2.5,1)),
    "13"=c(-1,1,-0.5,-1,-1.5,0.5,1,1.5),
    "14"=c(-31/21,17/21,41/21,53/21,59/21,62/21),
    "15"=c(-15/7,-3/7,9/7,16/7,18/7,20/7),
    "16"=c(5,15,25,35,45,55,65,75,85,95))}
```

```

mwsdeviations = function(d){
switch(d,
"1"=c(1),
"2"=c(1,2/3,5/9),
"3"=c(1,2/3,4/9,8/27,16/81,32/243,64/729,128/2187),
"4"=c(1,0.1),
"5"=c(1,0.1),
"6"=c(2/3,2/3),
"7"=c(0.5,0.5),
"8"=c(1,1/3),
"9"=c(0.6,0.6,0.25),
"10"=c(1,rep(0.1,times=5)),
"11"=c(2/3,2/3,rep(0.01,times=7)),
"12"=c(1,0.4,0.2,0.1,0.05,0.025),
"13"=c(rep(2/3,times=2),rep(0.01,times=3),rep(0.07,times=3)),
"14"=c(32/63,16/63,8/63,4/63,2/63,1/63),
"15"=c(rep(2/7,times=3),rep(1/21,times=3)),
"16"=c(rep(1,times=10)))}

#####
#Densidade de probabilidade de Mistura de p normais
#####
#p-vector de pesos w_1,w_2,...,w_p (w_1+...+w_p=1)
#med-vector de médias
#sd-vector de desvios-padrão
dmnorm<-function(x,w,med,sd){
dimw<-length(w)
dimx<-length(x)
fd<-array(dim=c(dimw,dimx))
for (i in 1:dimw) {fd[i,]<-w[i]*dnorm((x-med[i])/sd[i])/sd[i]}
colSums(fd)}

#gráficos das misturas de densidade
grafico = function(dist){
if (dist<10) di<-paste(0,dist,sep="") else di<-dist
pdf(paste("MWgrafico",di,".pdf",sep=""),width=5.5,height=5)
par(mar=c(2,2.2,3,1)) #parametros - baixo, esquerda, cima, direita
titulo <- paste("Distribuição MW #",dist,sep="")
x <- seq(-3.1,3.1,0.005)
y <- dmnorm(x,mwweights(dist),mwmeans(dist),mwsdeviations(dist))
plot(x,y,type="l",lty=1,lwd=2,xlab="",ylab="",xlim=c(-3.1,3.1),
ylim=c(0,max(y)*1.01),cex.axis=1.5,main=titulo,cex.main=1.5)
dev.off()}

for (i in 1:15) grafico(i)

#polinomio de Hermite de ordem r
Hermite = function(r,x){
if (r==0) aux <- rep(1,times=length(x))
else if (r==1) aux <- x else{
aux1 <- 1
aux2 <- x
for (j in 2:r){

```

```

aux <- x*aux2-(j-1)*aux1
aux1 <- aux2
aux2 <- aux}]
return(aux)}

# Derivada de ordem r da densidade normal standard
Deriv.dnorm<-function(r,x){
(-1)^r*Hermite(r,x)*dnorm(x,0,1)}

# Valor do psi_r = \int f^(r)(x) f(x) dx = (-1)^(r/2) \int {f^(r/2)(x)}^2 dx,
r=0,2,4,6 when f is a mixture of normal densities
psir.mixt = function(r,mu,sig2,w){
k <- length(w)
SS <- matrix(rep(sig2,rep(k,k)),ncol=k,byrow=TRUE)
SS <- SS+t(SS)
MU <- matrix(rep(mu,rep(k,k)),ncol=k,byrow=TRUE)
MU <- MU-t(MU)
A <- SS^(-(r+1)/2)*sapply(MU/sqrt(SS),Hermite,r=r)*dnorm(MU/sqrt(SS))
drop(t(w)%*A%*%w)}

#Função U'(h;s=0,q) do Teorema 5.1 de Marron e Wand (1992)
Deriv.U = function(h,q,w,mu,sig2){
difmeans <- outer(mu, mu, "-")
desp <- sqrt(outer(sig2, sig2, "+") + q*h^2)
A <- q*h*Deriv.dnorm(r=2,x=difmeans/desp)/desp^3
return(drop(t(w)%*A%*%w))}

#Função MISE'(h) do Teorema 5.1 de Marron e Wand (1992)
Deriv.MISE = function(h,n,w,mu,sig2){
-1/(2*sqrt(pi)*n*h^2)+ (1-1/n)*Deriv.U(h,q=2,w,mu,sig2)-2*Deriv.U(h,q=1,w,mu,sig2)}

h0 = function(n,w,mu,sig2,d=1,lower=0.001,upper=2){
if(missing(mu)||missing(sig2)||missing(w)) {
w <- mwweights(d)
mu <- mwmeans(d)
sig2 <- mwsdeviations(d)^2}
tol = 0.1*lower
iter <- 1
while( Deriv.MISE(h=lower,n,w,mu,sig2)*Deriv.MISE(h=upper,n,w,mu,sig2)>0 ) {
if (iter > 99) stop("no solution in the specified range of bandwidths")
upper <- upper * 1.2
lower <- lower/1.2
iter<- iter+1}
uniroot(Deriv.MISE,c(lower, upper),n=n,w=w,mu=mu,sig2=sig2,tol=tol)$root}

g <- function(u){dnorm(u)^2}
RK <- integrate(g,lower=-Inf, upper=+Inf)$value
g1 <- function(u){u^2*dnorm(u)}
mu2K <- integrate(g1,lower=-Inf, upper=+Inf)$value
g2 <- function(u){u^4*dnorm(u)}
mu4K <- integrate(g2,lower=-Inf, upper=+Inf)$value

h1 = function(n,d=1){
w <- mwweights(d)

```

```

mu <- mwmeans(d)
sig2 <- mwsdeviations(d)^2
RK^(1/5)*mu2K^(-2/5)*psir.mixt(r=4,mu,sig2,w)^(-1/5)*n^(-1/5)}

h2 = function(n,d=1){
w <- mwweights(d)
mu <- mwmeans(d)
sig2 <- mwsdeviations(d)^2
h1(n,d) + (1/20)*RK^(3/5)*mu2K^(-22/5)*mu4K*psir.mixt(r=4,mu,sig2,w)^(-8/5)
*abs(psir.mixt(r=6,mu,sig2,w))*n^(-3/5)}

graficoh = function(dist){
if (dist<10) di<-paste(0,dist,sep="") else di<-dist
pdf(paste("hMWgrafico",di,".pdf",sep=""),width=5.5,height=5) #,horizontal = FALSE)
par(mar=c(5,5,3,1)) #parametros - baixo, esquerda, cima, direita
titulo <- paste("Distribuição MW #",dist,sep="")
xs <- 2:6
ns <- 10^(xs)
h0s <- array(dim=length(ns))
for (j in 1:length(ns)) h0s[j] <- h0(n=ns[j],d=dist)
y1 <- h1(n=ns,d=dist)/h0s
y2 <- h2(n=ns,d=dist)/h0s
plot(xs,y1,type="b",lty=1,lwd=2,xlab=expression(log[10](n)),
ylab=expression(h/h[list(0,n)]),ylim=c(min(c(y1,y2))*0.95,max(c(y1,y2))*1.05),
cex.lab=1.5,cex.axis=1.5,main=titulo,cex.main=1.5,xaxt = "n")
lines(xs,y2,type="b",lty=5,lwd=2)
abline(h=1,lty=3,lwd=3)
axis(1, at=xs, labels=xs,cex.axis=1.5)
legend("bottom",legend=c(expression(h==h[list(1,n)]),expression(h==h[list(2,n)]))
,lty=c(1,5),lwd=2,bty="n",cex=1.25)
#legend("bottom",leg=c(1,2),lty=c(1,5),lwd=2,bty="n")
dev.off()
}

#Estudo do Capítulo 5
library(ks)

#usar mwweights, mwmeans, mwsdeviations

rmnorm = function(n,w,med,sd){
dimw<-length(w)
alf<-array(1:dimw)
if (dimw==1) soma<-rnorm(n,med,sd) else{
alf[1]<-w[1]
for (i in 2:dimw) {
alf[i]<-alf[i-1]+w[i]}
unive<-runif(n,0,1)
norve<-rnorm(n,0,1)
soma<-array(1:n)
pp<-1*(unive<alf[1])
soma<-(pp*norve)*sd[1]+pp*med[1]

for (i in 2:dimw){
pp<-1*(unive<alf[i])*(unive>=alf[i-1])

```

```

soma<-soma+(pp*norve)*sd[i]+pp*med[i]}}
soma}

#Simulação das distribuições de Marron and Wand (1992)
rmw = function(n,d){
rmnorm(n,mwweights(d),mwmeans(d),mwsdeviations(d))}

#Calculo de R(f) para f mistura de normais
#for f=\sum_{i=1}^k w_i dnorm(x,m_i,s_i) with w_1+\dots + w_k=1,

L2 = function(w,m,s){
difmeans <- outer(m, m, "-")
sig2 <- s^2
sumvar <- outer(sig2, sig2, "+")
desp <- sqrt(sumvar)
A <- dnorm(difmeans/desp)/desp
drop(t(w)%*%A%*%w)}

L2f = function(d){
L2(mwweights(d),mwmeans(d),mwsdeviations(d))}

ISE = function(x,d,h){
n <- length(x)
w <- c(mwweights(d),rep(-1/n,times=n))
m <- c(mwmeans(d),x)
s <- c(mwsdeviations(d),rep(h,times=n))
L2(w,m,s)}

g <- function(u){dnorm(u)^2}
RK <- integrate(g,lower=-Inf, upper=+Inf)$value
mu2K <- 1
g2 <- function(u){u^4*dnorm(u)}
mu4K <- 3

#Escolha de m
ISES = function(d,n,m.max=50,rep=500){
if (d<10) di <- paste(0,d,sep="") else di <- d
if (n<100) ni <- paste(0,n,sep="") else ni <- n

escrever <- paste("MWd",di,"n",ni,".pdf",sep="")
ISEm <- array(dim=c(rep,m.max))

for (j in 1:rep){
x <- rmw(n,d)
hk <- matrix(nc=n,nr=m.max+1)
hk[1,] <- (pi)^(-1/4)*exp(-x^2/2)
hk[2,] <- sqrt(2)*x*hk[1,]
for (i in 2:m.max)
hk[i+1,] <- sqrt(2/i)*x*hk[i,]-sqrt((i-1)/i)*hk[i-1,]

dev.h <- matrix(nc=n, nr=m.max+1)
dev.h[1,] <- -x*(pi)^(-1/4)*exp(-x^2/2)
for (i in 2:(m.max+1))
dev.h[i,]<- sqrt(2*(i-1))*hk[i-1,]-x*hk[i,]

```

```

dev2.h <- matrix(nc=n, nr=m.max+1)
dev2.h[1,] <- (pi)^(-1/4)*exp(-x^2/2)*(x^2-1)
for (i in 2:(m.max+1))
dev2.h[i,]<-sqrt(2*(i-1))*dev.h[i-1,]-x*dev.h[i,]-hk[i,]

est.a2k <- rowMeans(dev2.h)

for (i in 1:m.max){
R12m <- sum(est.a2k[1:(i+1)]^2)
h1m <- RK^(1/5)*mu2K^(-2/5)*R12m^(-1/5)*n^(-1/5)
ISEm[j,i] <- ISE(x,d,h=h1m)}
}
write.table(ISEm,file=escrever)}

grafbox = function(di){
if (d<10) di <- paste(0,d,sep="") else di <- d
if (n<100) ni <- paste(0,n,sep="") else ni <- n
ises <- read.table(paste("MWd",di,"n",ni,".pdf",sep=""))
pdf(paste("MWboxplot",di,"n", ni,".pdf",sep=""),width=5.5,height=5)
boxplot(ises, main=paste("Distribuição MW #",di,sep=""), xlab="m", ylab="ISE",
names=1:100)
dev.off()}

ISE_h = function(d,n,rep=500){
set.seed(989865)
if (d<10) di <- paste(0,d,sep="") else di <- d
if (n<100) ni <- paste(0,n,sep="") else ni <- n
escrever1 <- paste("MW","d",di,"n",ni,"valoresdem1",".txt",sep="")
escrever2 <-paste("MW","d",di,"n",ni,"hs",".txt",sep="")
escrever3 <- paste("MW","d",di,"n",ni,"ISE",".txt",sep="")
escrever4 <- paste("MW","d",di,"n",ni,"valoresdem2",".txt",sep="")
m.max <- round(20*n^(1/6))
hk <- matrix(nc=n,nr=m.max+1)
Hnm1 <- array(dim=m.max+1)
Hnm2 <- array(dim=m.max+1)
n_est <- 10
hs<- array(dim=c(rep,n_est))
ISEs <- array(dim=c(rep,n_est))
ms1 <- array(dim=rep)
ms2 <- array(dim=rep)
for (j in 1:rep){
x <- rmw(n,d)
hk[1,] <- (pi)^(-1/4)*exp(-x^2/2)
hk[2,] <- sqrt(2)*x*hk[1,]
for (i in 2:m.max)
hk[i+1,] <- sqrt(2/i)*x*hk[i,]-sqrt((i-1)/i)*hk[i-1,]

#Soma dos hk^2 variando o x
soma.hk2.x <- rowMeans(hk^2)
#Soma dos hk variando o x
soma.hk.x <- rowMeans(hk)

Hnm1[1] <- soma.hk2.x[1]/n - 0.5*(n+1)*soma.hk.x[1]^2/n

```

```

for (i in 2:(m.max+1))
Hnm1[i] <- Hnm1[i-1] + soma.hk2.x[i]/n - 0.5*(n+1)*soma.hk.x[i]^2/n

Hnm2[1] <- soma.hk2.x[1]/n
- 0.5*(n+1)*soma.hk.x[1]^2/(n-1)+0.5*(n+1)*soma.hk2.x[1]/n^2
for (i in 2:(m.max+1))
Hnm2[i] <- Hnm2[i-1] + soma.hk2.x[i]/n
- 0.5*(n+1)*soma.hk.x[i]^2/(n-1)+0.5*(n+1)*soma.hk2.x[i]/n^2

m1 <- which.min(Hnm1)-1
ms1[j] <- m1
m2 <- which.min(Hnm2)-1
ms2[j] <- m2

dev.h_m1 <- matrix(nc=n, nr=m1+1)
dev.h_m1[1,] <- -x*(pi)^(-1/4)*exp(-x^2/2)
if (m1>0)
for (i in 2:(m1+1))
dev.h_m1[i,]<- sqrt(2*(i-1))*hk[i-1,]-x*hk[i,]

dev.h_m2 <- matrix(nc=n, nr=m2+1)
dev.h_m2[1,] <- -x*(pi)^(-1/4)*exp(-x^2/2)
if (m2>0)
for (i in 2:(m2+1))
dev.h_m2[i,]<- sqrt(2*(i-1))*hk[i-1,]-x*hk[i,]

dev2.h_m1 <- matrix(nc=n, nr=m1+1)
dev2.h_m1[1,] <- (pi)^(-1/4)*exp(-x^2/2)*(x^2-1)
if (m1>0)
for (i in 2:(m1+1))
dev2.h_m1[i,]<-sqrt(2*(i-1))*dev.h_m1[i-1,]-x*dev.h_m1[i,]-hk[i,]

dev2.h_m2 <- matrix(nc=n, nr=m2+1)
dev2.h_m2[1,] <- (pi)^(-1/4)*exp(-x^2/2)*(x^2-1)
if (m2>0)
for (i in 2:(m2+1))
dev2.h_m2[i,]<-sqrt(2*(i-1))*dev.h_m2[i-1,]-x*dev.h_m2[i,]-hk[i,]

dev3.h_m1 <- matrix(nc=n, nr=m1+1)
dev3.h_m1[1,] <- (pi)^(-1/4)*exp(-x^2/2)*(3*x-x^3)
if (m1>0)
for (i in 2:(m1+1))
dev3.h_m1[i,]<-sqrt(2*(i-1))*dev2.h_m1[i-1,]-x*dev2.h_m1[i,]-2*dev.h_m1[i,]

dev3.h_m2 <- matrix(nc=n, nr=m2+1)
dev3.h_m2[1,] <- (pi)^(-1/4)*exp(-x^2/2)*(3*x-x^3)
if (m2>0)
for (i in 2:(m2+1))
dev3.h_m2[i,]<-sqrt(2*(i-1))*dev2.h_m2[i-1,]-x*dev2.h_m2[i,]-2*dev.h_m2[i,]

est.a2k <- rowMeans(dev2.h_m1)
R12m <- sum(est.a2k^2)
est.a2kquad <- (n/(n-1))*(est.a2k^2-(1/n)*rowMeans(dev2.h_m1^2))
R22m <- sum(est.a2kquad)

```

```

if (R22m > 0) R_2m <- R22m else R_2m <- R12m

est.a2k.2 <- rowMeans(dev2.h_m2)
R12m2 <- sum(est.a2k.2^2)
est.a2kquad.2 <- (n/(n-1))*(est.a2k.2^2-(1/n)*rowMeans(dev2.h_m2^2))
R22m2 <- sum(est.a2kquad.2)
if (R22m2 > 0) R_2m2 <- R22m2 else R_2m2 <- R12m2

est.a3k <- -rowMeans(dev3.h_m1)
R13m <- sum(est.a3k^2)
est.a3kquad <- (n/(n-1))*(est.a3k^2-(1/n)*rowMeans(dev3.h_m1^2))
R23m <- sum(est.a3kquad)
if (R23m > 0) R_3m <- R23m else R_3m <- R13m

est.a3k.2 <- -rowMeans(dev3.h_m2)
R13m2 <- sum(est.a3k.2^2)
est.a3kquad.2 <- (n/(n-1))*(est.a3k.2^2-(1/n)*rowMeans(dev3.h_m2^2))
R23m2 <- sum(est.a3kquad.2)
if (R23m2 > 0) R_3m2 <- R23m2 else R_3m2 <- R13m

# hm is based on the maximal smoothing principle by Terrel (1990)
ck <- (RK/mu2K^2)^(1/5)
hm <- max(1.47*sd(x),0.98*IQR(x))*ck*n^(-1/5)

h1a1 <- RK^(1/5)*mu2K^(-2/5)*R12m^(-1/5)*n^(-1/5)
h1b1 <- RK^(1/5)*mu2K^(-2/5)*R_2m^(-1/5)*n^(-1/5)
h1a2 <- RK^(1/5)*mu2K^(-2/5)*R12m2^(-1/5)*n^(-1/5)
h1b2 <- RK^(1/5)*mu2K^(-2/5)*R_2m2^(-1/5)*n^(-1/5)
#if (h1b > 2*hm) h1b <- h1a

ISEs[j,1] <- ISE(x,d,h=h1a1)
ISEs[j,2] <- ISE(x,d,h=h1b1)
ISEs[j,3] <- ISE(x,d,h=h1a2)
ISEs[j,4] <- ISE(x,d,h=h1b2)
hs[j,1] <- h1a1
hs[j,2] <- h1b1
hs[j,3] <- h1a2
hs[j,4] <- h1b2

h2a1 <- h1a1 + (1/20)*RK^(3/5)*mu2K^(-22/5)*mu4K*R12m^(-8/5)*R13m*n^(-3/5)
h2b1 <- h1b1 + (1/20)*RK^(3/5)*mu2K^(-22/5)*mu4K*R_2m^(-8/5)*R_3m*n^(-3/5)
h2a2 <- h1a2 + (1/20)*RK^(3/5)*mu2K^(-22/5)*mu4K*R12m2^(-8/5)*R13m2*n^(-3/5)
h2b2 <- h1b2 + (1/20)*RK^(3/5)*mu2K^(-22/5)*mu4K*R_2m2^(-8/5)*R_3m2*n^(-3/5)

ISEs[j,5] <- ISE(x,d,h=h2a1)
ISEs[j,6] <- ISE(x,d,h=h2b1)
ISEs[j,7] <- ISE(x,d,h=h2a2)
ISEs[j,8] <- ISE(x,d,h=h2b2)
hs[j,5] <- h2a1
hs[j,6] <- h2b1
hs[j,7] <- h2a2
hs[j,8] <- h2b2

```

```

H_PI<- hpi(x, nstage=2, binned=TRUE, deriv.order=0)
ISEs[j,9]<- ISE(x,d,h=H_PI)
hs[j,9] <- H_PI

HCVs<- hucv(x)
ISEs[j,10] <- ISE(x,d,h=HCVs)
hs[j,10] <- HCVs

}

write.table(ms1, file=escrever1)
write.table(ms2, file=escrever4)
write.table(hs, file=escrever2)
write.table(ISEs,file=escrever3)
}
for (i in c(1,2,3,6,7,8,9,12)) for (j in c(50,100,150,200)) ISE_h(d=i,n=j,rep=500)

graphics <- function(d,met=c(3,4,7,8))
{
if (d<10) di <- paste(0,d,sep="") else di<-d

ises050 <- read.table(paste("MWd",di,"n050ISE.txt",sep=""))
ises100 <- read.table(paste("MWd",di,"n100ISE.txt",sep=""))
ises150 <- read.table(paste("MWd",di,"n150ISE.txt",sep=""))
ises200 <- read.table(paste("MWd",di,"n200ISE.txt",sep=""))
L2.ises050 <- sqrt(colMeans(ises050^2))
L2.ises100 <- sqrt(colMeans(ises100^2))
L2.ises150 <- sqrt(colMeans(ises150^2))
L2.ises200 <- sqrt(colMeans(ises200^2))

maximo <- max(L2.ises050[c(met,9,10)],L2.ises100[c(met,9,10)],
L2.ises150[c(met,9,10)], L2.ises200[c(met,9,10)])
minimo <- min(L2.ises050[c(met,9,10)],L2.ises100[c(met,9,10)],
L2.ises150[c(met,9,10)], L2.ises200[c(met,9,10)])

L2.PI <- c(L2.ises050[9],L2.ises100[9],L2.ises150[9],L2.ises200[9])
L2.CV <- c(L2.ises050[10],L2.ises100[10],L2.ises150[10],L2.ises200[10])
L2.01 <- c(L2.ises050[met[1]],L2.ises100[met[1]],L2.ises150[met[1]],
L2.ises200[met[1]])
L2.02 <- c(L2.ises050[met[2]],L2.ises100[met[2]],L2.ises150[met[2]],
L2.ises200[met[2]])
L2.03 <- c(L2.ises050[met[3]],L2.ises100[met[3]],L2.ises150[met[3]],
L2.ises200[met[3]])
L2.04 <- c(L2.ises050[met[4]],L2.ises100[met[4]],L2.ises150[met[4]],
L2.ises200[met[4]])

n <- c(50,100,150,200)
pdf(paste("GMW",di,".pdf",sep=""),width=5.5,height=5) # ,horizontal = FALSE)
par(mar=c(4,5,2,1)) #parametros - baixo, esquerda, cima, direita
plot(1:length(n),L2.01,type="b",lwd=1.5,lty=1,
main=paste("Distribuição MW #",di,sep=""),cex.main=1.5,
xlab="tamanho da amostra",ylab=expression(n*o*r*m*a~L[2]~d*o~I*S*E),axes=FALSE,
cex.lab=1.3,cex.axis=1.3,ylim=c(minimo,maximo))
axis(1,at=1:length(n),labels=n,cex.axis=1.5)

```

```
axis(2,cex.axis=1.5)
lines(1:length(n),L2.O2,type="b",lwd=1.5,lty=5)
lines(1:length(n),L2.O3,type="b",lwd=3,lty=3)
lines(1:length(n),L2.O4,type="b",lwd=1.5,lty=2)
lines(1:length(n),L2.CV,type="b",lwd=1.5,lty=4,col="red")
lines(1:length(n),L2.PI,type="b",lwd=1.5,lty=6,col="blue")
legend("topright",leg=c(expression(hat(h)[1*","*1*","*n]),
expression(hat(h)[1*","*2*","*n]),expression(hat(h)[2*","*1*","*n]),
expression(hat(h)[2*","*2*","*n]),
expression(h[c*v]),expression(h[p*i])),
lty=c(1,5,3,2,4,6),cex=1.5,lwd=c(1.5,1.5,3,1.5,1.5,1.5),bty="n",
col=c("black","black","black","black","red","blue"))
dev.off()
}

for (i in 1:15) graphics(d=i,c(3,4,7,8))
```

Bibliografía

- Bogachev, V.I. (2007). *Measure Theory*. Springer-Verlag.
- Bosq, D., Lecoutre, J.P. (1978). *Theorie de l'estimation fonctionnelle*. Economica
- Butzer, P.L., Nessel, R.J. (1971). *Fourier analysis and approximation V.1: one dimensional theory*. Birkhauser
- Chacón, J.E. (2004). *Estimación de densidades: algunos resultados exactos y asintóticos*. PhD dissertation. Universidad de Extremadura.
- Chacón, J.E., Montanero, J., Nogales, A.G., Pérez, P. (2007). On the existence and limit behavior of the optimal bandwidth in kernel density estimation. *Statist. Sinica* 17, 289–300.
- Chiu, S.-T. (1991). Bandwidth selection for kernel density estimation. *The Annals of Statistics* 19, 1883–1905.
- Cohn, D.L. (1980). *Measure theory*. Birkhauser.
- Greblicki, W., Pawlak, M. (1984). Hermite series estimates of a probability density and its derivatives. *Journal of multivariate analysis*. 15, 174–182.
- Hall, P. (1983). Large samples optimality of least squares cross-validation in density estimation. *Ann. Statist.* 11, 1156–1174.
- Hart, J.D. (1985). On the choice of truncation point in fourier series density estimation. *Journal of statistical computation and simulation*. 21, 95–116.
- Hall, P., Marron, J.S. (1987). Extent to which least-squares cross-validation minimizes integrated square error in nonparametric density estimation. *Probability theory and related fields*. 74, 567–581.
- Hall, P., Sheather, S.J., Jones, M.C., Marron, J.S. (1991). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*. 78, 263–269.
- Jones, M.C., Sheather, S.J. (1991) Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statistics & Probability Letters*. 11, 511–514.
- Lee, A. J. (1990). *U-statistics*. CRC Press.
- Lima, E.L. (1995). *Curso de análise, Vol. I*. IMPA.
- Marron, J.S., Wand, M.P. (1992). Exact mean integrated squared error. *Ann. Statist.* 20, 712–736.
- Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* 33, 1065–1076.
- Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. *Ann. Math. Statist.* 27, 832–837.

- Rudin, W. (1990). *Fourier analysis on groups*. New York: Wiley.
- Schwartz, S. C. (1967). Estimation of a probability density by an orthogonal series. *Ann. Math. Statist.* 27, 1261–1265.
- Szegő, G. (1939). *Orthogonal Polynomials*. Amer. Math. Soc. Coll. Publ., Providence, R.I.
- Tenreiro, C. (2011). Fourier series-based direct plug-in bandwidth selectors for kernel density estimation. *Journal of Nonparametric Statistics*. 23, 533–545
- Walter, G. G. (1977). Properties of Hermite series estimation of probability density. *Ann. Statist.* 5, 1258–1264
- Wand, M. P., Jones, M. C. (1995). *Kernel smoothing*. London: Chapman & Hall.