

1 2 9 0



UNIVERSIDADE D
COIMBRA

João Nuno Campos Nunes de Freitas

**ESTIMAÇÃO NÃO PARAMÉTRICA DE UMA
FUNÇÃO DE PROBABILIDADE PELO MÉTODO
DO NÚCLEO**

**Dissertação no âmbito do Mestrado em Matemática, Ramo de Estatística,
Otimização e Matemática Financeira orientada pelos Professores Doutores Ana
Cristina Martins Rosa e Carlos Manuel Rebelo Tenreiro da Cruz e apresentada ao
Departamento de Matemática da Faculdade de Ciências e Tecnologia.**

Junho de 2022

Nonparametric Estimation of a Probability Mass Function by the Kernel Method

João Nuno Campos Nunes de Freitas



UNIVERSIDADE D
COIMBRA

Mestrado em Matemática

Master in Mathematics

Dissertação de Mestrado | MSc Dissertation

Junho 2022 | June 2022

Agradecimentos

Aos Professores Doutores Ana Cristina Martins Rosa e Carlos Manuel Rebelo Tenreiro da Cruz pela orientação, disponibilidade, perseverança e destreza em todos os momentos,
À minha Mãe e ao meu Pai pelo apoio incondicional.

Resumo

A presente dissertação é dedicada à estimação não paramétrica de uma função de probabilidade desconhecida, por intermédio do método do núcleo. Começamos por definir o estimador e exibir alguns exemplos típicos de núcleos, nomeadamente os núcleos de Dirac, de Aitchison e Aitken, de Wang e Van Ryzin, a versão simétrica do núcleo triangular clássico, bem como uma proposta mais recente baseada na distribuição de Conway-Maxwell-Poisson. Segue-se a apresentação das propriedades básicas de convergência local dos estimadores não normalizado e normalizado. No que diz respeito ao primeiro, são analisados os comportamentos assintóticos do viés, da variância e do erro quadrático médio, sendo depois estabelecida a convergência quase certa e a normalidade assintótica do estimador. Relativamente ao estimador normalizado, a dedução dos resultados de convergência em probabilidade e quase certa passa por analisar o comportamento assintótico da sucessão das constantes de normalização, cujo estudo é feito previamente. Posteriormente, apresentamos as propriedades globais de ambos os estimadores, com ênfase para a norma ℓ_1 . No caso do estimador normalizado, a convergência quase certa no sentido desta norma permitirá obter um resultado do tipo Glivenko-Cantelli para a correspondente função de repartição. Ainda no âmbito das propriedades globais, é obtido o desenvolvimento assintótico do erro quadrático médio integrado para o estimador não normalizado usando a expansão de Newton, o que constitui uma abordagem original, tanto quanto sabemos. Por fim, comparamos os desempenhos dos estimadores empírico e de Aitchison e Aitken, a distância finita, através de um breve estudo de simulação.

Conteúdo

Lista de Figuras	ix
1 Introdução	1
2 Definições e exemplos	3
2.1 Definição de núcleo	3
2.2 Exemplos de núcleos	4
2.3 Definição de estimador do núcleo	10
3 Propriedades locais de convergência	13
3.1 Estimador não normalizado	13
3.2 Estimador normalizado	19
3.2.1 Convergência em probabilidade	19
3.2.2 Convergência quase certa	21
4 Propriedades globais de convergência	23
4.1 Convergência ℓ_1	24
4.1.1 Estimador não normalizado	24
4.1.2 Estimador normalizado	27
4.2 Desenvolvimento assintótico do erro quadrático médio integrado	29
5 O estimador de Aitchison e Aitken	41
5.1 Janela ótima do estimador	41
5.2 Escolha prática da janela	43
5.3 Alguns resultados de simulação	44
6 Considerações finais	51
Bibliografia	53
Apêndice A Códigos R	57

Lista de Figuras

2.1	Configurações gráficas do núcleo apresentado no Exemplo 3.	7
2.2	Configurações gráficas do núcleo apresentado no Exemplo 4.	8
2.3	Configurações gráficas do núcleo apresentado no Exemplo 5.	11
5.1	Representações gráficas das funções de probabilidade usadas nos cenários em estudo.	45
5.2	Diagramas de extremos e quartis das distribuições dos EQI com base em amostras com distribuição no cenário 1.	46
5.3	Diagramas de extremos e quartis das distribuições dos EQI com base em amostras com distribuição no cenário 2.	47
5.4	Diagramas de extremos e quartis das distribuições dos EQI com base em amostras com distribuição no cenário 3.	48
5.5	Diagramas de extremos e quartis das distribuições dos EQI com base em amostras com distribuição no cenário 4.	49
5.6	Aproximações da norma- L_2 do EQI perante os vários cenários em estudo.	50

Capítulo 1

Introdução

A estimação não paramétrica de uma função densidade de probabilidade mereceu, nas últimas décadas, o interesse de inúmeros autores, tendo sido abordada através de vários métodos e diferentes perspectivas. De entre os métodos existentes, destaca-se o denominado método do núcleo, introduzido por Parzen [30] e Rosenblatt [34], cujas simplicidade e eficiência justificam a sua popularidade, quer a nível teórico, quer a nível das aplicações nas mais diversas áreas do conhecimento (vejam-se, por exemplo, as obras de Silverman [38], Wand e Jones [20], Bosq e Lecoutre [6], Simonoff [39] e, mais recentemente, Tenreiro [40], Gramacki [13]).

Em contrapartida, a atenção dedicada ao método do núcleo para estimar uma função de probabilidade tem sido incomparavelmente menor, o que talvez se deva à ausência de uma ideia simples e intuitiva de suavização (*smoothing*) discreta. No entanto, existem na literatura algumas propostas neste sentido, tais como os trabalhos de Aitchison e Aitken [3], Wang e Van Ryzin [43], Titterington [41], Rajagopalan e Lall [33], Titterington e Bowman [42], Burman [8], Ahmad e Cerrito [2], Dong e Simonoff [11], Hall e Titterington [14] e Ouyang et al. [28]. Mas, salvo raras exceções, essas propostas reduzem-se a uma mera discretização do caso absolutamente contínuo, mesmo quando as variáveis subjacentes são ordinais.

No que diz respeito a dados de natureza intrinsecamente discreta, encontramos uma abordagem mais coerente nos artigos de Kokonendji e Zocchi [26], Kokonendji e Abdous [1], Kokonendji et al. [23], Kokonendji e Senga Kiessé [24], Racine et al. [32], Chu et al. [9] e Harfouche et al. [15].

Naturalmente, tendo em mente a ideia de suavização subjacente ao método do núcleo no caso absolutamente contínuo (assente em propriedades de regularidade da função a estimar), a primeira questão que se coloca prende-se com a motivação e a utilidade desta ideia no caso discreto. Com efeito, é sabido que, neste contexto, o estimador não paramétrico de referência, definido por

$$\hat{f}_E(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{x\}}(X_i),$$

onde (X_1, \dots, X_n) denota uma amostra aleatória da variável em causa, constitui a via mais natural para estimar uma função de probabilidade sobre \mathbb{R} , além de possuir boas propriedades teóricas. Contudo, quando estamos em presença de dados esparsos (i.e., quando a dimensão da amostra é pequena comparativamente com o número de pontos do suporte da variável), as estimativas assim construídas podem atribuir frequência nula (ou praticamente nula) a valores da função de probabilidade com uma

ordem de grandeza não desprezável (cf. Hall e Titterington [14], Racine et al. [32] e Huang et al. [18], por exemplo). Uma forma de contornar esta dificuldade passa por construir um estimador que atribua frequência positiva não apenas aos valores observados mas também a pontos próximos desses valores, o que pode ser interpretado como uma técnica de suavização. O método do núcleo é, pois, uma forma de concretizar tal objetivo já que corresponde a uma versão suavizada da função de probabilidade empírica (através da janela h_n) e para ela converge quando a dimensão da amostra aumenta. Acresce ainda que a sua utilização é recomendada noutros domínios tais como a estimação da menor região de cobertura¹ de uma lei de probabilidade discreta (cf. O'Neill [29]).

À semelhança do caso contínuo, foram estabelecidos, nos artigos referidos no terceiro parágrafo, resultados de convergência pontual e global dos correspondentes estimadores, bem como propriedades relativas à sua distribuição assintótica ou à escolha da janela. O presente trabalho é dedicado ao estudo desta classe de estimadores, com a apresentação das suas propriedades fundamentais. Mais precisamente, no Capítulo 2, é introduzido o estimador e exibidas algumas classes usuais de núcleos (designados, na literatura, por núcleos associados discretos), nomeadamente os núcleos de Dirac, de Aitchison e Aitken, de Wang e Van Ryzin, a versão simétrica do núcleo triangular clássico, bem como um exemplo mais recente inspirado na distribuição de Conway-Maxwell-Poisson. O Capítulo 3 começa por estabelecer as propriedades básicas de convergência local do estimador não normalizado, nomeadamente o comportamento assintótico do viés, da variância e do erro quadrático médio. São depois abordadas a convergência quase certa e a normalidade assintótica do estimador. Seguidamente, estudamos as propriedades assintóticas da sucessão formada pelas constantes de normalização, que permitem deduzir a convergência em probabilidade e quase certa do estimador normalizado. O Capítulo 4 aborda as propriedades globais de ambos os estimadores. A primeira parte inicia-se com a obtenção das convergências em probabilidade e quase certa no sentido da norma ℓ_1 e termina com um resultado do tipo Glivenko-Cantelli relativo à função de repartição associada ao estimador normalizado. A segunda parte é inteiramente dedicada ao desenvolvimento assintótico do erro quadrático médio no caso do estimador não normalizado. No Capítulo 5, apresentamos os resultados obtidos num breve estudo de simulação que visa comparar o comportamento a distância finita do estimador de Aitchison e Aitken com o do estimador empírico, através das distribuições dos erros quadráticos integrados empíricos.

Por fim, gostaríamos de referir que, apesar de não estar explicitamente referido no texto, os resultados que apresentamos nos Teoremas 3, 6 e 8 relativos à convergência quase certa, são também válidos para a convergência quase completa.

¹Termo adaptado da nomenclatura inglesa “Highest density region”.

Capítulo 2

Definições e exemplos

Neste capítulo, são introduzidos os conceitos e exemplos fundamentais usados no trabalho, com particular destaque para as definições de núcleo associado discreto e estimador do núcleo, propostas por [1].

Seja X uma variável aleatória real discreta, de suporte $S \subseteq \mathbb{Z}$ e função de probabilidade f .

2.1 Definição de núcleo

Definição 1. *Seja $x \in S$ fixo e $(h_n)_{n \in \mathbb{N}}$ uma sucessão de números reais estritamente positivos tal que $h_n \rightarrow 0$ quando $n \rightarrow +\infty$.*

A família de funções de probabilidade $\{K_{x,h_n}, n \in \mathbb{N}\}$, com suporte $S_x \subseteq \mathbb{Z}$, diz-se um núcleo associado discreto se verifica as seguintes condições:

$$x \in S_x, \quad (2.1)$$

$$E(Z) = x + A(x, h_n), \quad (2.2)$$

$$\text{Var}(Z) = B(x, h_n), \quad (2.3)$$

onde, para cada n e x , $Z := Z(x, h_n)$ é uma variável aleatória discreta com função de probabilidade K_{x,h_n} e $\lim_{n \rightarrow +\infty} A(x, h_n) = \lim_{n \rightarrow +\infty} B(x, h_n) = 0$.

Para compreender a ideia que preside a esta definição, analisemos o comportamento assintótico de $P(Z = y)$, $y \in S_x$, quando n tende para $+\infty$.

Como $S_x \subseteq \mathbb{Z}$, dado $x \in S_x$, existe $\delta_x > 0$ tal que

$$]x - \delta_x, x + \delta_x[\cap S_x = \{x\}.$$

Consequentemente,

$$P(Z = x) = P_Z(]x - \delta_x, x + \delta_x[\cap S_x) = P_Z(]x - \delta_x, x + \delta_x[), \quad (2.4)$$

pois $P(Z \notin S_x) = 0$.

Mas

$$P_Z([x - \delta_x, x + \delta_x]) = P(|Z - x| < \delta_x) = 1 - P(|Z - x| \geq \delta_x), \quad (2.5)$$

pelo que, usando a desigualdade de Markov juntamente com as condições (2.2) e (2.3) da Definição 1, vem

$$\lim_{n \rightarrow +\infty} P(|Z - x| \geq \delta_x) \leq \lim_{n \rightarrow +\infty} \delta_x^{-2} E((Z - x)^2) = 0 \quad (2.6)$$

Portanto,

$$\forall x \in S_x, \lim_{n \rightarrow +\infty} P(Z = x) = 1 \quad (2.7)$$

e

$$\forall y \in S_x, y \neq x, \lim_{n \rightarrow +\infty} P(Z = y) = 0.$$

Concluimos, deste modo, que

$$\forall x \in S, \forall y \in S_x, \lim_{n \rightarrow +\infty} K_{x, h_n}(y) = \delta_{\{x\}}(y), \quad (2.8)$$

onde $\delta_{\{x\}}$ é a função de probabilidade da lei de Dirac no ponto x (cf. Exemplo 1).

Assim, as condições (2.2) e (2.3) asseguram que a probabilidade atribuída pelo núcleo ao ponto x aproxima-se de 1 à medida que a sucessão $(h_n)_{n \in \mathbb{N}}$ se aproxima de zero.

2.2 Exemplos de núcleos

Vejam os exemplos mais populares de núcleos associados discretos propostos na literatura (cf. Abdous e Kokonendji [1]).

Exemplo 1. A função de probabilidade da Lei de Dirac é o exemplo mais simples de núcleo associado discreto. Com efeito, sendo

$$\forall x \in S, \forall n \in \mathbb{N}, K_{x, h_n}(z) = \delta_{\{x\}}(z) = \mathbb{1}_{\{z=x\}},$$

verifica-se trivialmente que são satisfeitas as condições da Definição 1, com

$$S_x = \{x\}, E(Z) = x \text{ e } \text{Var}(Z) = 0.$$

Exemplo 2. No caso em que $S = \{0, 1, \dots, c-1\}$, com $c \in \mathbb{N} \setminus \{1\}$, o núcleo introduzido por Aitchison e Aitken [3] é definido por

$$K_{x, h_n}(z) = (1 - h_n) \mathbb{1}_{\{z=x\}} + \frac{h_n}{c-1} \mathbb{1}_{\{z \neq x\}}, \quad z \in S_x, h_n \in (0, 1), \quad (2.9)$$

onde $S_x = S$.

A validade das condições (2.2) e (2.3) da Definição 1 decorre facilmente da expressão de $E(Z^j)$, para $j = 1, 2$. De facto,

$$\begin{aligned} E(Z^j) &= \sum_{z \in S_x} z^j K_{x, h_n}(z) = x^j(1 - h_n) + \frac{h_n}{c-1} \sum_{z \in S_x \setminus \{x\}} z^j \\ &= x^j(1 - h_n) + \frac{h_n}{c-1} \left(\sum_{z \in S_x} z^j - x^j \right) \\ &= x^j + h_n \left(-\frac{c}{c-1} x^j + \frac{1}{c-1} \sum_{z \in S_x} z^j \right). \end{aligned}$$

Consequentemente,

$$E(Z) = x + h_n \left(-\frac{c}{c-1} x + \frac{c}{2} \right)$$

e

$$E(Z^2) = x^2 + h_n \left(-\frac{c}{c-1} x^2 + \frac{c(2c-1)}{6} \right),$$

pelo que

$$\begin{aligned} \text{Var}(Z) &= E(Z^2) - [E(Z)]^2 \\ &= h_n \left(\frac{c}{c-1} x^2 - xc + \frac{c(2c-1)}{6} \right) - h_n^2 \left(-\frac{c}{c-1} x + \frac{c}{2} \right)^2. \end{aligned}$$

Exemplo 3. Esta classe de núcleos foi proposta por Kokonendji et al. [23] e constitui uma versão do núcleo triangular simétrico. Sejam S um conjunto arbitrário, $m \in \mathbb{N}$ e $x \in S$ fixos.

Consideremos que a variável Z tem suporte $S_x = \{x, x \pm 1, \dots, x \pm m\}$ e função de probabilidade dada por

$$K_{x, h_n}(z) = \frac{(m+1)^{h_n} - |z-x|^{h_n}}{P_m(h_n)}, \quad z \in S_x, \quad h_n \in (0, 1),$$

onde

$$P_m(h_n) = (2m+1)(m+1)^{h_n} - 2 \sum_{k=1}^m k^{h_n}.$$

Pela simetria da distribuição de Z em relação a x temos que $E(Z) = x$.

Tendo em consideração que o suporte S_x possui $2m+1$ pontos, analisemos agora a variância de Z .

Temos que

$$\begin{aligned}
E(Z^2) &= \sum_{z \in S_x} z^2 K_{x, h_n}(z) \\
&= P_m^{-1}(h_n) \left[(m+1)^{h_n} \sum_{z \in S_x} z^2 - \sum_{z \in S_x} z^2 |z-x|^{h_n} \right] \\
&= P_m^{-1}(h_n) \left[(m+1)^{h_n} \left((2m+1)x^2 + 2 \sum_{k=1}^m k^2 \right) - 2 \sum_{k=1}^m x^2 k^{h_n} - 2 \sum_{k=1}^m k^{h_n+2} \right] \\
&= x^2 + P_m^{-1}(h_n) \left(3^{-1} m(2m+1)(m+1)^{h_n+1} - 2 \sum_{k=1}^m k^{h_n+2} \right).
\end{aligned}$$

Consequentemente,

$$\text{Var}(Z) = P_m^{-1}(h_n) \left(3^{-1} m(2m+1)(m+1)^{h_n+1} - 2 \sum_{k=1}^m k^{h_n+2} \right).$$

Como $c^{h_n} = \exp(h_n \ln c)$, para $c > 0$, temos¹

$$(m+1)^{h_n} = 1 + h_n \ln(m+1) + O_m(h_n^2) \tag{2.10}$$

e

$$k^{h_n} = 1 + h_n \ln k + O_k(h_n^2), \quad k = 1, \dots, m. \tag{2.11}$$

Então,

$$\begin{aligned}
P_m(h_n) &= (2m+1) \left(1 + h_n \ln(m+1) + O_m(h_n^2) \right) - 2 \sum_{k=1}^m \left(1 + h_n \ln k + O_k(h_n^2) \right) \\
&= 1 + \left((2m+1) \ln(m+1) - 2 \sum_{k=1}^m \ln k \right) h_n + O_m(h_n^2) \\
&= 1 + \zeta_m h_n + O_m(h_n^2).
\end{aligned}$$

Analogamente,

$$\begin{aligned}
&3^{-1} m(2m+1)(m+1)^{h_n+1} - 2 \sum_{k=1}^m k^{h_n+2} = \\
&= 3^{-1} m(2m+1)(m+1) \left(1 + h_n \ln(m+1) + O_m(h_n^2) \right) - 2 \sum_{k=1}^m k^2 \left(1 + h_n \ln k + O_k(h_n^2) \right) \\
&= \left(3^{-1} m(2m+1)(m+1) \ln(m+1) - 2 \sum_{k=1}^m k^2 \ln k \right) h_n + O_m(h_n^2) \\
&= \xi_m h_n + O_m(h_n^2).
\end{aligned}$$

¹Em Sedgewick e Flajolet [35] são definidas as expansões assintóticas da série exponencial e da série geométrica, derivadas do desenvolvimento em série de Taylor. Além disso, também são definidas as aproximações $o(\cdot)$ e $O(\cdot)$.

Obtemos, deste modo,¹

$$\text{Var}(Z) = \frac{\xi_m h_n + O_m(h_n^2)}{1 + \zeta_m h_n + O_m(h_n^2)} = O_m(h_n).$$

Na Figura 2.1 podemos ver duas representações gráficas deste núcleo, fixando $m = 6$, $x = 4$ e considerando diferentes valores da janela h_n .

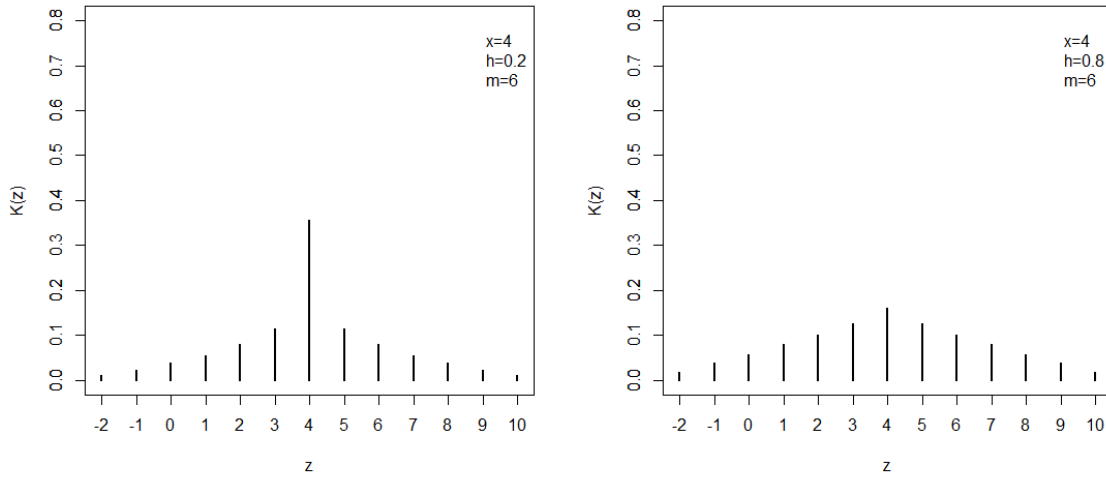


Figura 2.1 Configurações gráficas do núcleo apresentado no Exemplo 3.

Exemplo 4. Considerando $S = \mathbb{Z}$, Wang e Van Ryzin [43] propuseram a seguinte extensão do núcleo definido em (2.9):

$$K_{x,h_n}(z) = (1 - h_n) \mathbb{1}_{\{z=x\}} + \frac{1}{2}(1 - h_n) h_n^{|z-x|} \mathbb{1}_{\{|z-x| \geq 1\}}, \quad z \in S_x, \quad h_n \in (0, 1),$$

com $S_x = \mathbb{Z}$.

Como $E(|Z|) < +\infty$, então $E(Z) = x$, pela simetria da distribuição de Z em relação a x .

Para verificar a condição (2.3) da Definição 1, basta determinar a expressão de $E(Z^2)$.

Para tal, comecemos por notar que

$$\begin{aligned} \sum_{k \in \mathbb{N}} k^2 h_n^k &= h_n \sum_{k \in \mathbb{N}} k^2 h_n^{k-1} = h_n \frac{d}{dh_n} \left(h_n \sum_{k \in \mathbb{N}} k h_n^{k-1} \right) \\ &= h_n \frac{d}{dh_n} \left(h_n \frac{d}{dh_n} \left(\sum_{k \in \mathbb{N}} h_n^k \right) \right) \\ &= \frac{h_n(1 + h_n)}{(1 - h_n)^3}. \end{aligned}$$

Então,

$$\begin{aligned}
E(Z^2) &= \sum_{z \in S_x} z^2 K_{x, h_n}(z) \\
&= x^2(1 - h_n) + \frac{1}{2}(1 - h_n) \sum_{\{z \in \mathbb{Z}: |z-x| \geq 1\}} z^2 h_n^{|z-x|} \\
&= x^2(1 - h_n) + \frac{1}{2}(1 - h_n) \left(\sum_{\{z \in \mathbb{Z}: z-x \geq 1\}} z^2 h_n^{z-x} + \sum_{\{z \in \mathbb{Z}: z-x \leq -1\}} z^2 h_n^{-(z-x)} \right) \\
&= x^2(1 - h_n) + \frac{1}{2}(1 - h_n) \left(\sum_{k \in \mathbb{N}} (x+k)^2 h_n^k + \sum_{k \in \mathbb{N}} (x-k)^2 h_n^k \right) \\
&= x^2(1 - h_n) + (1 - h_n) \left(x^2 \sum_{k \in \mathbb{N}} h_n^k + \sum_{k \in \mathbb{N}} k^2 h_n^k \right) \\
&= x^2 + \frac{h_n(1 + h_n)}{(1 - h_n)^2}.
\end{aligned}$$

Portanto, vem

$$\text{Var}(Z) = \frac{h_n(1 + h_n)}{(1 - h_n)^2}.$$

Na Figura 2.2 apresentamos duas representações gráficas deste núcleo, fixando $x = 4$ e considerando diferentes valores da janela h_n .

Em alternativa, podemos sempre definir uma versão truncada do núcleo de Wang e Van Ryzin. Com efeito, para $m \in \mathbb{N}$ previamente fixado, basta considerar $S_x = \{x, x \pm 1, \dots, x \pm m\}$, $x \in S$, e

$$K_{x, h_n}(z) = (1 - h_n)(1 - h_n^{m+1})^{-1} [\mathbb{1}_{\{z=x\}} + \frac{1}{2} h_n^{|z-x|} \mathbb{1}_{\{|z-x| \geq 1\}}], \quad z \in S_x, \quad h_n \in (0, 1).$$

De modo análogo, somos conduzidos a $E(Z) = x$ e $\text{Var}(Z) = O_m(h_n)$.

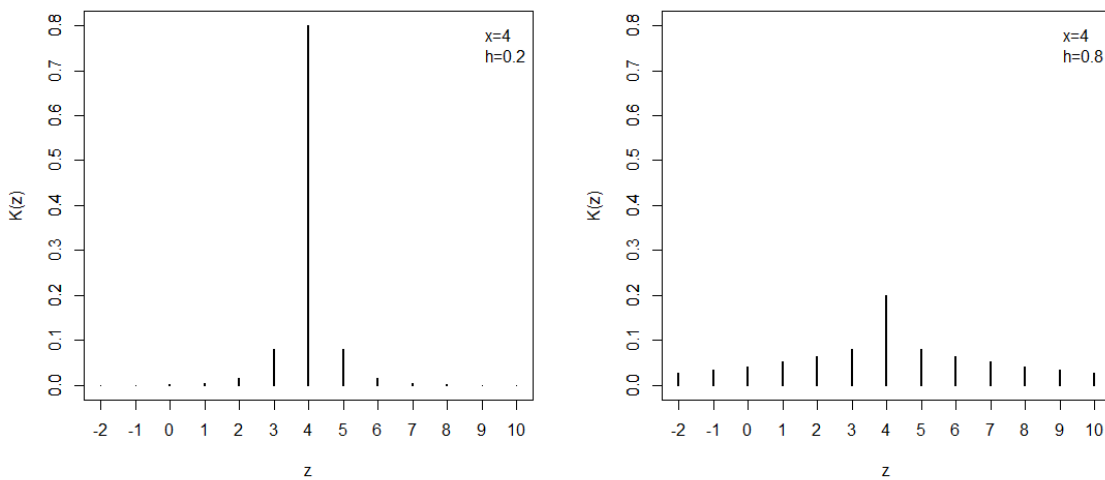


Figura 2.2 Configurações gráficas do núcleo apresentado no Exemplo 4.

Por fim, referimos a classe de núcleos recentemente proposta por Huang et al.² [18], construída a partir da distribuição de Conway-Maxwell-Poisson. Esta lei de probabilidade, que constitui uma generalização da lei de Poisson, tem suporte \mathbb{N}_0 e função de probabilidade dada por

$$\frac{1}{C(\lambda, \nu)} \frac{\lambda^y}{(y!)^\nu}, y \in \mathbb{N}_0, \quad (2.12)$$

onde

$$C(\lambda, \nu) = \sum_{y \in \mathbb{N}_0} \frac{\lambda^y}{(y!)^\nu}$$

é a constante de normalização e λ e ν são dois parâmetros tais que $(\lambda, \nu) \in \mathbb{R}^+ \times \mathbb{R}^+ \cup]0, 1[\times \{0\}$. Encontramos em Daly e Gaunt [10] uma compilação das suas propriedades fundamentais. O facto de não ser conhecida uma forma fechada de $C(\lambda, \nu)$ impossibilita a obtenção de expressões exatas para os seus momentos simples, por exemplo, e, conseqüentemente, as relações exatas entre tais momentos e os parâmetros λ e ν . Ainda assim, é possível, cf. Sellers et al. [36], interpretar ν como um parâmetro de dispersão, por comparação com a dispersão da lei de Poisson (que corresponde a $\nu = 1$), sendo que valores de ν superiores (inferiores) a 1 traduzem subdispersão (sobredispersão) relativamente à lei de Poisson. A utilização deste modelo no presente contexto é conseguida através de uma reparametrização em que a média da distribuição intervém diretamente como parâmetro e é funcionalmente independente de ν . Denotando esse valor médio por μ , Huang [17], introduz a distribuição de Conway-Maxwell-Poisson com média $\mu \geq 0$ e dispersão $\nu \geq 0$, que denotamos por $CMP(\mu, \nu)$, substituindo na expressão (2.12) a constante λ pela função de μ e ν , $\lambda(\mu, \nu)$, que satisfaz a igualdade

$$\sum_{y \in \mathbb{N}_0} \frac{[\lambda(\mu, \nu)]^y}{(y!)^\nu} (y - \mu) = 0.$$

Deste modo, se $Y \sim CMP(\mu, \nu)$, então $E(Y) = \mu$. Por outro lado, a proposição 3.1 de Huang et al. [18] estabelece, sem demonstração, que

$$\forall \mu \in \mathbb{N}_0, \lim_{\nu \rightarrow +\infty} Var(Y) = 0.$$

A definição de um núcleo associado discreto a partir desta construção é, agora, evidente.

Exemplo 5. *A seguinte classe de núcleos foi introduzida por Huang [17], e é baseada na distribuição de Conway-Maxwell-Poisson. Sejam $S = S_x = \mathbb{N}_0$ e*

$$K_{x, h_n}(z) = \frac{1}{C(\lambda(x, h_n^{-1}), h_n^{-1})} \frac{[\lambda(x, h_n^{-1})]^z}{(z!)^{h_n^{-1}}}, z \in S_x, h_n \in (0, 1),$$

com

$$C(\lambda(x, h_n^{-1}), h_n^{-1}) = \sum_{z \in \mathbb{N}_0} \frac{[\lambda(x, h_n^{-1})]^z}{(z!)^{h_n^{-1}}},$$

²E também abordada em Esstafa et al. [12]

onde $\lambda(x, h_n^{-1})$ satisfaz a equação

$$\sum_{y \in \mathbb{N}_0} \frac{[\lambda(x, h_n^{-1})]^y}{(y!)^{h_n^{-1}}} (y - x) = 0.$$

Na Figura 2.3 podemos ver duas representações gráficas deste núcleo, fixando $x = 4$ e considerando dois valores da janela h_n . Notemos que os diferentes valores da constante C decorrem dos diferentes valores de h_n , pois x está fixo, não sendo portanto escolhidos.

Concluimos esta secção com as seguintes observações:

1. Em rigor, um núcleo associado discreto diz-se de segunda ordem quando $\lim_{n \rightarrow +\infty} B(x, h_n) = 0$ e de primeira ordem no caso contrário; veja-se, por exemplo, Kokonendji e Kiessé [24] e Senga Kiessé [37]. Exemplos desta última classe são os núcleos construídos a partir das leis de probabilidade usuais de Poisson, binomial e binomial-negativa. Tal como argumentam os autores, os estimadores construídos com base em núcleos de primeira ordem, apesar de não serem consistentes, apresentam bons comportamentos na prática, em estudos de simulação envolvendo amostras de pequena ou média dimensão, quando comparados com os estimadores definidos a partir dos núcleos de segunda ordem, o que pode justificar a sua utilização;
2. Tanto quanto é do nosso conhecimento, não existe um método geral que permita construir núcleos associados discretos, pelo que a sua construção é efetuada caso a caso. Notemos ainda que, apesar de as condições (2.2) e (2.3) aparentarem ser bastante simples, são escassos os exemplos de núcleos que as verificam. Por outro lado, a condição (2.1) obriga a que $S \subseteq \bigcup_{x \in S} S_x$ e, portanto, a escolha do núcleo deve ter em conta o suporte da lei de probabilidade subjacente à amostra. Quando $S \subset \bigcup_{x \in S} S_x$, o desempenho do estimador a distância finita fica comprometido devido aos denominados *efeitos de fronteira*³. Tal acontece, por exemplo, se S for limitado (inferiormente ou superiormente) e o estimador for construído com base num núcleo do tipo triangular. Com efeito, como o suporte do núcleo é da forma $S_x = \{x, x \pm 1, \dots, x \pm m\}$, $m \in \mathbb{N}$, ele atribui probabilidade positiva a pontos que não pertencem a S , o que conduz à subestimação de $f(x_0)$ para x_0 próximo dos limites inferior ou superior de S . Este problema é discutido e solucionado por Kokonendji e Zocchi [26], que propõem a utilização de versões assimétricas ou simétricas do núcleo triangular consoante a distância entre x_0 e os limites inferior e superior de S .

2.3 Definição de estimador do núcleo

Definição 2. *Seja (X_1, \dots, X_n) uma amostra aleatória da variável aleatória real X . Dado um núcleo associado discreto $\{K_{x, h_n}, n \in \mathbb{N}\}$, o estimador do núcleo da função de probabilidade desconhecida f , baseado na amostra (X_1, \dots, X_n) , é definido por*

$$\forall x \in S, \widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{x, h_n}(X_i). \quad (2.13)$$

³Também sobejamente conhecidos no contexto da estimação não paramétrica de uma função densidade, cf. Karunamuni e Alberts [21].

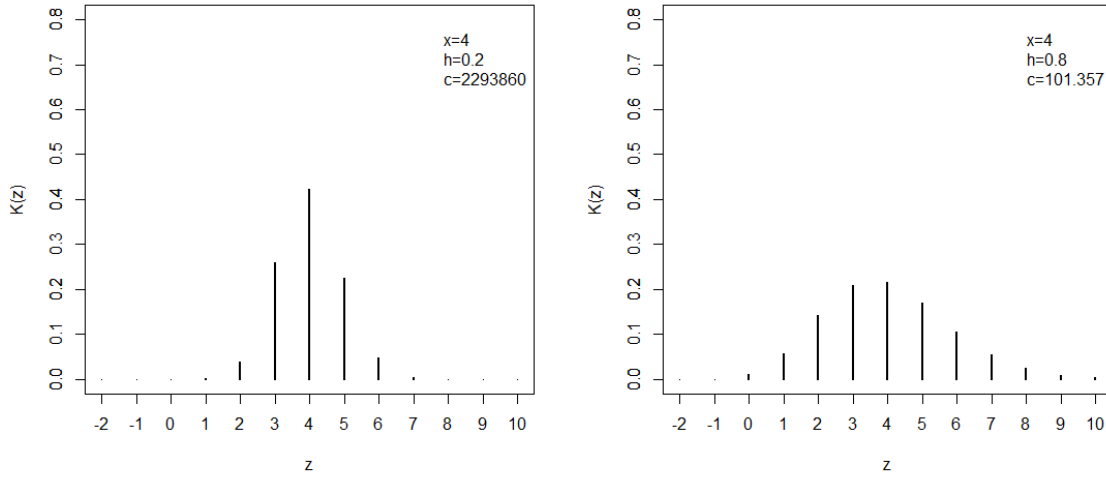


Figura 2.3 Configurações gráficas do núcleo apresentado no Exemplo 5.

É de salientar que as condições impostas ao núcleo não garantem que $\widehat{f}_n(x)$ seja uma função de probabilidade sobre S . Com efeito, tem-se $\widehat{f}_n(x) \in [0, 1]$, $x \in S$, e

$$\sum_{x \in S} \widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n Y_{i,h_n} := C_n, \quad (2.14)$$

com

$$Y_{i,h_n} = \sum_{x \in S} K_{x,h_n}(X_i) \quad (2.15)$$

e C_n uma variável aleatória (*q.c.*) finita se

$$\forall y \in S, \sum_{x \in S} K_{x,h_n}(y) < +\infty.$$

Notemos que C_n é (*q.c.*) positiva uma vez que

$$\forall y \in S, \sum_{x \in S} K_{x,h_n}(y) = \sum_{x \in S \cap S_x} K_{x,h_n}(y)$$

e

$$\forall y \in S, 0 < K_{x,h_n}(x) \leq \sum_{x \in S \cap S_x} K_{x,h_n}(y).$$

Para os núcleos apresentados nos Exemplos 1, 2 e 4 verifica-se que $C_n = 1$ *q.c.*. Em relação ao núcleo apresentado no Exemplo 3 verifica-se que C_n é finito (*q.c.*) se S é finito e, em particular, $C_n = 1$ *q.c.* se $S = S_x$. Mais geralmente, dada uma família de núcleos associados discretos simétricos em relação a x e tal que $S = S_x = \mathbb{Z}$, facilmente verificamos que $C_n = 1$ *q.c.*.

Contudo, quando $C_n \neq 1$, podemos normalizar o estimador, considerando

$$f_n^*(x) = \frac{\widehat{f}_n(x)}{C_n}, x \in \mathcal{S}. \quad (2.16)$$

Mas, de um modo geral, $(C_n)_{n \in \mathbb{N}}$ é uma sucessão de variáveis aleatórias e há que estudar o seu comportamento assintótico, bem como o do estimador normalizado. Esse estudo será abordado na secção 3.2.

Salientamos ainda que no caso do núcleo referido no Exemplo 5 não foi possível tirar conclusões sobre a constante C_n , pois a função λ não é definida explicitamente.

Capítulo 3

Propriedades locais de convergência

Dedicar-nos-emos, neste capítulo, ao estudo dos comportamentos assintóticos pontuais dos estimadores. Começaremos por estudar o viés, a variância e o erro quadrático médio do estimador não normalizado, terminando com as propriedades do estimador normalizado. Os resultados apresentados foram obtidos por [1] e [25].

3.1 Estimador não normalizado

Lema 1. Para todo o $x \in S$ e $n \in \mathbb{N}$ tem-se

$$E(\widehat{f}_n(x)) = E(f(Z)),$$

onde Z é uma variável aleatória discreta com função de probabilidade K_{x,h_n} .

Demonstração. Basta ter em conta que

$$\begin{aligned} \forall x \in S, E(\widehat{f}_n(x)) &= E\left(\frac{1}{n} \sum_{i=1}^n K_{x,h_n}(X_i)\right) \\ &= E(K_{x,h_n}(X_1)) \\ &= \sum_{y \in S} K_{x,h_n}(y) P(X_1 = y) \\ &= \sum_{y \in S \cap S_x} P(Z = y) f(y) \\ &= E(f(Z)), \end{aligned}$$

uma vez que $k_{x,h_n}(y)f(y) = 0$ para $y \in \overline{S \cap S_x}$. ■

O teorema que se segue permite afirmar que o estimador é assintoticamente cêntrico.

Teorema 1. Sob as condições (2.1)-(2.3) da Definição 1, tem-se

$$\forall x \in S, \lim_{n \rightarrow +\infty} E(\widehat{f}_n(x)) = f(x).$$

Demonstração. Seja $x \in S$ fixo.

$$\begin{aligned}
 E(\widehat{f}_n(x)) - f(x) &= \sum_{y \in S \cap S_x} f(y)P(Z=y) - f(x) \sum_{y \in S_x} P(Z=y) \\
 &= \sum_{y \in S \cap S_x} f(y)P(Z=y) - f(x) \left[\sum_{y \in S \cap S_x} P(Z=y) + \sum_{y \in \bar{S} \cap S_x} P(Z=y) \right] \\
 &= \sum_{y \in S \cap S_x \setminus \{x\}} [f(y) - f(x)]P(Z=y) - f(x) \sum_{y \in \bar{S} \cap S_x} P(Z=y) \\
 &:= B_{1,n} + B_{2,n}.
 \end{aligned}$$

Como S e S_x são subconjuntos de \mathbb{Z} , existe $\delta_x > 0$ tal que $|y - x| \geq \delta_x$ para todo o $y \in S \cap S_x \setminus \{x\}$. Logo

$$\begin{aligned}
 |B_{1,n}| &\leq \sum_{y \in S \cap S_x \setminus \{x\}} (|f(y)| + |f(x)|)P(Z=y) \\
 &\leq 2 \sum_{y \in S \cap S_x \setminus \{x\}} P(Z=y) \\
 &= 2P(Z \in S \cap S_x \setminus \{x\}) \\
 &= 2P(Z \in S \cap S_x \setminus \{x\}, |Z - x| \geq \delta_x) \\
 &\leq 2P(|Z - x| \geq \delta_x),
 \end{aligned}$$

donde se deduz, usando (2.6), que

$$\lim_{n \rightarrow +\infty} |B_{1,n}| = 0.$$

No que diz respeito a $B_{2,n}$, vem

$$\begin{aligned}
 |B_{2,n}| &\leq f(x) \sum_{y \in \bar{S} \cap S_x} P(Z=y) \\
 &= f(x)P(Z \in \bar{S} \cap S_x) \\
 &= f(x)P(Z \in \bar{S} \cap S_x \setminus \{x\}) \\
 &\leq f(x)P(Z \in S_x \setminus \{x\}) \\
 &\leq f(x)P(|Z - x| \geq \delta'_x),
 \end{aligned}$$

para algum $\delta'_x > 0$.

Consequentemente, um raciocínio análogo ao anterior permite concluir que $\lim_{n \rightarrow +\infty} |B_{2,n}| = 0$. ■

Analisemos agora o comportamento assintótico da variância do estimador.

Teorema 2. *Sob as condições (2.1)-(2.3) da Definição 1, tem-se*

$$\forall x \in S, n\text{Var}(\widehat{f}_n(x)) \xrightarrow{n \rightarrow +\infty} f(x)(1 - f(x)).$$

Demonstração. Para x arbitrariamente fixo, escrevamos

$$\begin{aligned} |n\text{Var}(\widehat{f}_n(x)) - f(x)(1 - f(x))| &= |\text{Var}(K_{x,h_n}(X_1)) - f(x)(1 - f(x))| \\ &= |E(K_{x,h_n}^2(X_1)) - [E(\widehat{f}_n(x))]^2 - f(x) + f^2(x)| \\ &\leq \left| \sum_{y \in S \cap S_x} K_{x,h_n}^2(y)f(y) - f(x) \right| + \left| [E(\widehat{f}_n(x))]^2 - f^2(x) \right| \\ &\leq f(x) |K_{x,h_n}^2(x) - 1| + \sum_{y \in S \cap S_x \setminus \{x\}} K_{x,h_n}^2(y)f(y) + |E(\widehat{f}_n(x))]^2 - f^2(x)|. \end{aligned}$$

Ora,

- $\forall n \in \mathbb{N}, \forall y \in S \cap S_x \setminus \{x\}, K_{x,h_n}^2(y)f(y) \leq f(y)$, com $\sum_{y \in S \cap S_x \setminus \{x\}} f(y) < 1$;
- $\forall y \in S \cap S_x \setminus \{x\}, \lim_{n \rightarrow +\infty} K_{x,h_n}^2(y) \leq \lim_{n \rightarrow +\infty} K_{x,h_n}(y) = 0$.

Assim, pelo teorema da convergência dominada de Lebesgue, vem

$$\lim_{n \rightarrow +\infty} \sum_{y \in S \cap S_x \setminus \{x\}} K_{x,h_n}^2(y)f(y) = \sum_{y \in S \cap S_x \setminus \{x\}} \lim_{n \rightarrow +\infty} K_{x,h_n}^2(y)f(y) = 0,$$

pelo que o segundo termo da desigualdade anterior converge para zero.

Como a convergência para zero do terceiro termo decorre diretamente do Teorema 1 e

$$\lim_{n \rightarrow +\infty} K_{x,h_n}^2(x) = 1,$$

a demonstração fica concluída. ■

Corolário 1. *Sob as condições (2.1)-(2.3) da Definição 1, tem-se*

$$\forall x \in S, \lim_{n \rightarrow +\infty} \text{Var}(\widehat{f}_n(x)) = 0.$$

A qualidade de um estimador não paramétrico num ponto x é frequentemente medida através do erro quadrático médio, definido por

$$EQM\{\widehat{f}_n(x)\} = E \left[(\widehat{f}_n(x) - f(x))^2 \right],$$

que admite a decomposição

$$EQM\{\widehat{f}_n(x)\} = \text{Var}(\widehat{f}_n(x)) + \left[\text{Viés}(\widehat{f}_n(x)) \right]^2.$$

O próximo resultado é, assim, uma consequência da decomposição anterior e dos Teoremas 1 e 2.

Corolário 2. *Sob as condições (2.1)-(2.3) da Definição 1, tem-se*

$$\forall x \in S, \lim_{n \rightarrow +\infty} EQM\{\widehat{f}_n(x)\} = 0.$$

Como é sabido, da relação entre as convergências em média quadrática e em probabilidade, concluímos também a seguinte propriedade.

Corolário 3. *Sob as condições (2.1)-(2.3) da Definição 1, tem-se*

$$\forall x \in \mathcal{S}, \widehat{f}_n(x) \xrightarrow[n \rightarrow +\infty]{P} f(x).$$

Um outro critério usado para avaliar a qualidade de um estimador é baseado na noção de convergência quase certa. No contexto da estimação não paramétrica, a obtenção de resultados deste tipo é conseguida a partir das denominadas desigualdades exponenciais.

O lema seguinte recorda a que usaremos no presente trabalho (cf. Hoeffding [16]).

Lema 2. *Sejam Y_1, \dots, Y_n variáveis aleatórias reais independentes e identicamente distribuídas com momento de segunda ordem finito. Se existem números reais a e b tais que $P(Y_i \in [a, b]) = 1$, então*

$$\forall \varepsilon > 0, P\left(\left|\frac{1}{n} \sum_{i=1}^n Y_i\right| \geq \varepsilon\right) \leq 2 \exp\left\{-\frac{n\varepsilon^2}{\varepsilon(b-a) + 2\text{Var}(Y_1)}\right\}.$$

Teorema 3. *Sob as condições (2.1)-(2.3) da Definição 1, tem-se*

$$\forall x \in \mathcal{S}, \widehat{f}_n(x) \xrightarrow[n \rightarrow +\infty]{q.c.} f(x).$$

Demonstração. Sejam $x \in \mathcal{S}$ e $n \in \mathbb{N}$ arbitrariamente fixos. Atendendo a que

$$\widehat{f}_n(x) - f(x) = \left[\widehat{f}_n(x) - E(\widehat{f}_n(x))\right] + \text{Viés}(\widehat{f}_n(x))$$

e ao resultado obtido no Teorema 1, basta mostrar que

$$\left[\widehat{f}_n(x) - E(\widehat{f}_n(x))\right] \xrightarrow[n \rightarrow +\infty]{q.c.} 0.$$

Para tal, notemos que

$$\widehat{f}_n(x) - E(\widehat{f}_n(x)) = \frac{1}{n} \sum_{i=1}^n Y_i, \quad (3.1)$$

com $Y_i = K_{x,h_n}(X_i) - E(K_{x,h_n}(X_i))$, $i = 1, \dots, n$, variáveis aleatórias independentes e identicamente distribuídas.

Ora, para todo o $i \in \{1, \dots, n\}$,

$$P(K_{x,h_n}(X_i) \in [0, 1]) = 1,$$

que implica

$$P(Y_i \in [-1, 1]) = 1$$

e

$$\text{Var}(Y_i) = \text{Var}(K_{x,h_n}(X_i)) \leq E(K_{x,h_n}^2(X_i)) \leq 1,$$

o que mostra que tais variáveis satisfazem as condições do lema anterior.

Então, aplicando o Lema 2, vem

$$\forall \varepsilon > 0, P\left(|\widehat{f}_n(x) - E(\widehat{f}_n(x))| \geq \varepsilon\right) = P\left(\left|\frac{1}{n} \sum_{i=1}^n Y_i\right| \geq \varepsilon\right) \leq 2 \exp\left\{\frac{-n\varepsilon^2}{2\varepsilon + 2}\right\}.$$

Portanto,

$$\forall \varepsilon > 0, \sum_{n \geq 1} P\left(|\widehat{f}_n(x) - E(\widehat{f}_n(x))| \geq \varepsilon\right) \leq 2 \sum_{n \geq 1} \exp\left\{\frac{-n\varepsilon^2}{2\varepsilon + 2}\right\} < +\infty.$$

O resultado pretendido decorre, pois, do lema de Borel-Cantelli (cf. [7], Lema 3.14). ■

Apresentamos agora uma propriedade relativa à normalidade assintótica do estimador. Antes, porém, convém recordar o seguinte resultado auxiliar, que podemos encontrar em Breiman ([7], Teorema 9.2).

Lema 3. *Seja $(X_n)_{n \in \mathbb{N}}$ uma sucessão de variáveis aleatórias independentes e identicamente distribuídas, centradas e com momento de terceira ordem. Seja ainda $s_n^2 = \sum_{k=1}^n E(X_k^2)$, $n \in \mathbb{N}$.*

Se

$$\lim_{n \rightarrow +\infty} s_n^{-3} \sum_{k=1}^n E(|X_k|^3) = 0,^1$$

então

$$s_n^{-1} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} N(0, 1).$$

Teorema 4. *Sob as condições (2.1)-(2.3) da Definição 1, tem-se*

$$\forall x \in S, \frac{\widehat{f}_n(x) - E(\widehat{f}_n(x))}{\sqrt{\text{Var}(\widehat{f}_n(x))}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} N(0, 1).$$

Demonstração. Seja $x \in S$ arbitrariamente fixo.

Começemos por escrever

$$\widehat{f}_n(x) - E(\widehat{f}_n(x)) = \sum_{i=1}^n Z_i,$$

com $Z_i = n^{-1}Y_i$, $i = 1, \dots, n$, onde Y_i são as variáveis definidas em (3.1), e provemos que a sucessão $(Z_n)_{n \in \mathbb{N}}$ satisfaz as condições do Lema 3.

Tais variáveis são claramente independentes e identicamente distribuídas, centradas, e verificam

$$E(|Z_i|^3) = \frac{1}{n^3} E(|Y_1|^3) = \frac{1}{n^3} E\left[|K_{x,h_n}(X_1) - E(K_{x,h_n}(X_1))|^3\right] \leq \frac{2^3}{n^3} < +\infty,$$

¹Note-se que $\lim_{n \rightarrow +\infty} u_n = 0 \Rightarrow \overline{\lim}_{n \rightarrow +\infty} u_n = \lim_{n \rightarrow +\infty} u_n = 0$.

uma vez que

$$|K_{x,h_n}(X_1) - E(K_{x,h_n}(X_1))| \leq |K_{x,h_n}(X_1)| + |E(K_{x,h_n}(X_1))| \leq 2, \text{ q.c.}$$

Logo,

$$\sum_{i=1}^n E(|Z_i|^3) = nE(|Z_1|^3) \leq \frac{2^3}{n^2}, n \in \mathbb{N}.$$

Além disso,

$$s_n^2 = \text{Var}(\widehat{f}_n(x)) = \frac{1}{n} \text{Var}(K_{x,h_n}(X_1)), n \in \mathbb{N}.$$

Consequentemente,

$$\frac{1}{s_n^3} \sum_{i=1}^n E(|Z_i|^3) \leq \frac{2^3}{\sqrt{n} [\text{Var}(K_{x,h_n}(X_1))]^{\frac{3}{2}}}, n \in \mathbb{N},$$

donde se obtém o resultado enunciado, invocando o Teorema 2. ■

A próxima propriedade, inspirada em Esstafa et al. ([12], Teorema 2.4), é uma consequência direta do teorema anterior.

Corolário 4. *Suponhamos que, nas condições (2.1)-(2.3) da Definição 1, se tem $A(x, h_n) = O_x(h_n)$ e $B(x, h_n) = O_x(h_n)$. Se a sucessão $(h_n)_{n \in \mathbb{N}}$ é tal que $\lim_{n \rightarrow +\infty} \sqrt{n}h_n = 0$, então*

$$\forall x \in S : f(x) < 1, \sqrt{n} \frac{\widehat{f}_n(x) - f(x)}{\sqrt{f(x)(1-f(x))}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} N(0, 1).$$

Demonstração. Para x nas condições dadas, temos

$$\sqrt{n} \frac{\widehat{f}_n(x) - f(x)}{\sqrt{f(x)(1-f(x))}} = \sqrt{n} \frac{\widehat{f}_n(x) - E(\widehat{f}_n(x))}{\sqrt{f(x)(1-f(x))}} + \sqrt{n} \frac{E(\widehat{f}_n(x)) - f(x)}{\sqrt{f(x)(1-f(x))}} = U_{1,n} + U_{2,n}.$$

Reescrevendo $U_{1,n}$ na forma

$$\frac{\widehat{f}_n(x) - E(\widehat{f}_n(x))}{\sqrt{\text{Var}(\widehat{f}_n(x))}} \frac{\sqrt{n \text{Var}(\widehat{f}_n(x))}}{\sqrt{f(x)(1-f(x))}},$$

concluimos, pelos Teoremas 2 e 4, que

$$U_{1,n} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} N(0, 1).$$

Relativamente a $U_{2,n}$, basta analisar o comportamento assintótico de $\sqrt{n}(E(\widehat{f}_n(x)) - f(x))$.

Para tal, notemos que

$$\begin{aligned} \sqrt{n}|E(\widehat{f}_n(x)) - f(x)| &= \sqrt{n} \left| \sum_{y \in S \cap S_x} K_{x,h_n}(y)f(y) - f(x) \right| \\ &\leq \sqrt{n}f(x)(1 - K_{x,h_n}(x)) + \sqrt{n} \sum_{y \in S \cap S_x \setminus \{x\}} K_{x,h_n}(y)f(y) \\ &\leq 2\sqrt{n}P(Z \neq x). \end{aligned}$$

Pela desigualdade de Markov, e atendendo a (2.4) e (2.5), vem

$$P(Z \neq x) = P(|Z - x| \geq \delta_x) \leq \delta_x^{-2} \left(\text{Var}(Z) + (E(Z - x))^2 \right) = O_x(h_n).$$

Logo, a condição imposta à sucessão $(h_n)_{n \in \mathbb{N}}$ assegura que

$$\sqrt{n}P(Z \neq x) \leq \sqrt{n}O_x(h_n) \xrightarrow{n \rightarrow +\infty} 0,$$

o que conclui a demonstração, pelo teorema de Slutsky. ■

3.2 Estimador normalizado

Nesta secção, que tem por base o artigo de Kokonendji e Varron [25], iremos analisar o comportamento assintótico do estimador definido em (2.16), sob a perspectiva das convergências em probabilidade e quase certa. Para isso, começaremos por estudar o comportamento assintótico da sucessão $(C_n)_{n \in \mathbb{N}}$ (cf. (2.14) e (2.15)), impondo as seguintes condições adicionais ao núcleo associado discreto:

- $HK_0(l)$: Existem uma família de funções de probabilidade $\{g_y, y \in S\}$ e uma função real mensurável η tais que $E(\eta^l(X)) < +\infty$ e $K_{x,h_n}(y) \leq \eta(y)g_y(x)$, para todos $(x, y) \in S^2$ e $n \in \mathbb{N}$.
- HK_1 : Para alguma constante $C > 0$, tem-se $\sup_{y \in S} \sup_{n \in \mathbb{N}} \sum_{x \in S} K_{x,h_n}(y) \leq C$.

3.2.1 Convergência em probabilidade

O nosso primeiro resultado diz respeito à convergência em média de ordem l da sucessão $(Y_{1,h_n})_{n \in \mathbb{N}}$.

Teorema 5. *Se o núcleo associado discreto satisfaz a hipótese $HK_0(l)$, para algum $l \in \mathbb{N}$, então*

$$\lim_{n \rightarrow +\infty} E(Y_{1,h_n}^l) = 1.$$

Demonstração. Seja $l \in \mathbb{N}$ arbitrariamente fixo. Pelo teorema da convergência monótona, vem

$$\begin{aligned} \forall n \in \mathbb{N}, E(Y_{1,h_n}^l) &= E\left(\left(\sum_{x \in S} K_{x,h_n}(X)\right)^l\right) = \sum_{(x_1, \dots, x_l) \in S^l} E\left(\prod_{i=1}^l K_{x_i, h_n}(X)\right) \\ &= \sum_{(x_1, \dots, x_l) \in S^l} \sum_{y \in S} \left[\prod_{i=1}^l K_{x_i, h_n}(y) \mathbb{1}_{S_{x_i}}(y)\right] f(y) \\ &\leq \sum_{y \in S} \sum_{(x_1, \dots, x_l) \in S^l} f(y) \prod_{i=1}^l K_{x_i, h_n}(y). \end{aligned}$$

Ora, sob a hipótese $HK_0(l)$, cada termo da (dupla) soma anterior é majorado por

$$F(y, x_1, \dots, x_l) = f(y) \eta^l(y) \prod_{i=1}^l g_y(x_i),$$

com

$$\begin{aligned} \sum_{y \in S} \sum_{(x_1, \dots, x_l) \in S^l} F(y, x_1, \dots, x_l) &= \sum_{y \in S} f(y) \eta^l(y) \sum_{(x_1, \dots, x_l) \in S^l} \prod_{i=1}^l g_y(x_i) \\ &= \sum_{y \in S} f(y) \eta^l(y) \\ &= E(\eta^l(X)) < +\infty. \end{aligned}$$

Assim, dado o resultado estabelecido em (2.8), obtemos

$$\begin{aligned} \lim_{n \rightarrow +\infty} E(Y_{1,h_n}^l) &= \sum_{y \in S} f(y) \sum_{(x_1, \dots, x_l) \in S^l} \lim_{n \rightarrow +\infty} \prod_{i=1}^l K_{x_i, h_n}(y) \mathbb{1}_{S_{x_i}}(y) \\ &= \sum_{y \in S} f(y) \sum_{(x_1, \dots, x_l) \in S^l} \prod_{i=1}^l \delta_{\{x_i\}}(y) \mathbb{1}_{S_{x_i}}(y) \\ &= \sum_{y \in S} f(y) \\ &= 1, \end{aligned}$$

pelo teorema da convergência dominada de Lebesgue. ■

Deste teorema, deduzimos facilmente a convergência em média quadrática da sucessão $(C_n)_{n \in \mathbb{N}}$.

Corolário 5. *Sob as condições da hipótese do teorema anterior, tem-se*

$$C_n \xrightarrow[n \rightarrow +\infty]{m.q.} 1.$$

Demonstração. Observemos que, pelas propriedades dos momentos, se a hipótese $HK_0(2)$ se verificar, o raciocínio efectuado no teorema anterior permite afirmar que

$$\lim_{n \rightarrow +\infty} E(Y_{1,h_n}^j) = 1, \quad j = 1, 2,$$

o que implica que

$$\lim_{n \rightarrow +\infty} \text{Var}(C_n - 1) = \lim_{n \rightarrow +\infty} n^{-1} \text{Var}(Y_{1,h_n}) = 0.$$

Deste modo, a fórmula de Kőenig conduz a

$$E((C_n - 1)^2) = \text{Var}(C_n - 1) + [E(Y_{1,h_n}) - 1]^2 \xrightarrow{n \rightarrow +\infty} 0.$$

■

Estamos agora em condições de apresentar a propriedade relativa à convergência em probabilidade do estimador normalizado.

Corolário 6. *Sob as condições da hipótese do Teorema 5, tem-se*

$$\forall x \in \mathcal{S}, f_n^*(x) \xrightarrow[n \rightarrow +\infty]{P} f(x).$$

Demonstração. Como a convergência em média quadrática implica a convergência em probabilidade, concluímos que $(C_n)_{n \in \mathbb{N}}$ converge em probabilidade para 1, quando $n \rightarrow +\infty$, pelo Corolário 5. Para concluir a demonstração, resta relembrar o resultado obtido no Corolário 3 e aplicar as propriedades operatórias da convergência em probabilidade.

■

3.2.2 Convergência quase certa

O próximo objetivo é estabelecer a convergência quase certa de $(C_n)_{n \in \mathbb{N}}$. Para o efeito, será usado o seguinte resultado auxiliar, que podemos encontrar em Pollard [31].

Lema 4 (Desigualdade de Bennett). *Sejam X_1, \dots, X_n variáveis aleatórias independentes, centradas e tais que $|X_i| \leq b$ q.c., $i = 1, \dots, n$, para algum $b > 0$. Sejam ainda*

$$S = \sum_{i=1}^n X_i \quad e \quad v \in \mathbb{R} : \sum_{i=1}^n \text{Var}(X_i) \leq v.$$

Então,

$$\forall t > 0, P(|S| > t) \leq 2 \exp \left\{ -\frac{v}{b^2} \phi(bt v^{-1}) \right\},$$

onde $\phi(\lambda) = (1 + \lambda) \log(1 + \lambda) - \lambda$, $\lambda > 0$.

Teorema 6. *Se o núcleo associado discreto satisfaz as hipóteses $HK_0(2)$ e HK_1 , tem-se*

$$C_n \xrightarrow[n \rightarrow +\infty]{q.c.} 1.$$

Demonstração. Atendendo a que

$$C_n - 1 = \{C_n - E(Y_{1,h_n})\} + E(Y_{1,h_n}) - 1$$

e ao resultado obtido no Teorema 5, basta provar que

$$\{C_n - E(Y_{1,h_n})\} \xrightarrow[n \rightarrow +\infty]{q.c.} 0.$$

Para tal, escrevamos

$$C_n - E(Y_{1,h_n}) = \frac{1}{n} \sum_{i=1}^n W_i,$$

com $W_i = Y_{i,h_n} - E(Y_{i,h_n})$, $i = 1, \dots, n$, variáveis aleatórias independentes e identicamente distribuídas. Ora, pela hipótese HK_1 , vem

$$Y_{i,h_n} \leq C \text{ q.c.}, \quad i = 1, \dots, n,$$

que implica

$$|W_i| \leq 2C \text{ q.c.}, \quad i = 1, \dots, n,$$

o que mostra que tais variáveis satisfazem as condições do Lema 4.

Por outro lado,

$$\sum_{i=1}^n \text{Var}(W_i) = \sum_{i=1}^n \text{Var}(Y_{i,h_n}) \leq \sum_{i=1}^n E(Y_{i,h_n}^2) = nE(Y_{1,h_n}^2) = nv_n,$$

com $v_n = E(Y_{1,h_n}^2) \xrightarrow[n \rightarrow +\infty]{} 1$.

Consequentemente, tomando $v = nv_n$ no Lema 4, somos conduzidos a

$$\begin{aligned} \forall \varepsilon > 0, \sum_{n \in \mathbb{N}} P\left(|C_n - E(Y_{1,h_n})| > \varepsilon\right) &= \sum_{n \in \mathbb{N}} P\left(\left|\sum_{i=1}^n W_i\right| > n\varepsilon\right) \\ &\leq 2 \sum_{n \in \mathbb{N}} \exp\left\{-\frac{nv_n}{4C^2} \phi(2C\varepsilon v_n^{-1})\right\} \\ &= 2 \sum_{n \in \mathbb{N}} \exp\left\{-\frac{nv_n u_n}{4C^2}\right\} < +\infty, \end{aligned}$$

pois $u_n = \phi(2C\varepsilon v_n^{-1}) \xrightarrow[n \rightarrow +\infty]{} a > 0$.

Novamente, a aplicação do lema de Borel-Cantelli conclui a demonstração. ■

Usando este teorema e o Teorema 3 obtém-se a convergência quase certa do estimador normalizado, tal como afirma o próximo resultado.

Corolário 7. *Sob as condições da hipótese do teorema anterior, tem-se*

$$\forall x \in S, f_n^*(x) \xrightarrow[n \rightarrow +\infty]{q.c.} f(x).$$

Capítulo 4

Propriedades globais de convergência

A avaliação da qualidade de um estimador não paramétrico passa também por estudar o comportamento assintótico de medidas globais da discrepância entre esse estimador e a função a estimar. No contexto em que se insere o presente trabalho, as medidas mais usadas são, em geral, traduzidas à custa das distâncias ℓ_k , $1 \leq k \leq \infty$. Com efeito, dado o estimador f_n (\hat{f}_n ou f_n^*) da função de probabilidade desconhecida f , consideramos¹

$$\|f_n - f\|_k = \left(\sum_{x \in \mathcal{S}} |f_n(x) - f(x)|^k \right)^{\frac{1}{k}}, \quad 1 \leq k < \infty$$

e

$$\|f_n - f\|_\infty = \sup_{x \in \mathcal{S}} |f_n(x) - f(x)|.$$

Repare-se que, em ambos os casos, as desigualdades

$$\forall n \in \mathbb{N}, \|f_n - f\|_\infty \stackrel{q.c.}{\leq} \|f_n - f\|_1$$

e

$$\forall k > 1, \forall n \in \mathbb{N}, \|f_n - f\|_k \stackrel{q.c.}{\leq} \|f_n - f\|_1,$$

permitem que nos restrinjamos ao estudo das convergências no sentido da distância ℓ_1 .

Assim, na primeira parte deste capítulo, apresentaremos as principais propriedades de convergência da sucessão $(\|f_n - f\|_1)_{n \in \mathbb{N}}$, tendo por base os trabalhos de Kokonendji e Varron [25] e Jankowski e Wellner [19]. Por outro lado, a forma mais popular de medir a qualidade global de um estimador não paramétrico é através do erro quadrático médio integrado. Entre outras razões para essa popularidade, está o facto de ele estar na base de um dos critérios fundamentais para a escolha da janela (h_n) em termos práticos. Há, pois, todo o interesse em obter a sua expressão (exata ou assintótica), que constitui o objetivo da segunda parte do capítulo, focada no estimador não normalizado.

¹Relembremos que f_n^* é uma função de probabilidade; no caso de \hat{f}_n , estas normas estão igualmente bem definidas uma vez que, para todo o $n \in \mathbb{N}$, tem-se $0 \leq \hat{f}_n(x) \leq 1$ q.c., $x \in \mathcal{S}$, e $\|\hat{f}_n\|_k < +\infty$, q.c., $k \geq 1$, sob as nossas hipóteses de trabalho.

4.1 Convergência ℓ_1

4.1.1 Estimador não normalizado

Começamos por apresentar, para o estimador não normalizado, dois resultados relativos à convergência em média da distância ℓ_1 .

Teorema 7. *Se o núcleo associado discreto satisfaz a hipótese $HK_0(1)$, tem-se*

$$E\left(\|\widehat{f}_n - f\|_1\right) \xrightarrow{n \rightarrow +\infty} 0.$$

Além disso, se $HK_0(2)$ é verificada e $\sum_{y \in S} \eta(y) \sqrt{f(y)} < +\infty$, então

$$E\left(\|\widehat{f}_n - E(K_{\cdot, h_n}(X))\|_1\right) = O(n^{-\frac{1}{2}}). \quad (4.1)$$

Demonstração. Somando e subtraindo a esperança do estimador e efetuando a decomposição usual, iremos provar que

$$\lim_{n \rightarrow +\infty} E(\|\widehat{f}_n - E(\widehat{f}_n)\|_1) = \lim_{n \rightarrow +\infty} (\|E(\widehat{f}_n) - f\|_1) = 0.$$

Para estabelecer a segunda igualdade, vejamos que, sob a hipótese $HK_0(1)$, temos

$$\begin{aligned} \forall x \in S, \forall n \in \mathbb{N}, |E(K_{x, h_n}(X)) - f(x)| &\leq \sum_{y \in S \cap S_x} K_{x, h_n}(y) f(y) + f(x) \\ &\leq \sum_{y \in S} f(y) \eta(y) g_y(x) + f(x) := H(y, x), \end{aligned}$$

com

$$\sum_{x \in S} H(y, x) = \sum_{y \in S} f(y) \eta(y) \sum_{x \in S} g_y(x) + 1 = E(\eta(X)) + 1 < +\infty.$$

Assim, pelo teorema da convergência dominada de Lebesgue e usando o Teorema 1, vem

$$\begin{aligned} \lim_{n \rightarrow +\infty} \|E(K_{\cdot, h_n}(X)) - f\|_1 &= \lim_{n \rightarrow +\infty} \sum_{x \in S} |E(K_{x, h_n}(X)) - f(x)| \\ &= \sum_{x \in S} \lim_{n \rightarrow +\infty} |E(\widehat{f}_n(x)) - f(x)| \\ &= 0. \end{aligned}$$

Provemos agora que $\lim_{n \rightarrow +\infty} E(\|\widehat{f}_n - E(K_{\cdot, h_n}(X))\|_1) = 0$.

O teorema da convergência monótona permite escrever

$$\begin{aligned} \lim_{n \rightarrow +\infty} E\left(\|\widehat{f}_n - E(K_{\cdot, h_n}(X))\|_1\right) &= \lim_{n \rightarrow +\infty} E\left(\sum_{x \in S} |\widehat{f}_n(x) - E(\widehat{f}_n(x))|\right) \\ &= \lim_{n \rightarrow +\infty} \sum_{x \in S} E\left(|\widehat{f}_n(x) - E(\widehat{f}_n(x))|\right). \end{aligned}$$

O nosso objetivo é, novamente, aplicar o teorema da convergência dominada de Lebesgue. Para tal, notemos que, para toda a variável aleatória positiva Z , tal que $E(|Z|^2) < +\infty$, é válida a seguinte desigualdade

$$E(|Z - E(Z)|) \leq \min \left\{ 2E(Z), \sqrt{\text{Var}(Z)} \right\}.$$

Fazendo $Z := \widehat{f}_n(x)$ e atendendo a que

$$\begin{aligned} \text{Var}(\widehat{f}_n(x)) &\leq \frac{1}{n} E(K_{x,h_n}^2(X_1)) = \frac{1}{n} \sum_{y \in \mathcal{S} \cap \mathcal{S}_x} [P(Z=y)]^2 f(y) \\ &\leq \frac{1}{n} \sum_{y \in \mathcal{S} \cap \mathcal{S}_x} f(y) \\ &\leq \frac{1}{n}, \end{aligned}$$

obtemos

$$\forall n \in \mathbb{N}, \forall x \in \mathcal{S}, 0 \leq E(|\widehat{f}_n(x) - E(\widehat{f}_n(x))|) \leq \min \left\{ 2E(K_{x,h_n}(X)), \frac{1}{\sqrt{n}} \right\} \leq \frac{1}{\sqrt{n}}.$$

Além disso, $E(|\widehat{f}_n(x) - E(\widehat{f}_n(x))|) \leq 2E(K_{x,h_n}(X))$, com

$$2 \sum_{x \in \mathcal{S}} E(K_{x,h_n}(X)) \leq 2E(\eta(X)) < +\infty,$$

pela hipótese $HK_0(1)$.

Nestas condições, o referido teorema conduz a

$$\begin{aligned} \lim_{n \rightarrow +\infty} E(\|\widehat{f}_n - E(K_{\cdot,h_n}(X))\|_1) &= \sum_{x \in \mathcal{S}} \lim_{n \rightarrow +\infty} E(|\widehat{f}_n(x) - E(\widehat{f}_n(x))|) \\ &\leq \sum_{x \in \mathcal{S}} \lim_{n \rightarrow +\infty} \frac{1}{\sqrt{n}} = 0, \end{aligned}$$

Por fim, para demonstrar (4.1), começamos por escrever

$$E(\|\widehat{f}_n - E(K_{\cdot,h_n}(X))\|_1) = \sum_{x \in \mathcal{S}} E(|\widehat{f}_n(x) - E(\widehat{f}_n(x))|) \leq \sum_{x \in \mathcal{S}} \sqrt{\text{Var}(\widehat{f}_n(x))},$$

pela desigualdade de Cauchy-Schwarz.

Logo,

$$\begin{aligned} \forall n \in \mathbb{N}, E(\|\widehat{f}_n - E(K_{\cdot,h_n}(X))\|_1) &\leq \frac{1}{\sqrt{n}} \sum_{x \in \mathcal{S}} \sqrt{\text{Var}(K_{x,h_n}(X))} \\ &\leq \frac{1}{\sqrt{n}} \sum_{x \in \mathcal{S}} \left[\sum_{y \in \mathcal{S} \cap \mathcal{S}_x} K_{x,h_n}^2(y) f(y) \right]^{\frac{1}{2}}. \end{aligned}$$

Tendo em conta $HK_0(2)$, obtemos ainda

$$\begin{aligned} E\left(\|\widehat{f}_n - E(K_{\cdot, h_n}(X))\|_1\right) &\leq \frac{1}{\sqrt{n}} \sum_{x \in \mathcal{S}} \left[\sum_{y \in \mathcal{S} \cap \mathcal{S}_x} \eta^2(y) g_y^2(x) f(y) \right]^{\frac{1}{2}} \\ &\leq \frac{1}{\sqrt{n}} \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S} \cap \mathcal{S}_x} \eta(y) g_y(x) \sqrt{f(y)} \\ &\leq \frac{1}{\sqrt{n}} \sum_{y \in \mathcal{S}} \eta(y) \sqrt{f(y)} \\ &= O(n^{-\frac{1}{2}}), \end{aligned}$$

uma vez que $\sum_{y \in \mathcal{S}} \eta(y) \sqrt{f(y)} < +\infty$, por hipótese. ■

Mais uma vez, da relação entre as convergências em média e em probabilidade decorre também que $\|\widehat{f}_n - f\|_1 \xrightarrow[n \rightarrow +\infty]{P} 0$, sob as condições da hipótese da primeira parte do teorema anterior.

O próximo teorema diz respeito à convergência quase certa da mesma norma. A dedução deste resultado usa uma desigualdade exponencial enunciada no lema seguinte, que pode ser encontrada em McDiarmid [27].

Lema 5 (Desigualdade de McDiarmid). *Sejam X_1, \dots, X_n variáveis aleatórias independentes, com $P(X_i \in L_i) = 1$, para $i = 1, \dots, n$. Seja $f : L_1 \times \dots \times L_n \rightarrow \mathbb{R}$ uma função tal que*

$$|f(x_1, \dots, x_n) - f(x'_1, \dots, x'_n)| \leq c_i,$$

para todos os vetores $(x_1, \dots, x_n), (x'_1, \dots, x'_n) \in L_1 \times \dots \times L_n$ que difiram apenas na i -ésima coordenada.²

Então

$$\forall t > 0, P\left(|f(X_1, \dots, X_n) - E(f(X_1, \dots, X_n))| \geq t\right) \leq 2 \exp\left\{-2t^2 \left[\sum_{i=1}^n c_i^2\right]^{-1}\right\}.$$

Teorema 8. *Se o núcleo associado discreto satisfaz as hipóteses $HK_0(1)$ e HK_1 , tem-se*

$$\|\widehat{f}_n - f\|_1 \xrightarrow[n \rightarrow +\infty]{q.c.} 0.$$

Demonstração. Escrevendo

$$\|\widehat{f}_n - f\|_1 = \|\widehat{f}_n - f\|_1 - E(\|\widehat{f}_n - f\|_1) + E(\|\widehat{f}_n - f\|_1), \quad n \in \mathbb{N},$$

e, tendo em conta o Teorema 7, basta provar que

$$\|\widehat{f}_n - f\|_1 - E(\|\widehat{f}_n - f\|_1) \xrightarrow[n \rightarrow +\infty]{q.c.} 0.$$

²Também conhecida como propriedade das diferenças (c_1, \dots, c_n) -limitadas.

Ora $\|\widehat{f}_n - f\|_1 = g_1(X_1, \dots, X_n)$, com

$$g_1 : S^n \longrightarrow \mathbb{R}$$

$$(z_1, \dots, z_n) \longmapsto \sum_{x \in S} \left| \frac{1}{n} \sum_{i=1}^n K_{x, h_n}(z_i) - f(x) \right|$$

uma função que verifica as condições do Lema 5.

De facto, para todos os vetores (z_1, \dots, z_n) e (z'_1, \dots, z'_n) que difiram apenas na i -ésima componente, vem

$$\begin{aligned} |g_1(z_1, \dots, z_n) - g_1(z'_1, \dots, z'_n)| &\leq \sum_{x \in S} \left| \left| \frac{1}{n} \sum_{i=1}^n K_{x, h_n}(z_i) - f(x) \right| - \left| \frac{1}{n} \sum_{j=1}^n K_{x, h_n}(z'_j) - f(x) \right| \right| \\ &\leq \frac{1}{n} \sum_{x \in S} \left| \sum_{i=1}^n K_{x, h_n}(z_i) - \sum_{j=1}^n K_{x, h_n}(z'_j) \right| \\ &\leq \frac{1}{n} \sum_{x \in S} |K_{x, h_n}(z_i) - K_{x, h_n}(z'_i)| \\ &\leq 2Cn^{-1}, \end{aligned}$$

usando a hipótese HK_1 .

Portanto, a aplicação do Lema 5 conduz a

$$\forall \varepsilon > 0, \forall n \in \mathbb{N}, P\left(\left|\|\widehat{f}_n - f\|_1 - E(\|\widehat{f}_n - f\|_1)\right| \geq \varepsilon\right) \leq 2 \exp\left\{-\frac{n\varepsilon^2}{2C^2}\right\},$$

donde se deduz o resultado pretendido, pelo lema de Borel-Cantelli (cf. [7], Lema 3.14), uma vez que

$$\forall \varepsilon > 0, \sum_{n \geq 1} P\left(\left|\|\widehat{f}_n - f\|_1 - E(\|\widehat{f}_n - f\|_1)\right| \geq \varepsilon\right) \leq 2 \sum_{n \geq 1} \exp\left\{-\frac{n\varepsilon^2}{2C^2}\right\} < +\infty.$$

■

4.1.2 Estimador normalizado

No que diz respeito ao estimador normalizado, as convergências em probabilidade e quase certa da distância ℓ_1 decorrem das propriedades estabelecidas na secção anterior, usando argumentos semelhantes aos que foram referidos no Capítulo 3, a propósito dos Corolários 6 e 7. Contudo, tirando partido do facto de $f_n^*, n \in \mathbb{N}$, ser uma função de probabilidade, é possível obter os mesmos resultados como consequência direta da convergência pontual do estimador, o que nos parece ser uma abordagem interessante. Por esta razão, apresentamos seguidamente a demonstração baseada nesta abordagem. O lema que se segue, consultado em Jankowski e Wellner [19], é fundamental para tal objetivo.

Lema 6. *Sejam $g_n, n \in \mathbb{N}$ e g funções de probabilidade sobre \mathbb{Z} tais que $\lim_{n \rightarrow +\infty} g_n(x) = g(x)$, para todo o $x \in \mathbb{Z}$. Então,*

$$\lim_{n \rightarrow +\infty} \|g_n - g\|_1 = 0.$$

Demonstração. Seja $\varepsilon > 0$ fixo. Sendo g uma função de probabilidade sobre Z , existe $\alpha > 0$, tal que

$$\sum_{|x| \leq \alpha} g(x) > 1 - \frac{\varepsilon}{4}. \quad (4.2)$$

Por outro lado, como $\lim_{n \rightarrow +\infty} g_n(x) = g(x)$, para todo o $x \in \mathbb{Z}$, é garantida a existência de $N \in \mathbb{N}$ tal que

$$\forall n \geq N, \sup_{|x| \leq \alpha} |g_n(x) - g(x)| \leq \frac{\varepsilon}{4(2\alpha + 1)}.$$

Portanto, para todo o $n \geq N$ e x tal que $|x| \leq \alpha$, vem

$$|g_n(x) - g(x)| \leq \sup_{|x| \leq \alpha} |g_n(x) - g(x)| \leq \frac{\varepsilon}{4(2\alpha + 1)}.$$

Desta desigualdade deduzimos que, para $n \geq N$,

$$\sum_{|x| \leq \alpha} |g_n(x) - g(x)| \leq \frac{\varepsilon}{4(2\alpha + 1)} (2\alpha + 1) = \frac{\varepsilon}{4},$$

o que implica

$$\sum_{|x| \leq \alpha} (g(x) - g_n(x)) \leq \sum_{|x| \leq \alpha} |g_n(x) - g(x)| \leq \frac{\varepsilon}{4}.$$

Então, vem

$$\sum_{|x| \leq \alpha} g_n(x) \geq \sum_{|x| \leq \alpha} g(x) - \frac{\varepsilon}{4} \geq 1 - \frac{\varepsilon}{2},$$

de acordo com (4.2).

Consequentemente,

$$\begin{aligned} \sum_{x \in \mathbb{Z}} |g_n(x) - g(x)| &= \sum_{|x| \leq \alpha} |g_n(x) - g(x)| + \sum_{|x| > \alpha} |g_n(x) - g(x)| \\ &\leq \sum_{|x| \leq \alpha} |g_n(x) - g(x)| + \sum_{|x| > \alpha} g_n(x) + \sum_{|x| > \alpha} g(x) \\ &\leq \sum_{|x| \leq \alpha} |g_n(x) - g(x)| + \left(1 - \sum_{|x| \leq \alpha} g_n(x)\right) + \left(1 - \sum_{|x| \leq \alpha} g(x)\right) \\ &\leq \varepsilon, \end{aligned}$$

o que estabelece o resultado pretendido. ■

O lema anterior juntamente com o Corolário 7 conduzem ao seguinte teorema.³

³Evidentemente, consideramos $f_n^*(x) = f(x) = 0$, para $x \in Z \setminus S$.

Teorema 9. *Se o núcleo associado discreto verifica as hipóteses $HK_0(2)$ e HK_1 , tem-se*

$$\|f_n^* - f\|_1 \xrightarrow[n \rightarrow +\infty]{q.c.} 0.$$

Observemos que apesar de a estimação da função de repartição estar fora do âmbito desta dissertação, é de referir que as hipóteses do Lema 6 garantem ainda que

$$\lim_{n \rightarrow +\infty} \sup_{x \in \mathbb{R}} |G_n(x) - G(x)| = 0,$$

onde G_n , $n \in \mathbb{N}$, e G denotam as funções de repartição de g_n , $n \in \mathbb{N}$, e g , respetivamente. Com efeito, denotando por $[x]$ o maior número inteiro inferior ou igual a $x \in \mathbb{R}$, pela definição de função de repartição de uma variável discreta, tem-se

$$\begin{aligned} \forall n \in \mathbb{N}, \forall x \in \mathbb{R}, |G_n(x) - G(x)| &= \left| \sum_{y \leq [x]} g_n(y) - \sum_{y \leq [x]} g(y) \right| \leq \sum_{y \leq [x]} |g_n(y) - g(y)| \\ &\leq \|g_n - g\|_1. \end{aligned}$$

Assim,

$$\forall n \in \mathbb{N}, \sup_{x \in \mathbb{R}} |G_n(x) - G(x)| \leq \|g_n - g\|_1 \xrightarrow[n \rightarrow +\infty]{} 0.$$

Consequentemente, sob as condições da hipótese do teorema anterior, é válido para a função de repartição definida a partir de f_n^* um resultado do tipo Glivenko-Cantelli.

4.2 Desenvolvimento assintótico do erro quadrático médio integrado

O erro quadrático médio integrado (*EQMI*) não é mais do que a extensão do *EQM* a todos os pontos de S , isto é, o *EQMI* de \hat{f}_n é definido por⁴

$$EQMI\{\hat{f}_n\} = \sum_{x \in S} Var(\hat{f}_n(x)) + \sum_{x \in S} Viés^2(\hat{f}_n(x)). \quad (4.3)$$

Não sendo fácil deduzir a sua expressão exata na grande maioria dos casos, limitamo-nos habitualmente em obter o correspondente desenvolvimento assintótico, o que constitui o objetivo desta secção. Contudo, em nossa opinião, as demonstrações apresentadas na literatura (cf. Kokonendji e Kiessé [24]) sobre este tópico carecem de rigor, o que nos levou a analisar apenas os casos particulares considerados nos quatro exemplos iniciais apresentados no Capítulo 2.

Para caso o núcleo trivial referido no Exemplo 1, como $S_x = \{x\}$, a expressão (4.3) reduz-se apenas a

$$EQMI\{\hat{f}_E\} = \sum_{x \in S} Var(\hat{f}_n(x)) = \frac{1}{n} \left(1 - \sum_{x \in S} f^2(x) \right).$$

⁴Trata-se da versão discreta do *EQMI* definido no contexto da estimação de uma função densidade de probabilidade (cf. Tenreiro [40]).

Relativamente ao núcleo definido no Exemplo 2, verificamos que, para todos os $n \in \mathbb{N}$ e $x \in S$,

$$\begin{aligned}
n\text{Var}(\widehat{f}_n(x)) &= \text{Var}(K_{x,h_n}(X_1)) \\
&= \sum_{y \in S_x} K_{x,h_n}^2(y) f(y) - \left[\sum_{y \in S_x} K_{x,h_n}(y) f(y) \right]^2 \\
&= (1-h_n)^2 f(x) + \frac{h_n^2}{(c-1)^2} \sum_{y \in S_x \setminus \{x\}} f(y) - \left[(1-h_n)f(x) + \frac{h_n}{c-1} \sum_{y \in S_x \setminus \{x\}} f(y) \right]^2 \\
&= (1-h_n)^2 f(x) + \frac{h_n^2}{(c-1)^2} (1-f(x)) - (1-h_n)^2 f^2(x) - \frac{h_n^2}{(c-1)^2} (1-f(x))^2 \\
&\quad - 2 \frac{h_n(1-h_n)}{c-1} f(x)(1-f(x)) \\
&= f(x)(1-f(x)) \left((1-h_n)^2 + \frac{h_n^2}{(c-1)^2} - 2 \frac{h_n(1-h_n)}{c-1} \right) \\
&= f(x)(1-f(x)) \left(1 - \frac{c}{c-1} h_n \right)^2,
\end{aligned}$$

logo

$$\begin{aligned}
\sum_{x \in S} \text{Var}(\widehat{f}_n(x)) &= \frac{1}{n} \left(1 - \frac{c}{c-1} h_n \right)^2 \sum_{x \in S} f(x)(1-f(x)) \\
&= \frac{1}{n} \left(1 - \sum_{x \in S} f^2(x) \right) + O_c(h_n n^{-1}).
\end{aligned}$$

No que diz respeito ao viés, tem-se

$$\begin{aligned}
\forall n \in \mathbb{N}, \forall x \in S, \text{Viés}(\widehat{f}_n(x)) &= \sum_{y \in S_x} K_{x,h_n}(y) f(y) - f(x) \\
&= (1-h_n)f(x) + \frac{h_n}{c-1} \sum_{y \in S_x \setminus \{x\}} f(y) - f(x) \\
&= -h_n f(x) + \frac{h_n}{c-1} (1-f(x)) \\
&= h_n \frac{1-cf(x)}{c-1},
\end{aligned}$$

pelo que

$$\begin{aligned}
\sum_{x \in S} \text{Viés}^2(\widehat{f}_n(x)) &= \frac{h_n^2}{(c-1)^2} \sum_{x \in S} (1-2cf(x) + c^2 f^2(x)) \\
&= \frac{h_n^2}{(c-1)^2} \left(c + c^2 \sum_{x \in S} f^2(x) - 2c \right) \\
&= \frac{ch_n^2}{(c-1)^2} \left(c \sum_{x \in S} f^2(x) - 1 \right) \\
&= O_c(h_n^2),
\end{aligned}$$

donde se conclui que (4.3) é assintoticamente igual a

$$\frac{1}{n} \left(1 - \sum_{x \in S} f^2(x) \right) + O(h_n n^{-1}) + O(h_n^2).$$

No caso dos Exemplos 3 e 4, o desenvolvimento assintótico de (4.3), por nós obtido, baseia-se na expansão de Newton, enunciada no próximo lema (cf. [4]). Admitimos que $S_x \subseteq S$, para todo o $x \in S$ e tiramos partido da simetria do núcleo em relação a x , assim como da igualdade $\text{Var}(Z) = O(h_n)$.

Lema 7. *Dada uma qualquer função limitada $h : \mathbb{Z} \rightarrow \mathbb{R}$, tem-se*

$$\forall j, l \in \mathbb{Z}, h(j+l) = \sum_{s=0}^2 \binom{l}{s} \Delta^s h(j) + R^+(j, l) \mathbb{1}_{\{l \geq 0\}} + R^-(j, l) \mathbb{1}_{\{l < 0\}},$$

com

$$R^+(j, l) = \begin{cases} 0 & , l = 0, 1, 2 \\ \sum_{s=0}^{l-3} \binom{l-s-1}{2} \Delta^3 h(j+s) & , l \geq 3 \end{cases} \quad (4.4)$$

e

$$R^-(j, l) = - \sum_{s=1}^{-l} \binom{l+s-1}{2} \Delta^3 h(j-s), \quad (4.5)$$

com as seguintes convenções:

- $\binom{n}{r} = 0$ se $r < 0$ ou $r > n$, $n \in \mathbb{N}_0$,
- $\binom{-n}{r} = \frac{1}{r!} (-n)(-n-1) \cdots (-n-r+1) = (-1)^r \binom{n+r-1}{r}$, $n \in \mathbb{N}$,
- $\Delta f(x) = f(x+1) - f(x)$, $\Delta^n f(x) = \Delta(\Delta^{n-1} f(x))$, $n \in \mathbb{N}$, onde $\Delta^0 f(x) = f(x)$.

Em relação ao Exemplo 3, o resultado decorre da proposição seguinte, de índole mais geral.

Proposição 1. *Seja $\{K_{x, h_n}, n \in \mathbb{N}\}$ um núcleo associado discreto simétrico em relação a x , de suporte $S_x = \{x, x \pm 1, \dots, x \pm m\}$, $m \in \mathbb{N}$, com $S_x \subseteq S$. Se $\text{Var}(Z) = O(h_n)$, a expressão assintótica de (4.3) é dada por*

$$EQMI\{\hat{f}_n\} = \frac{1}{n} \left(1 - \sum_{x \in S} f^2(x) \right) + O(h_n n^{-1}) + O(h_n^2).$$

Demonstração. Primeiramente, a simetria do núcleo associado discreto em relação a x permite escrever: $K_{x, h_n}(y) := K_{h_n}(y-x)$, $(x, y) \in S \times S_x$, onde $\{K_{h_n}, n \in \mathbb{N}\}$ é uma família de funções de probabilidade com suporte $\{0, \pm 1, \dots, \pm m\}$, $m \in \mathbb{N}$.

Atendendo a esta relação, vem, para $n \in \mathbb{N}$ e $x \in S$ arbitrariamente fixos,

$$\begin{aligned} E(\widehat{f}_n(x)) &= \sum_{y \in S_x} K_{h_n}(y-x)f(y) \\ &= K_{h_n}(0)f(x) + \sum_{\{y \in S_x: y \neq x\}} K_{h_n}(y-x)f(y) \\ &= K_{h_n}(0)f(x) + \sum_{i=1}^m K_{h_n}(i)(f(x+i) + f(x-i)). \end{aligned}$$

Aplicando a expansão de Newton a $f(x+i)$ e $f(x-i)$, obtemos

$$\begin{aligned} E(\widehat{f}_n(x)) &= K_{h_n}(0)f(x) + \sum_{i=1}^m K_{h_n}(i)(2f(x) + i^2\Delta^2 f(x) + R_n^+(x, i) + R_n^-(x, -i)) \\ &= K_{h_n}(0)f(x) + 2f(x) \sum_{i=1}^m K_{h_n}(i) + \Delta^2 f(x) \sum_{i=1}^m i^2 K_{h_n}(i) + A_n(x) \\ &= f(x) + \frac{1}{2}\Delta^2 f(x)Var(Z) + A_n(x), \end{aligned}$$

com

$$A_n(x) = \sum_{i=1}^m K_{h_n}(i)(R_n^+(x, i) + R_n^-(x, -i)),$$

onde $R_n^+(x, i)$ e $R_n^-(x, -i)$ são dados, respetivamente, por (4.4) e (4.5).

Consequentemente,

$$Viés(\widehat{f}_n(x)) = \frac{1}{2}\Delta^2 f(x)Var(Z) + A_n(x).$$

Resta estudar o comportamento assintótico do último termo.

Ora, para todos os $i \in \{1, \dots, m\}$, temos

$$\begin{aligned} |R_n^+(x, i) + R_n^-(x, -i)| &\leq |R_n^+(x, i)| + |R_n^-(x, -i)| \\ &\leq \frac{1}{2} \sum_{s=0}^{i-3} (i-s-1)(i-s-2) |\Delta^3 f(x+s)| \\ &\quad + \frac{1}{2} \sum_{s=1}^i |(s-i-1)(s-i-2)| |\Delta^3 f(x-s)| \\ &\leq \frac{1}{2}(i+1)^2 g_i(x), \end{aligned}$$

com

$$g_i(x) = \sum_{s=0}^{i-3} |\Delta^3 f(x+s)| + \sum_{s=1}^i |\Delta^3 f(x-s)|,$$

o que, majorando $(i+1)^2$, conduz a

$$|A_n(x)| \leq \frac{1}{2}(m+1)^2 \sum_{i=1}^m K_{h_n}(i)g_i(x). \quad (4.6)$$

Deste modo,

$$\begin{aligned} \sum_{x \in S} \text{Viés}^2(\widehat{f}_n(x)) &= \sum_{x \in S} \left(\frac{1}{2} \Delta^2 f(x) \text{Var}(Z) + A_n(x) \right)^2 \\ &= \frac{1}{4} \text{Var}^2(Z) \sum_{x \in S} [\Delta^2 f(x)]^2 + \text{Var}(Z) \sum_{x \in S} \Delta^2 f(x) A_n(x) + \sum_{x \in S} A_n^2(x). \end{aligned}$$

Mas, de (4.6), vem

$$\begin{aligned} A_n^2(x) &\leq \frac{1}{4} (m+1)^4 \left(\sum_{i=1}^m K_{h_n}(i) g_i(x) \right)^2 \\ &\leq \frac{1}{4} (m+1)^4 \left(\sum_{i=1}^m K_{h_n}^2(i) \right) \left(\sum_{i=1}^m g_i^2(x) \right), \end{aligned}$$

pela desigualdade de Cauchy-Schwarz.

Atendendo a que

$$K_{h_n}(0) = K_{x, h_n}(x) = P(Z = x),$$

de (2.4) e (2.5), juntamente com a desigualdade de Markov, vem

$$1 - K_{h_n}(0) = P(|Z - x| \geq \delta_x) \leq \delta_x^{-2} \text{Var}(Z) = O(h_n),$$

com $\delta_x > 0$.

Assim, para todo $1 \leq i \leq m$, vem

$$K_{h_n}(i) \leq \sum_{i=1}^m K_{h_n}(i) = \frac{1}{2} (1 - K_{h_n}(0)) \leq O(h_n),$$

pelo que

$$K_{h_n}^2(i) \leq O(h_n^2).$$

Logo,

$$\sum_{i=1}^m K_{h_n}^2(i) \leq O(h_n^2).$$

Portanto,

$$\sum_{x \in S} A_n^2(x) \leq \frac{1}{4} (m+1)^4 O(h_n^2) \sum_{i=1}^m \sum_{x \in S} g_i^2(x) = O(h_n^2),$$

uma vez que

$$\sum_{x \in S} g_i^2(x) < +\infty.$$

Logo,

$$\sum_{x \in S} \text{Viés}^2(\widehat{f}_n(x)) = \frac{1}{4} \text{Var}^2(Z) \sum_{x \in S} [\Delta^2 f(x)]^2 + \text{Var}(Z) \sum_{x \in S} \Delta^2 f(x) A_n(x) + O(h_n^2).$$

Por um raciocínio análogo ao anterior, temos

$$\begin{aligned} \sum_{y \in S_x} K_{h_n}^2(y-x)f(y) &= K_{h_n}^2(0)f(x) + \sum_{\{y \in S_x: y \neq x\}} K_{h_n}^2(y-x)f(y) \\ &= K_{h_n}^2(0)f(x) + \sum_{i=1}^m K_{h_n}^2(i)(f(x+i) + f(x-i)) \\ &= K_{h_n}^2(0)f(x) + \sum_{i=1}^m K_{h_n}^2(i)(2f(x) + i^2 \Delta^2 f(x) + R_n^+(x, i) + R_n^-(x, -i)) \\ &= K_{h_n}^2(0)f(x) + 2f(x) \sum_{i=1}^m K_{h_n}^2(i) + \Delta^2 f(x) \sum_{i=1}^m i^2 K_{h_n}(i) + B_n(x), \end{aligned}$$

com

$$B_n(x) = \sum_{i=1}^m K_{h_n}^2(i)(R_n^+(x, i) + R_n^-(x, -i)), \quad n \in \mathbb{N}, \quad x \in S.$$

Então,

$$\begin{aligned} n \sum_{x \in S} \text{Var}(\widehat{f}_n(x)) &= \sum_{x \in S} \left[\sum_{y \in S_x} K_{h_n}^2(y-x)f(y) - \left(\text{Viés}(\widehat{f}_n(x)) + f(x) \right)^2 \right] = \\ &= K_{h_n}^2(0) + 2 \sum_{i=1}^m K_{h_n}(i) + \sum_{i=1}^m i^2 K_{h_n}(i) \sum_{x \in S} \Delta^2 f(x) + \sum_{x \in S} B_n(x) \\ &\quad - \sum_{x \in S} f^2(x) - \sum_{x \in S} \text{Viés}^2(\widehat{f}_n(x)) - 2 \sum_{x \in S} f(x) \text{Viés}(\widehat{f}_n(x)). \end{aligned}$$

Notemos agora que

$$\sum_{x \in S} |B_n(x)| \leq \frac{1}{2}(m+1)^2 \sum_{i=1}^m K_{h_n}^2(i) \sum_{x \in S} g_i(x) = O(h_n^2),$$

pois

$$|B_n(x)| \leq \sum_{i=1}^m K_{h_n}^2(i) |R_n^+(x, i) + R_n^-(x, -i)| \leq \frac{1}{2}(m+1)^2 \sum_{i=1}^m K_{h_n}^2(i) g_i(x).$$

Além disso,

$$1 - K_{h_n}^2(0) = (1 + K_{h_n})(1 - K_{h_n}) \leq 2(1 - K_{h_n}) = O(h_n)$$

e

$$\sum_{x \in S} f(x) \text{Viés}(\widehat{f}_n(x)) = \frac{1}{2} \text{Var}(Z) \sum_{x \in S} f(x) \Delta^2 f(x) + \sum_{x \in S} f(x) A_n(x),$$

com

$$\left| \sum_{x \in S} f(x) A_n(x) \right| \leq \sum_{x \in S} |A_n(x)| \leq \frac{1}{2} (m+1)^2 \sum_{i=1}^m K_{h_n}(i) \sum_{x \in S} g_i(x) = O(h_n),$$

de acordo com a majoração (4.6).

Conjugando estes resultados, somos conduzidos a

$$\begin{aligned} n \sum_{x \in S} \text{Var}(\widehat{f}_n(x)) &= 1 - \sum_{x \in S} f^2(x) - \frac{1}{4} \text{Var}^2(Z) \sum_{x \in S} [\Delta^2 f(x)]^2 - \text{Var}(Z) \sum_{x \in S} \Delta^2 f(x) A_n(x) \\ &\quad - \text{Var}(Z) \sum_{x \in S} f(x) \Delta^2 f(x) + O(h_n) \\ &= 1 - \sum_{x \in S} f^2(x) - \text{Var}(Z) \sum_{x \in S} \Delta^2 f(x) \left[\frac{1}{4} \text{Var}(Z) \Delta^2 f(x) + f(x) + A_n(x) \right] + O(h_n), \end{aligned}$$

ou ainda

$$\begin{aligned} \sum_{x \in S} \text{Var}(\widehat{f}_n(x)) &= \frac{1}{n} \left[1 - \sum_{x \in S} f^2(x) - \text{Var}(Z) \sum_{x \in S} \Delta^2 f(x) \left(\frac{1}{4} \text{Var}(Z) \Delta^2 f(x) + f(x) + A_n(x) \right) \right] \\ &\quad + O(h_n n^{-1}). \end{aligned}$$

Em conclusão,

$$\begin{aligned} EQMI\{\widehat{f}_n\} &= \frac{1}{n} \left[1 - \sum_{x \in S} f^2(x) - \text{Var}(Z) \sum_{x \in S} \Delta^2 f(x) \left(\frac{1}{4} \text{Var}(Z) \Delta^2 f(x) + f(x) + A_n(x) \right) \right] \\ &\quad + \frac{1}{4} \text{Var}^2(Z) \sum_{x \in S} [\Delta^2 f(x)]^2 + \text{Var}(Z) \sum_{x \in S} \Delta^2 f(x) A_n(x) + O(h_n^2) + O(h_n n^{-1}) \end{aligned}$$

Para estabelecer o resultado enunciado, basta notar que $\text{Var}(Z) = O(h_n)$ e efetuar a seguinte majoração

$$\begin{aligned} \left| \text{Var}(Z) \sum_{x \in S} \Delta^2 f(x) A_n(x) \right| &\leq \text{Var}(Z) \sum_{x \in S} |\Delta^2 f(x)| |A_n(x)| \\ &\leq 2 \text{Var}(Z) (m+1)^2 \sum_{i=1}^m K_{h_n}(i) \sum_{x \in S} g_i(x) \\ &= O(h_n^2), \end{aligned}$$

tendo em conta que $\sum_{x \in S} g_i(x) < +\infty$.

■

Por fim, analisemos a construção do desenvolvimento assintótico de (4.3) para o estimador de Wang e Van Ryzin. Como o raciocínio é muito semelhante ao da demonstração da proposição anterior, destacaremos apenas os passos que permitem contornar o facto de agora trabalharmos com um suporte infinito.

Pois bem, temos que

$$\forall n \in \mathbb{N}, \forall x \in \mathbb{Z}, E(\widehat{f}_n(x)) = f(x) + \frac{1}{2}\Delta^2 f(x)\text{Var}(Z) + A_n(x),$$

onde

$$\begin{aligned} A_n(x) &= \frac{1}{2}(1-h_n) \sum_{i \in \mathbb{N}} h_n^i (R_n^+(x, i) + R_n^-(x, -i)) \\ &= \frac{1}{2}(1-h_n) \sum_{i \in \mathbb{N}} h_n^i R_n^+(x, i) + \frac{1}{2}(1-h_n) \sum_{i \in \mathbb{N}} h_n^i R_n^-(x, -i) \\ &= A_n^+(x) + A_n^-(x). \end{aligned}$$

Vejamos agora que

$$\begin{aligned} A_n^+(x) &= \frac{1}{4}(1-h_n) \sum_{i \geq 3} h_n^i \sum_{s=0}^{i-3} (i-s-2)(i-s-1)\Delta^3 f(x+s) \\ &= \frac{1}{4}(1-h_n) \sum_{j \geq 0} h_n^{j+3} \sum_{s=0}^j (j-s+1)(j-s+2)\Delta^3 f(x+s) \\ &= \frac{1}{4}(1-h_n)h_n^3 \sum_{s \geq 0} \Delta^3 f(x+s) \sum_{j \geq s} h_n^j (j-s+1)(j-s+2) \\ &= \frac{1}{4}(1-h_n)h_n^3 \sum_{s \geq 0} \Delta^3 f(x+s) \sum_{m \geq 0} h_n^{m+s} (m+1)(m+2) \\ &= \frac{1}{2}(1-h_n)^{-2} h_n^3 \sum_{s \geq 0} \Delta^3 f(x+s) h_n^s \end{aligned}$$

e

$$\begin{aligned} A_n^-(x) &= -\frac{1}{4}(1-h_n) \sum_{i \geq 1} h_n^i \sum_{s=1}^i (i-s+1)(i-s+2)\Delta^3 f(x-s) \\ &= -\frac{1}{2}(1-h_n)^{-2} \sum_{s \geq 1} h_n^s \Delta^3 f(x-s), \end{aligned}$$

pelo que

$$\begin{aligned} \sum_{x \in \mathbb{Z}} |A_n^+(x)| &\leq \frac{1}{2}(1-h_n)^{-2} h_n^3 \sum_{s \geq 0} h_n^s \sum_{x \in \mathbb{Z}} |\Delta^3 f(x+s)| \\ &\leq 4(1-h_n)^{-2} h_n^3 \sum_{s \geq 0} h_n^s \\ &= O(h_n^3) \end{aligned}$$

e

$$\sum_{x \in \mathbb{Z}} |A_n^-(x)| \leq 4(1-h_n)^{-2} \sum_{s \geq 1} h_n^s = O(h_n).$$

Desta forma, vem

$$\sum_{x \in \mathbb{Z}} |A_n(x)| = O(h_n).$$

Mais ainda,

$$\begin{aligned} |A_n^+(x)|^2 &\leq \frac{1}{4}(1-h_n)^{-4}h_n^6 \left(\sum_{s \geq 0} |\Delta^3 f(x+s)|h_n^s \right)^2 \\ &\leq \frac{1}{4}(1-h_n)^{-4}h_n^6 \left(\sum_{s \geq 0} h_n^s \right) \left(\sum_{s \geq 0} h_n^s |\Delta^3 f(x+s)|^2 \right) \\ &\leq \frac{1}{4}(1-h_n)^{-5}h_n^6 \sum_{s \geq 0} h_n^s |\Delta^3 f(x+s)|^2 \end{aligned}$$

e

$$|A_n^-(x)|^2 \leq \frac{1}{4}h_n(1-h_n)^{-5} \sum_{s \geq 1} h_n^s |\Delta^3 f(x-s)|^2,$$

pela desigualdade de Cauchy-Schwarz.

Logo,

$$\begin{aligned} \sum_{x \in \mathbb{Z}} |A_n^+(x)|^2 &\leq \frac{1}{4}(1-h_n)^{-5}h_n^6 \sum_{s \geq 0} h_n^s \sum_{x \in \mathbb{Z}} |\Delta^3 f(x+s)|^2 \\ &\leq 16(1-h_n)^{-6}h_n^6 \\ &= O(h_n^6) \end{aligned}$$

e

$$\sum_{x \in \mathbb{Z}} |A_n^-(x)|^2 \leq 16h_n^2(1-h_n)^{-6} = O(h_n^2).$$

Ora,

$$|A_n(x)|^2 \leq 2|A_n^+(x)|^2 + 2|A_n^-(x)|^2,$$

implica que

$$\sum_{x \in \mathbb{Z}} |A_n(x)|^2 = O(h_n^2).$$

Assim,

$$\sum_{x \in \mathbb{Z}} \text{Viés}^2(\widehat{f}_n(x)) = \frac{1}{4}\text{Var}^2(Z) \sum_{x \in \mathbb{Z}} [\Delta^2 f(x)]^2 + \text{Var}(Z) \sum_{x \in \mathbb{Z}} \Delta^2 f(x)A_n(x) + O(h_n^2).$$

Analisemos agora o momento de segunda ordem. Para todos os $n \in \mathbb{N}$ e $x \in \mathbb{Z}$,

$$\begin{aligned} \sum_{y \in \mathbb{Z}} K_{x, h_n}^2(y) f(y) &= \\ &= (1 - h_n)^2 f(x) + \frac{1}{2} (1 - h_n)^2 \frac{h_n^2}{1 - h_n^2} f(x) + \frac{1}{4} (1 - h_n)^2 \frac{h_n^2 (1 + h_n^2)}{(1 - h_n^2)^3} \Delta^2 f(x) + B_n(x) \\ &= f(x) \psi_1(h_n) + \frac{1}{4} \Delta^2 f(x) \psi_2(h_n) + B_n(x), \end{aligned}$$

com

$$\begin{aligned} B_n(x) &= \frac{1}{4} (1 - h_n)^2 \sum_{i \in \mathbb{N}} h_n^{2i} (R_n^+(x, i) + R_n^-(x, -i)) \\ &= \frac{1}{4} (1 - h_n)^2 \sum_{i \in \mathbb{N}} h_n^{2i} R_n^+(x, i) + \frac{1}{4} (1 - h_n)^2 \sum_{i \in \mathbb{N}} h_n^{2i} R_n^-(x, -i) \\ &= B_n^+(x) + B_n^-(x). \end{aligned}$$

Uma vez que

$$B_n^+(x) = \frac{1}{4} (1 - h_n)^{-1} (1 + h_n)^{-3} h_n^6 \sum_{s \geq 0} \Delta^3 f(x + s) h_n^{2s}$$

e

$$B_n^-(x) = -\frac{1}{4} (1 - h_n)^{-1} (1 + h_n)^{-3} \sum_{s \geq 1} \Delta^3 f(x - s) h_n^{2s}$$

obtemos

$$\sum_{x \in \mathbb{Z}} |B_n^+(x)| \leq 2(1 - h_n)^{-1} (1 + h_n)^{-3} h_n^6 \sum_{s \geq 0} h_n^{2s} = O(h_n^6)$$

e

$$\sum_{x \in \mathbb{Z}} |B_n^-(x)| \leq 2(1 - h_n)^{-1} (1 + h_n)^{-3} \sum_{s \geq 1} h_n^{2s} = O(h_n^2),$$

pelo que

$$\sum_{x \in \mathbb{Z}} |B_n(x)| = O(h_n^2).$$

Deste modo,

$$\begin{aligned} n \sum_{x \in \mathbb{Z}} \text{Var}(\widehat{f}_n(x)) &= \psi_1(h_n) + \frac{1}{4} \psi_2(h_n) \sum_{x \in \mathbb{Z}} \Delta^2 f(x) - \sum_{x \in \mathbb{Z}} f^2(x) - \sum_{x \in \mathbb{Z}} \text{Viés}^2(\widehat{f}_n(x)) \\ &\quad - 2 \sum_{x \in \mathbb{Z}} f(x) \text{Viés}(\widehat{f}_n(x)) + O(h_n^2). \end{aligned}$$

Como

$$\psi_1(h_n) = (1 - \frac{1}{2} h_n^2) (1 - h_n) (1 + h_n)^{-1} = 1 + O(h_n)$$

e

$$\sum_{x \in \mathbb{Z}} \Delta^2 f(x) = \sum_{x \in \mathbb{Z}} (f(x+2) - 2f(x+1) + f(x)) = 0,$$

concluimos que, também neste caso, o desenvolvimento assintótico do *EQMI* é dado por

$$\frac{1}{n} \left(1 - \sum_{x \in S} f^2(x) \right) + O(h_n n^{-1}) + O(h_n^2),$$

uma vez que $\text{Var}(Z) = O(h_n)$.

Capítulo 5

O estimador de Aitchison e Aitken

Neste capítulo é feito um breve estudo de simulação que compara o desempenho do estimador de Aitchison e Aitken (Exemplo 2), com o estimador empírico. Para tal, é feita a escolha prática da janela h_n , com base no *EQMI* do estimador e simulados conjuntos de amostras de diferentes tamanhos, tendo por base quatro cenários distintos. Por fim, terminamos com uma análise aos resultados obtidos. O código R usado nesta simulação pode ser consultado no Apêndice A.

5.1 Janela ótima do estimador

Tendo em conta as expressões obtidas na secção 4.2, relativamente ao estimador de Aitchison e Aitken, é possível obter um desenvolvimento exato do seu *EQMI* a partir do qual se pode exibir a janela ótima (exata) de tal estimador. Vejamos o seguinte resultado.

Teorema 10. *O EQMI do estimador de Aitchison e Aitken é dado por*

$$EQMI\{\widehat{f}_n; h_n\} = \frac{1}{n} \left(1 - \sum_{x \in S} f^2(x) \right) - \frac{h_n}{n} B_f + h_n^2 C_f + \frac{h_n^2}{n} D_f, \quad (5.1)$$

onde

$$B_f = \frac{2c}{c-1} \left(1 - \sum_{x \in S} f^2(x) \right),$$

$$C_f = \frac{c}{(c-1)^2} \left(c \sum_{x \in S} f^2(x) - 1 \right)$$

e

$$D_f = \left(\frac{c}{c-1} \right)^2 \left(1 - \sum_{x \in S} f^2(x) \right).$$

Além disso, a janela ótima no sentido do *EQMI* é dada por

$$h_n^{opt} = \frac{1}{2} \frac{B_f}{nC_f + D_f}. \quad (5.2)$$

Demonstração. De facto, para obtermos a expressão apresentada em (5.1), basta ter em consideração que

$$\begin{aligned} \sum_{x \in S} \text{Var}(\hat{f}_n(x)) &= \frac{1}{n} \left(1 - \frac{c}{c-1} h_n\right)^2 \sum_{x \in S} f(x)(1-f(x)) \\ &= \frac{1}{n} \left(1 - \frac{2c}{c-1} h_n + \left(\frac{c}{c-1}\right)^2 h_n^2\right) \left(1 - \sum_{x \in S} f^2(x)\right) \\ &= \frac{1}{n} \left(1 - \sum_{x \in S} f^2(x)\right) - \frac{h_n}{n} \frac{2c}{c-1} \left(1 - \sum_{x \in S} f^2(x)\right) + \frac{h_n^2}{n} \left(\frac{c}{c-1}\right)^2 \left(1 - \sum_{x \in S} f^2(x)\right) \\ &= \frac{1}{n} \left(1 - \sum_{x \in S} f^2(x)\right) - \frac{h_n}{n} B_f + \frac{h_n^2}{n} D_f \end{aligned}$$

e

$$\sum_{x \in S} \text{Viés}^2(\hat{f}_n(x)) = h_n^2 \frac{c}{(c-1)^2} \left(c \sum_{x \in S} f^2(x) - 1\right) = h_n^2 C_f.$$

Logo, a janela que minimiza (5.1) é dada por

$$h_n^{opt} = \frac{1}{2} \frac{B_f}{nC_f + D_f}.$$

■

Uma vez que

$$EQMI\{\hat{f}_E\} = \frac{1}{n} \left(1 - \sum_{x \in S} f^2(x)\right), \quad (5.3)$$

podemos concluir que o estimador em estudo exibirá melhores resultados que o estimador empírico, mediante uma escolha adequada da janela, conforme a proposição seguinte.

Proposição 2. Quando $0 < h_n < \frac{B_f}{nC_f + D_f}$, tem-se

$$EQMI\{\hat{f}_n; h_n\} < EQMI\{\hat{f}_E\}.$$

Demonstração. Atendendo a (5.1) e (5.3), é suficiente mostrar que

$$\forall h_n \in (0, 2h_n^{opt}), G(h_n) > 0,$$

onde

$$G(h_n) = \frac{h_n}{n} B_f - h_n^2 C_f - \frac{h_n^2}{n} D_f.$$

Como G é uma função quadrática com maximizante no ponto h_n^{opt} e $G(0) = 0$, o resultado é imediato.

■

5.2 Escolha prática da janela

Tal como na estimação de uma função densidade pelo método do núcleo, a escolha da janela h_n é também, no nosso contexto, uma questão fundamental na construção prática dos estimadores. Nos trabalhos de Kiessé [22] e Kokonendji e Kiessé [24] é sugerido o método clássico da validação cruzada. Outras propostas que encontramos na literatura são a abordagem Bayesiana (cf. Zougab et al. [45]) e, mais recentemente, um método baseado na minimização da distância em variação total entre o estimador e a função a estimar (cf. Kokonendji e Varron [25]). Dada a sua popularidade, descrevemos brevemente o método da validação cruzada, que apesar de tudo não será usado no presente trabalho. Neste método, a janela é dada por

$$h_{CV} = \arg \min_{h_n > 0} CV(h_n),$$

com

$$CV(h_n) = \sum_{x \in S} \widehat{f}_n^2(x) - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{n,-i}(X_i),$$

onde

$$\widehat{f}_{n,-i}(X_i) = \frac{1}{n-1} \sum_{j \neq i} K_{X_i, h_n}(X_j)$$

é o estimador de f construído com base nas observações $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$. Na base deste método está o facto de a função CV ser um estimador centrado de

$$EQMI\{\widehat{f}_n; h_n\} - \sum_{x \in S} f^2(x).$$

De facto, basta ter em consideração que

$$\begin{aligned} E\left(\frac{1}{n} \sum_{i=1}^n \widehat{f}_{n,-i}(X_i)\right) &= E(K_{X_1, h_n}(X_2)) = \sum_{x \in S} \sum_{y \in S} K_{x, h_n}(y) f(x) f(y) \\ &= \sum_{x \in S} E(\widehat{f}_n(x)) f(x) \\ &= E\left(\sum_{x \in S} \widehat{f}_n(x) f(x)\right) \end{aligned}$$

e que $\sum_{x \in S} \widehat{f}_n^2(x)$ é, evidentemente, um estimador centrado da sua esperança.

Além de computacionalmente pesado, o método da validação cruzada apresenta a grande desvantagem de não ser garantida a existência de um único mínimo local. Em contrapartida, o método *plug-in* para a escolha da janela é baseado na ideia de substituir as quantidades desconhecidas, que surgem na expressão da janela assintoticamente ótima¹, por estimadores convergentes.

No nosso caso, uma vez que conseguimos exibir uma expressão exata para a janela ótima h_n^{opt} , apresentada em (5.2), usaremos a ideia deste método para a construção do estimador, considerando a

¹Obtida através da minimização do erro quadrático médio integrado assintótico.

janela da forma

$$\widehat{h}_n^{opt} = \frac{1}{2} \frac{\widehat{B}_f}{n\widehat{C}_f + \widehat{D}_f},$$

com

$$\widehat{B}_f = \frac{2c}{c-1} \left(1 - \sum_{x \in S} \widehat{f}_E^2(x)\right), \widehat{C}_f = \frac{c}{(c-1)^2} \left(c \sum_{x \in S} \widehat{f}_E^2(x) - 1\right) \text{ e } \widehat{D}_f = \frac{1}{2} \frac{c}{c-1} \widehat{B}_f.$$

5.3 Alguns resultados de simulação

Nesta secção faremos um breve estudo de simulação, começando por comparar a eficiência dos estimadores através da distribuição do erro quadrático integrado (*EQI*) para diferentes tamanhos de amostra e quatro cenários gerados por misturas de distribuições binomiais com suportes $\{0, 1, \dots, m\}$, com m fixo tal que $c = m + 1$, conforme o suporte S_x do núcleo de Aitchison e Aitken. Por último, para uma análise mais global do comportamento dos estimadores usaremos, como medida de qualidade do estimador, a norma- L_2 do $EQI\{\widehat{f}_n; h_n\}$ dada por

$$\sqrt{\text{Var}(EQI\{\widehat{f}_n; h_n\}) + EQMI\{\widehat{f}_n; h_n\}}.$$

Esta medida tem em conta não só o centro da distribuição do $EQI\{\widehat{f}_n; h_n\}$ mas também a variabilidade dessa distribuição. Na prática, para cada um dos cenários em estudo usaremos uma aproximação para esta norma da forma

$$\sqrt{\frac{1}{M} \sum_{i=1}^M EQI^2(i)},$$

onde M é o número de replicações de amostras de tamanho n e $EQI(i)$ denota o *EQI* empírico do estimador com base nas janelas em estudo.

Os cenários considerados neste estudo são denotados por 1, 2, 3 e 4. O cenário 1 é gerado por uma distribuição binomial assimétrica de parâmetros 50 e 0.2; o cenário 2 é gerado por uma distribuição binomial simétrica de tamanho 50; o cenário 3 deriva de uma mistura de distribuições binomiais da forma

$$\frac{2}{5} \binom{50}{y} 0.4^y 0.6^{50-y} + \frac{3}{5} \binom{50}{y} 0.6^y 0.4^{50-y}, 0 \leq y \leq 50;$$

e o cenário 4 vem de uma mistura de distribuições binomiais dada por

$$\frac{2}{5} \binom{50}{y} 0.3^y 0.7^{50-y} + \frac{3}{5} \binom{50}{y} 0.7^y 0.3^{50-y}, 0 \leq y \leq 50.$$

Na Figura 5.1 podemos ver as representações gráficas dos quatro cenários descritos.

Nos diagramas de extremos de quartis apresentados nas Figuras 5.2, 5.3, 5.4 e 5.5, encontramos a distribuição do *EQI* dos estimadores para os cenários 1, 2, 3 e 4, respetivamente; construídos a partir de 500 amostras de tamanho n indicado. Começemos por observar que, independentemente do tamanho da amostra e do cenário considerado, a distribuição do *EQI* do estimador de Aitchison e Aitken com a janela \widehat{h}_n^{opt} apresenta sempre melhores resultados relativamente ao estimador empírico.

É também visível os bons resultados obtidos pela janela \hat{h}_n^{opt} quando comparada com os resultados (de controlo) dados pelo estimador ótimo de Aitchison e Aitken. Logo, concluímos que o estimador com a janela \hat{h}_n^{opt} se comporta de forma semelhante ao estimador ótimo de Aitchison e Aitken. Claro que, à medida que o tamanho da amostra vai aumentando, melhor será a precisão do estimador em causa, sendo mais visível essa evolução nos cenários 3 e 4.

Relativamente à Figura 5.6, temos uma análise mais global do comportamento do estimador e, tal como acontecia nos gráficos das Figuras 5.2-5.5, o comportamento do estimador em estudo é sempre melhor que o estimador empírico, para todos os cenários e todos os tamanhos de amostra. Para amostras de tamanho superior a 100, o estimador de Aitchison e Aitken com a janela \hat{h}_n^{opt} tem um bom comportamento assintótico, não havendo diferenças significativas com o estimador ótimo. No entanto, para amostras de tamanho inferior a 100, notamos uma maior diferença perante os dois cenários com mistura de binomiais, relativamente ao estimador ótimo. Desta forma, é de salientar a qualidade da janela \hat{h}_n^{opt} . Novamente, a construção destes últimos quatro gráficos foram baseados em 500 amostras de tamanho n .

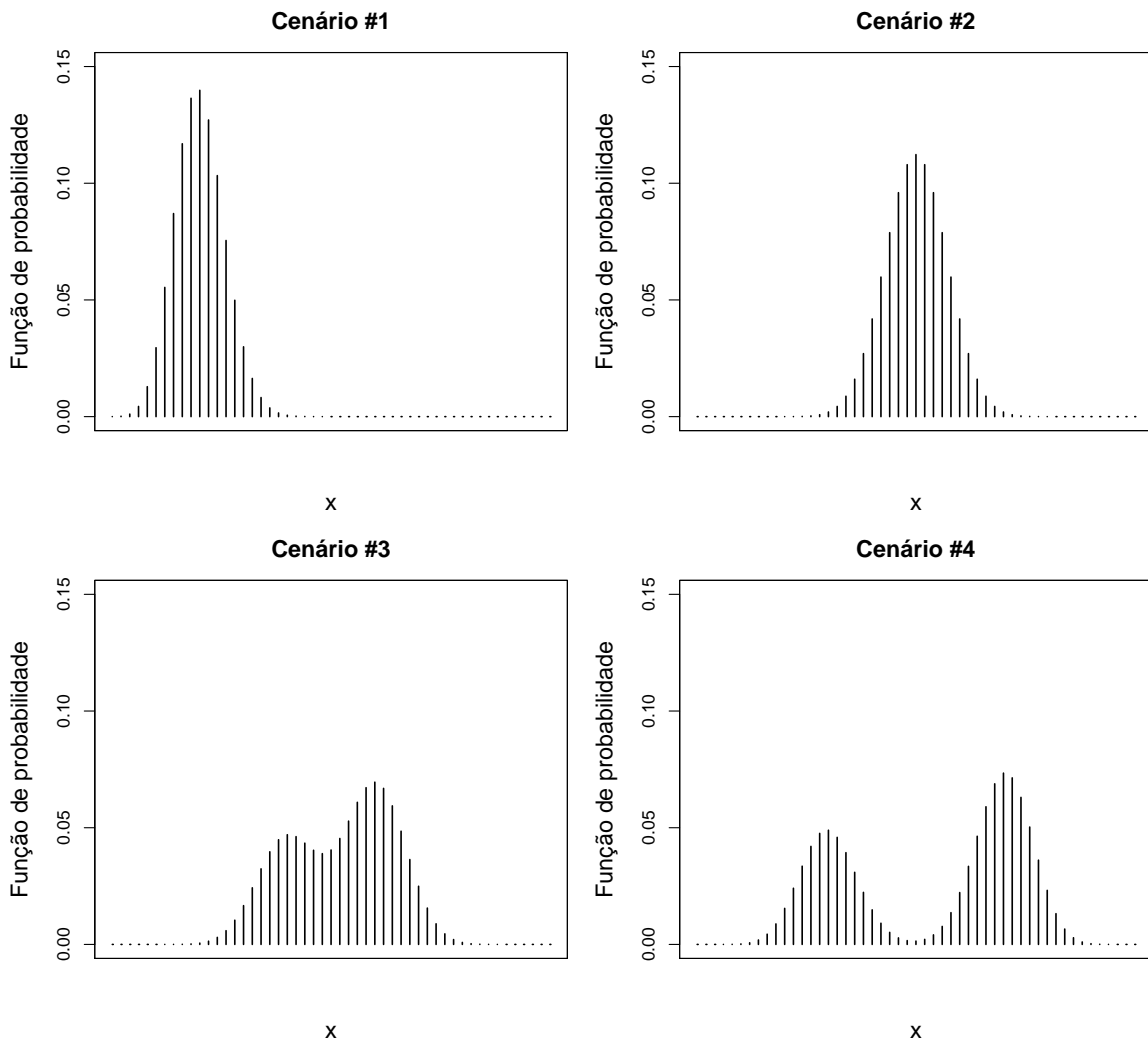


Figura 5.1 Representações gráficas das funções de probabilidade usadas nos cenários em estudo.

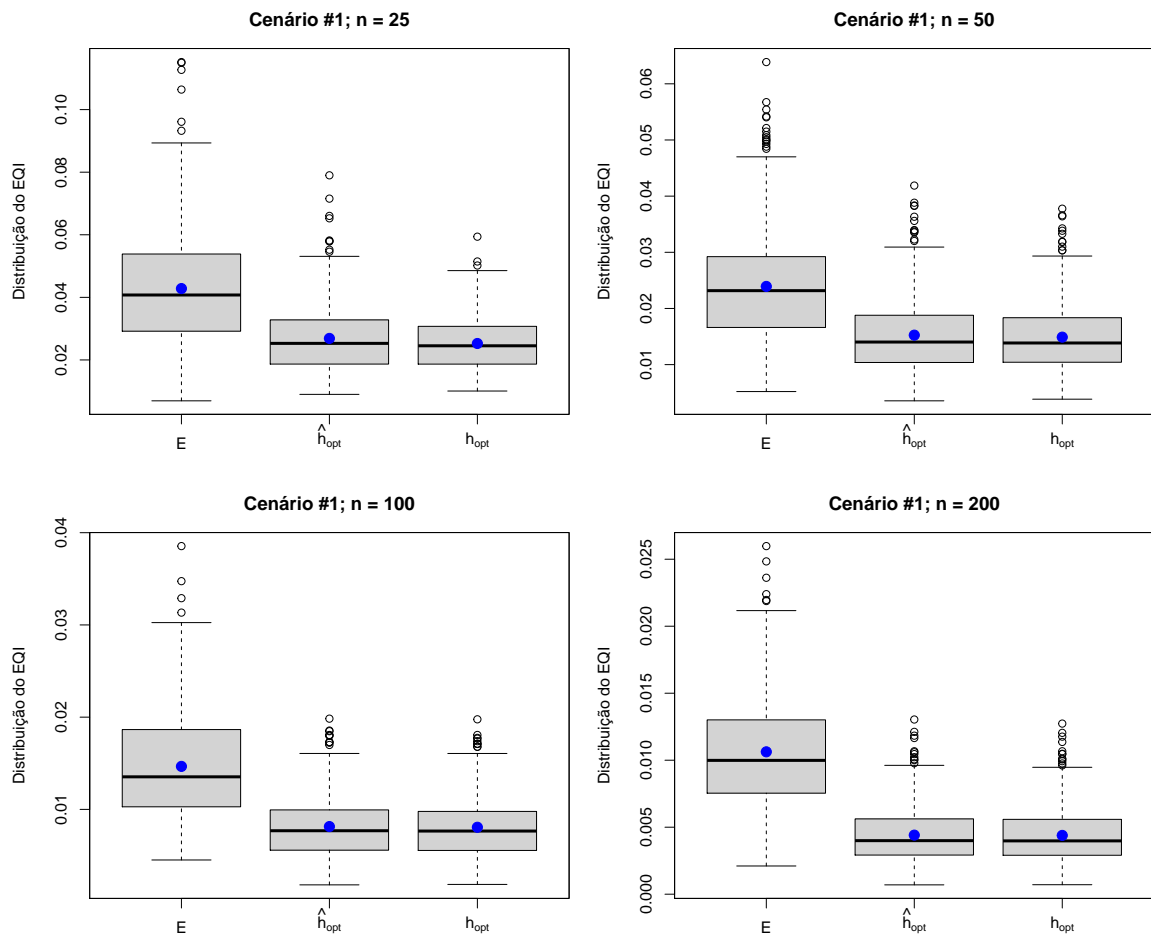


Figura 5.2 Diagramas de extremos e quartis das distribuições dos EQI com base em amostras com distribuição no cenário 1.

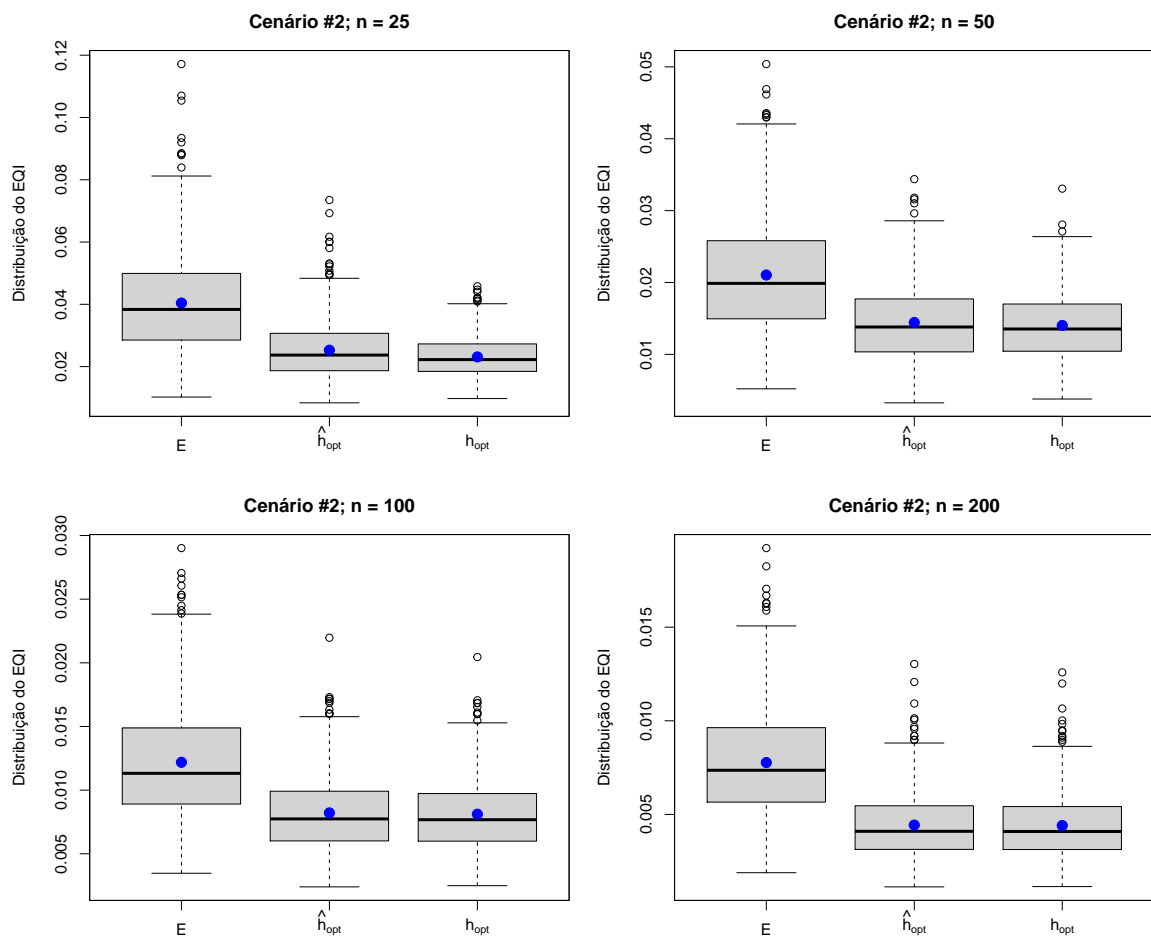


Figura 5.3 Diagramas de extremos e quartis das distribuições dos EQI com base em amostras com distribuição no cenário 2.

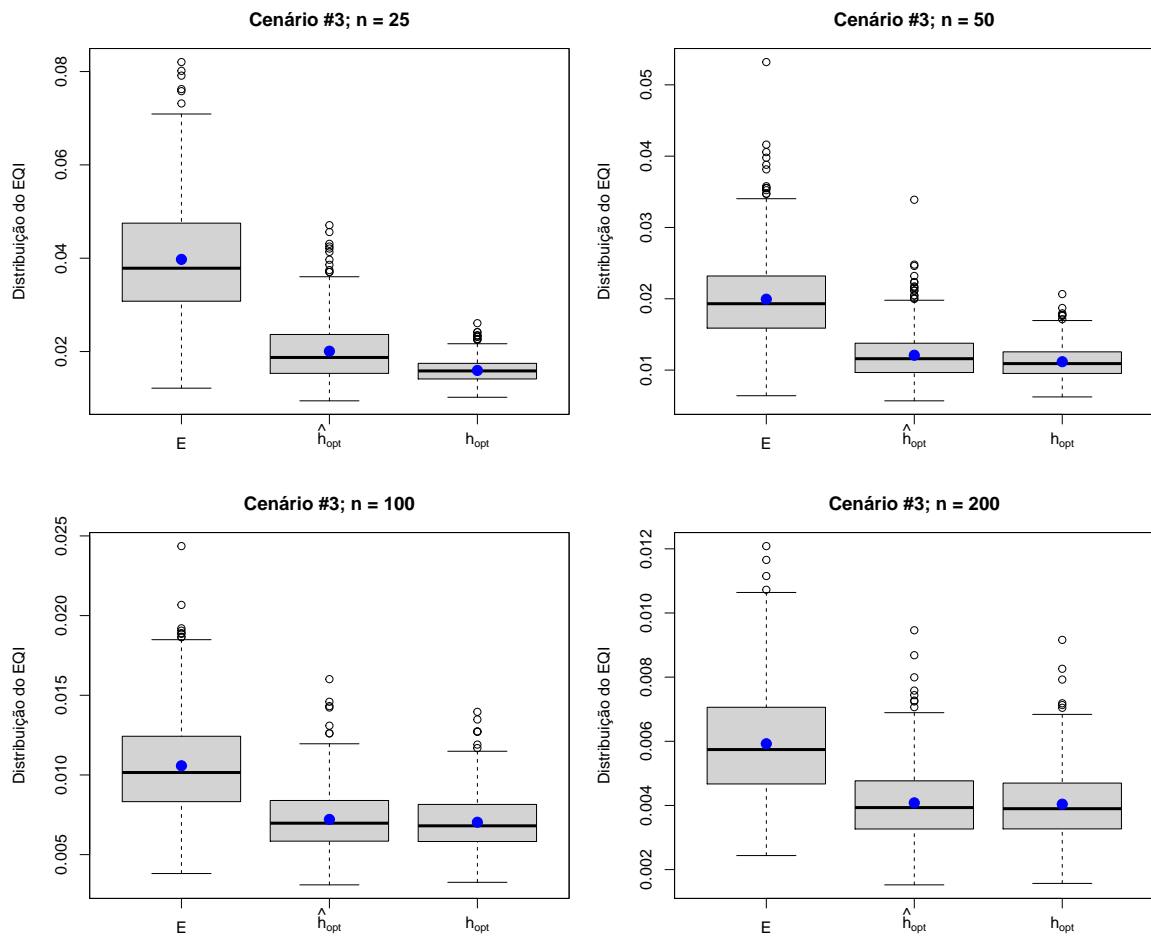


Figura 5.4 Diagramas de extremos e quartis das distribuições dos EQI com base em amostras com distribuição no cenário 3.

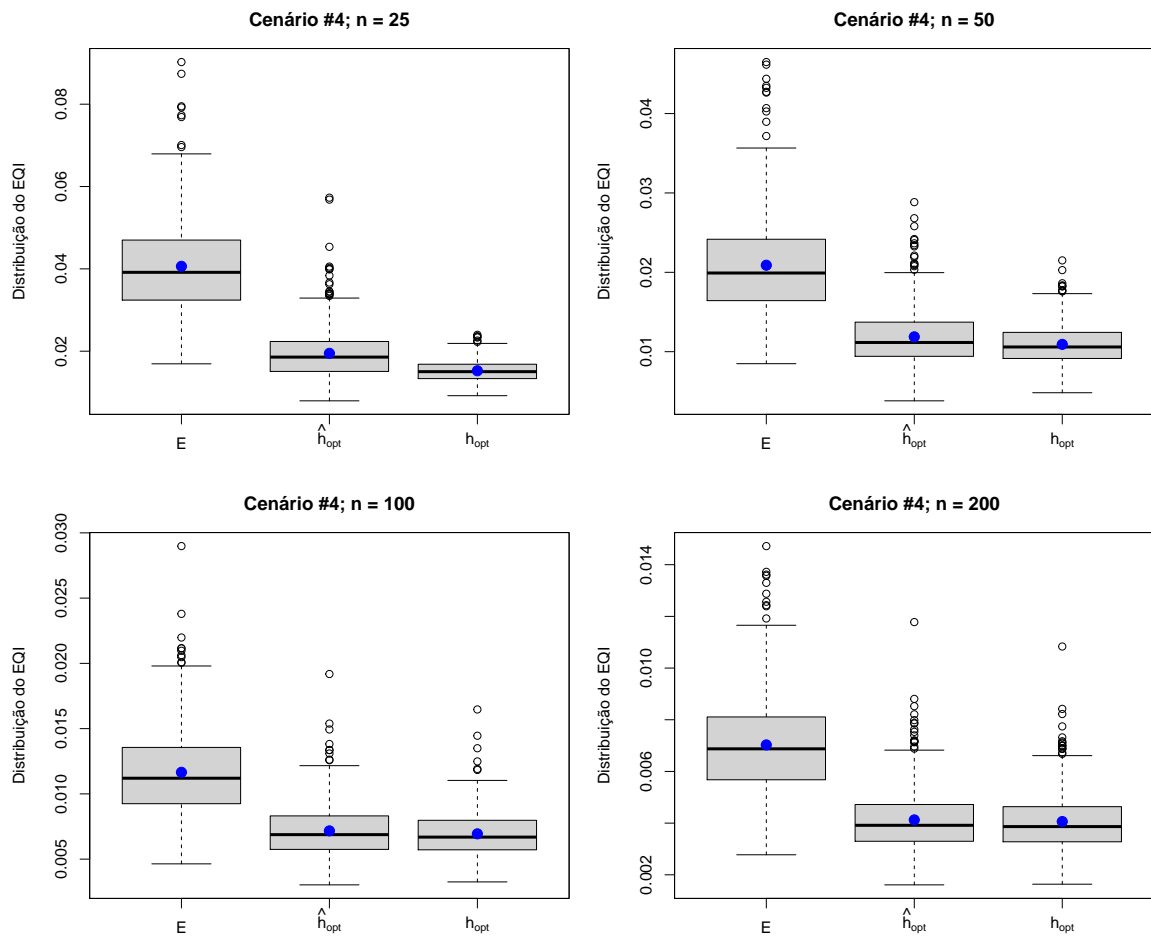


Figura 5.5 Diagramas de extremos e quartis das distribuições dos EQI com base em amostras com distribuição no cenário 4.

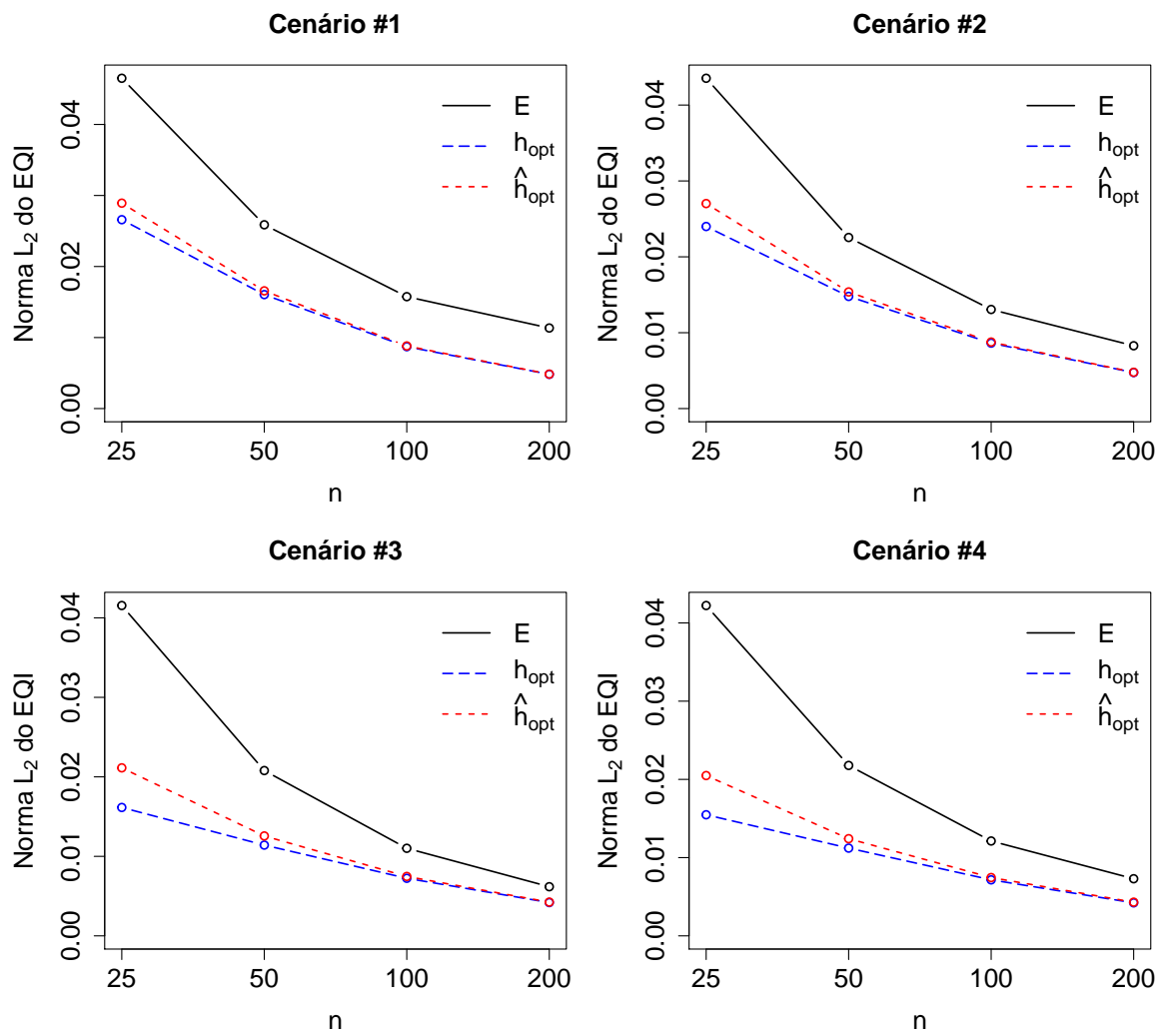


Figura 5.6 Aproximações da norma- L_2 do EQI perante os vários cenários em estudo.

Capítulo 6

Considerações finais

O tema do presente trabalho, sendo relativamente recente, permanece ainda objeto de investigação para muitos autores e tem sido essencialmente focado em dois tópicos: a generalização do método do núcleo ao caso multidimensional e a aplicação de métodos Bayesianos para a determinação da janela (cf. Belaid et al. [5], Zougab et al. [44]). Mesmo no caso unidimensional, são vários os aspetos que, em nossa opinião, permanecem em aberto, de entre os quais salientamos: desenvolver o estudo do núcleo apresentado no Exemplo 5, já iniciado por Huang et al. [18], pois a definição implícita (e complexa) do parâmetro λ nele interveniente não permitiu aprofundar as propriedades do estimador correspondente, em particular no que diz respeito ao *EQMI*; procurar, à semelhança de Esstafa et al. [12], condições menos restritivas que assegurem a convergência de $(C_n)_{n \in \mathbb{N}}$ e, conseqüentemente, as propriedades de consistência do estimador normalizado; procurar hipóteses adequadas sobre o núcleo ou a função a estimar de modo a obter o desenvolvimento assintótico do *EQMI* no caso geral; comparar, através de estudos de simulação, os métodos de escolha da janela h_n propostos na literatura, com ênfase para o método da validação cruzada; comparar, em termos práticos, o desempenho dos estimadores construídos a partir dos diferentes núcleos perante dados reais.

Finalmente, não podemos deixar de lamentar o facto de encontrarmos publicados ou pré-publicados resultados que nos parecem manifestamente incorretos, como é o caso do desenvolvimento assintótico do *EQMI* apresentado por Kokonendji e Kiessé [24], assim como as demonstrações das proposições 2.1 e 2.5 que figuram em Esstafa [12].

Bibliografia

- [1] Abdous, B. and Kokonendji, C. C. (2009). Consistency and asymptotic normality for discrete associated-kernel estimator. *African Diaspora Journal of Mathematics*, 8(2):63–70.
- [2] Ahmad, I. A. and Cerrito, P. B. (1994). Nonparametric estimation of joint discrete-continuous probability densities with applications. *Journal of Statistical Planning and Inference*, 41(3):349–364.
- [3] Aitchison, J. and Aitken, C. G. (1976). Multivariate binary discrimination by the kernel method. *Biometrika*, 63(3):413–420.
- [4] Barbour, A. D. and Čekanavičius, V. (2002). Total variation asymptotics for sums of independent integer random variables. *The Annals of Probability*, 30(2):509–545.
- [5] Belaid, N., Adjabi, S., Zougab, N., and Kokonendji, C. C. (2016). Bayesian bandwidth selection in discrete multivariate associated kernel estimators for probability mass functions. *Journal of the Korean Statistical Society*, 45(4):557–567.
- [6] Bosq, D. and Lecoutre, J.-P. (1987). *Théorie de l'estimation fonctionnelle*. Collection économie et statistiques avancées. *Economica, Paris*.
- [7] Breiman, L. P. (1992). Probability, Classics Appl. Math. 7, Philadelphia. *Society for Industrial and Applied Mathematics*.
- [8] Burman, P. (1987). Smoothing sparse contingency tables. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 24–36.
- [9] Chu, C.-Y., Henderson, D. J., and Parmeter, C. F. (2017). On discrete epanechnikov kernel functions. *Computational statistics & data analysis*, 116:79–105.
- [10] Daly, F. and Gaunt, R. E. (2015). The conway-maxwell-poisson distribution: distributional theory and approximation. *Preprint arXiv:1503.07012*.
- [11] Dong, J. and Simonoff, J. S. (1994). The construction and properties of boundary kernels for smoothing sparse multinomials. *Journal of Computational and Graphical Statistics*, 3(1):57–66.
- [12] Esstafa, Y., Kokonendji, C. C., and Somé, S. M. (2022). Asymptotic properties of the normalized discrete associated-kernel estimator for probability mass function. *Preprint arXiv:2202.10078*.
- [13] Gramacki, A. (2018). *Nonparametric kernel density estimation and its computational aspects*. Cham Springer International Publishing, Springer.
- [14] Hall, P. and Titterton, D. M. (1987). On smoothing sparse multinomial data. *Australian Journal of Statistics*, 29(1):19–37.
- [15] Harfouche, L., Adjabi, S., Zougab, N., and Funke, B. (2018). Multiplicative bias correction for discrete kernels. *Statistical Methods & Applications*, 27(2):253–276.

- [16] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30.
- [17] Huang, A. (2017). Mean-parametrized conway–maxwell–poisson regression models for dispersed counts. *Statistical Modelling*, 17(6):359–380.
- [18] Huang, A., Sippel, L., and Fung, T. (2022). Consistent second-order discrete kernel smoothing using dispersed conway–maxwell–poisson kernels. *Computational Statistics*, 37(2):551–563.
- [19] Jankowski, H. K. and Wellner, J. A. (2009). Estimation of a discrete monotone distribution. *Electronic journal of statistics*, 3:1567.
- [20] Jones, W. M. (1995). Kernel Smoothing. Monographs on Statistics and Applied, Probability, Chapman & hall.
- [21] Karunamuni, R. J. and Alberts, T. (2005). On boundary correction in kernel density estimation. *Statistical Methodology*, 2(3):191–212.
- [22] Kiessé, T. S. (2008). *Approche non-paramétrique par noyaux associés discrets des données de dénombrement*. PhD thesis, Université de Pau et des Pays de l’Adour.
- [23] Kokonendji, C., Senga Kiessé, T., and Zocchi, S. S. (2007). Discrete triangular distributions and non-parametric estimation for probability mass function. *Journal of Nonparametric Statistics*, 19(6-8):241–254.
- [24] Kokonendji, C. C. and Kiessé, T. S. (2011). Discrete associated kernels method and extensions. *Statistical Methodology*, 8(6):497–516.
- [25] Kokonendji, C. C. and Varron, D. (2016). Performance of discrete associated kernel estimators through the total variation distance. *Statistics & Probability Letters*, 110:225–235.
- [26] Kokonendji, C. C. and Zocchi, S. S. (2010). Extensions of discrete triangular distributions and boundary bias in kernel estimation for discrete functions. *Statistics & Probability Letters*, 80(21-22):1655–1662.
- [27] McDiarmid, C. (1998). Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, pages 195–248. Springer.
- [28] Ouyang, D., Li, Q., and Racine, J. (2006). Cross-validation and the estimation of probability distributions with categorical data. *Journal of Nonparametric Statistics*, 18(1):69–100.
- [29] O’Neill, B. (2022). Smallest covering regions and highest density regions for discrete distributions. *Computational Statistics*, pages 1–26.
- [30] Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076.
- [31] Pollard, D. (2012). *Convergence of stochastic processes*. Springer Science & Business Media.
- [32] Racine, J. S., Li, Q., and Yan, K. X. (2020). Kernel smoothed probability mass functions for ordered datatypes. *Journal of Nonparametric Statistics*, 32(3):563–586.
- [33] Rajagopalan, B. and Lall, U. (1995). A kernel estimator for discrete distributions. *Journaltitle of Nonparametric Statistics*, 4(4):409–426.
- [34] Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, 27:832–837.
- [35] Sedgewick, R. and Flajolet, P. (2013). *An introduction to the analysis of algorithms*. Pearson Education India.

- [36] Sellers, K. F., Swift, A. W., and Weems, K. S. (2017). A flexible distribution class for count data. *Journal of Statistical Distributions and Applications*, 4(1):1–21.
- [37] Senga Kiessé, T. (2017). On finite sample properties of nonparametric discrete asymmetric kernel estimators. *Statistics*, 51(5):1046–1060.
- [38] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.
- [39] Simonoff, J. S. (2012). *Smoothing methods in statistics*. Springer Science & Business Media.
- [40] Tenreiro, C. (2010). *Uma introdução à estimação não-paramétrica de densidade*. In *XVIII Congresso Anual da Sociedade Portuguesa de Estatística*.
- [41] Titterington, D. (1980). A comparative study of kernel-based density estimates for categorical data. *Technometrics*, 22(2):259–268.
- [42] Titterington, D. and Bowman, A. (1985). A comparative study of smoothing procedures for ordered categorical data. *Journal of Statistical Computation and Simulation*, 21(3-4):291–312.
- [43] Wang, M.-C. and Van Ryzin, J. (1981). A class of smooth estimators for discrete distributions. *Biometrika*, 68(1):301–309.
- [44] Zougab, N., Adjabi, S., and Kokonendji, C. (2013a). Adaptive smoothing in associated kernel discrete functions estimation using bayesian approach. *Journal of Statistical Computation and Simulation*, 83(12):2219–2231.
- [45] Zougab, N., Adjabi, S., and Kokonendji, C. C. (2013b). A bayesian approach to bandwidth selection in univariate associate kernel estimation. *Journal of Statistical Theory and Practice*, 7(1):8–23.

Apêndice A

Códigos R

```
m=50
c <- m+1

# Cenario 1

pf1 = function(x)
{
  dbinom(x, size=m, prob=0.2)
}

r1 = function(n)
{
  rbinom(n, size=m, prob=0.2)
}

# Cenario 2

pf2 = function(x)
{
  dbinom(x, size=m, prob=0.5)
}

r2 = function(n)
{
  rbinom(n, size=m, prob=0.5)
}

# Cenario 3

pf3 = function(x)
{
  w <- 2/5
  w*dbinom(x, size=m, prob=0.4)+(1-w)*dbinom(x, size=m, prob=0.6)
}

r3 = function(n)
{
```

```

    u <- runif(n)
    w <- 2/5
    (u<w)*rbinom(n, size=m, prob=0.4)+(u>=w)*rbinom(n, size=m, prob=0.6)
  }

# Cenario 4

pf4 = function(x)
{
  w <- 2/5
  w*dbinom(x, size=m, prob=0.3)+(1-w)*dbinom(x, size=m, prob=0.7)
}

r4 = function(n)
{
  u <- runif(n)
  w <- 2/5
  (u<w)*rbinom(n, size=m, prob=0.3)+(u>=w)*rbinom(n, size=m, prob=0.7)
}

# Guarda os graficos dos cenarios em pdf

graph.ps = function(d)
{
  if (d==1) {pfd <- function(x){return(pf1(x))}}
  if (d==2) {pfd <- function(x){return(pf2(x))}}
  if (d==3) {pfd <- function(x){return(pf3(x))}}
  if (d==4) {pfd <- function(x){return(pf4(x))}}

  pdf(paste("graf",d,".pdf",sep=""),width=5.5,height=5)
  par(mar=c(4,4,3,1))
  t <- 0:m
  y <- pfd(t)
  plot(t,y,type='h',lwd=1.5,xaxt='n',ylim=c(0,0.15),xlab="x",ylab="Funcao de
    probabilidade",main=paste("Cenario #",d,sep=""),cex.main=1.3,cex.lab=1.3)
  box()
  dev.off()
}

# Nucleo do Exemplo 2

K = function(x,z,h,c=m+1)
{
  ((1-h)*(z==x)+(h/(c-1))*(z!=x))*(z>=0)*(z<=(c-1))
}

# Estimador fn

fn_K = function(s,x,h,c=m+1)
{
  n <- length(s)
  kk <- NULL

```

```

    for (i in 1:n)
      kk[i] <- K(x=x, z=s[i], h=h, c=c)
    mean(kk)
  }

# Estimador empirico

fn_E = function(s, x)
{
  n <- length(s)
  kk <- NULL
  for (i in 1:n)
    kk[i] <- 1*(s[i]==x)
  mean(kk)
}

# EQI

ISE = function(n, rep, d, m)
{
  if (d==1) {pfd <- function(x){return(pf1(x))}; rd <- function(n){return(r1(n))}}
  if (d==2) {pfd <- function(x){return(pf2(x))}; rd <- function(n){return(r2(n))}}
  if (d==3) {pfd <- function(x){return(pf3(x))}; rd <- function(n){return(r3(n))}}
  if (d==4) {pfd <- function(x){return(pf4(x))}; rd <- function(n){return(r4(n))}}

  ises <- array(dim=c(rep, 3))
  set.seed(346762)

  Bf <- (2*c/(c-1))*(1-sum(pfd(0:m)^2))
  Cf <- (c/(c-1)^2)*(c*sum(pfd(0:m)^2)-1)
  Df <- (c/(c-1))*Bf/2

  estE <- array(dim=m+1)

  for (j in 1:rep)
  {
    s <- rd(n)

    for (i in 1:(m+1))
      estE[i] <- fn_E(s, i)

    hat.Bf <- (2*c/(c-1))*(1-sum(estE^2))
    hat.Cf <- (c/(c-1)^2)*(c*sum(estE^2)-1)
    hat.Df <- (c/(c-1))*hat.Bf/2

    # Empirico
    soma <- 0
    for (i in 0:m)

```

```

    soma <- soma + (estE[i+1]-pfd(i))^2
    ises[j,1] <- soma

    hat.h.opt <- hat.Bf/(2*(n*hat.Cf+hat.Df))
    soma <- 0
    for (i in 0:m)
      soma <- soma + (fn_K(s,i,h=hat.h.opt,c=m+1)-pfd(i))^2
    ises[j,2] <- soma

    # Controlo
    h.opt <- Bf/(2*(n*Cf+Df))
    soma <- 0
    for (i in 0:m)
      soma <- soma + (fn_K(s,i,h=h.opt,c=m+1)-pfd(i))^2
    ises[j,3] <- soma

  }

  if (n<100) ni <- paste("0",n,sep="") else ni <- n
  write.table(ises, file=paste("ISEd",d,"n",ni,".tex",sep=""))
}

# BoxPlot's

grafico.boxplot <- function(d,n)
{
  if (n<100) ni <- paste("0",n,sep="") else ni <- n
  ises <- read.table(paste("ISEd",d,"n",ni,".tex",sep=""))
  pdf(paste("boxplotsd",d,"n",ni,".pdf",sep=""), width=6,height=5)
  par(mar=c(3,4.1,3,1)) #baixo, esquerda, cima, direita
  boxplot(ises,
    main=paste("Cenario #",d,"; n = ",n,sep=""),
    ylab=expression(D*i*s*t*r*i*b*u*i*c*a*o~d*o~E*Q*I), names=c(expression(E),
      expression(hat(h)[opt]), expression(h[opt])))
  points(x=c(1,2,3), y=colMeans(ises), pch=19,col="blue",cex=1.4)
  box()
  dev.off()
}

# Graficos da norma L2

grafico.ns <- function(d)
{
  ises025 <- read.table(paste("ISEd",d,"n025.tex",sep=""))
  ises050 <- read.table(paste("ISEd",d,"n050.tex",sep=""))
  ises100 <- read.table(paste("ISEd",d,"n100.tex",sep=""))
  ises200 <- read.table(paste("ISEd",d,"n200.tex",sep=""))
  L2.ises025 <- sqrt(colMeans(ises025^2))
  L2.ises050 <- sqrt(colMeans(ises050^2))
  L2.ises100 <- sqrt(colMeans(ises100^2))
  L2.ises200 <- sqrt(colMeans(ises200^2))
}

```

```

L2.Emp      <- c(L2.ises025 [1],L2.ises050 [1],L2.ises100 [1],L2.ises200 [1])
L2.hat.h.opt <- c(L2.ises025 [2],L2.ises050 [2],L2.ises100 [2],L2.ises200 [2])
L2.h.opt    <- c(L2.ises025 [3],L2.ises050 [3],L2.ises100 [3],L2.ises200 [3])
maximo <- max(c(L2.Emp,L2.h.opt,L2.hat.h.opt))
n <- c(25,50,100,200)

pdf(paste("grafnsd",d,".pdf",sep=""),width=5.5,height=5)
par(mar=c(4,5,4,0.5)) #baixo, esquerda, cima, direita
plot(1:length(n),L2.Emp,type="b",lwd=1.5,lty=1, #log="y",
main=paste("Cenario #",d,sep=""),cex.main=1.5,
xlab="n",ylab=expression(N*o*r*m*a~L[2]~d*o~E*Q*I),axes=FALSE,
cex.lab=1.5,cex.axis=1.5,ylim=c(0,maximo))
axis(1,at=1:length(n),labels=n,cex.axis=1.5)
axis(2,cex.axis=1.5)
lines(1:length(n),L2.h.opt,type="b",lwd=1.5,lty=5,col="blue")
lines(1:length(n),L2.hat.h.opt,type="b",lwd=1.5,lty=2,col="red")
legend("topright",leg=c(expression(E),expression(h[opt]),expression(hat(h)[
  opt])),
lty=c(1,5,2),cex=1.5,lwd=1.5,bty="n",col=c("black","blue","red"))
box()
dev.off()
}

```