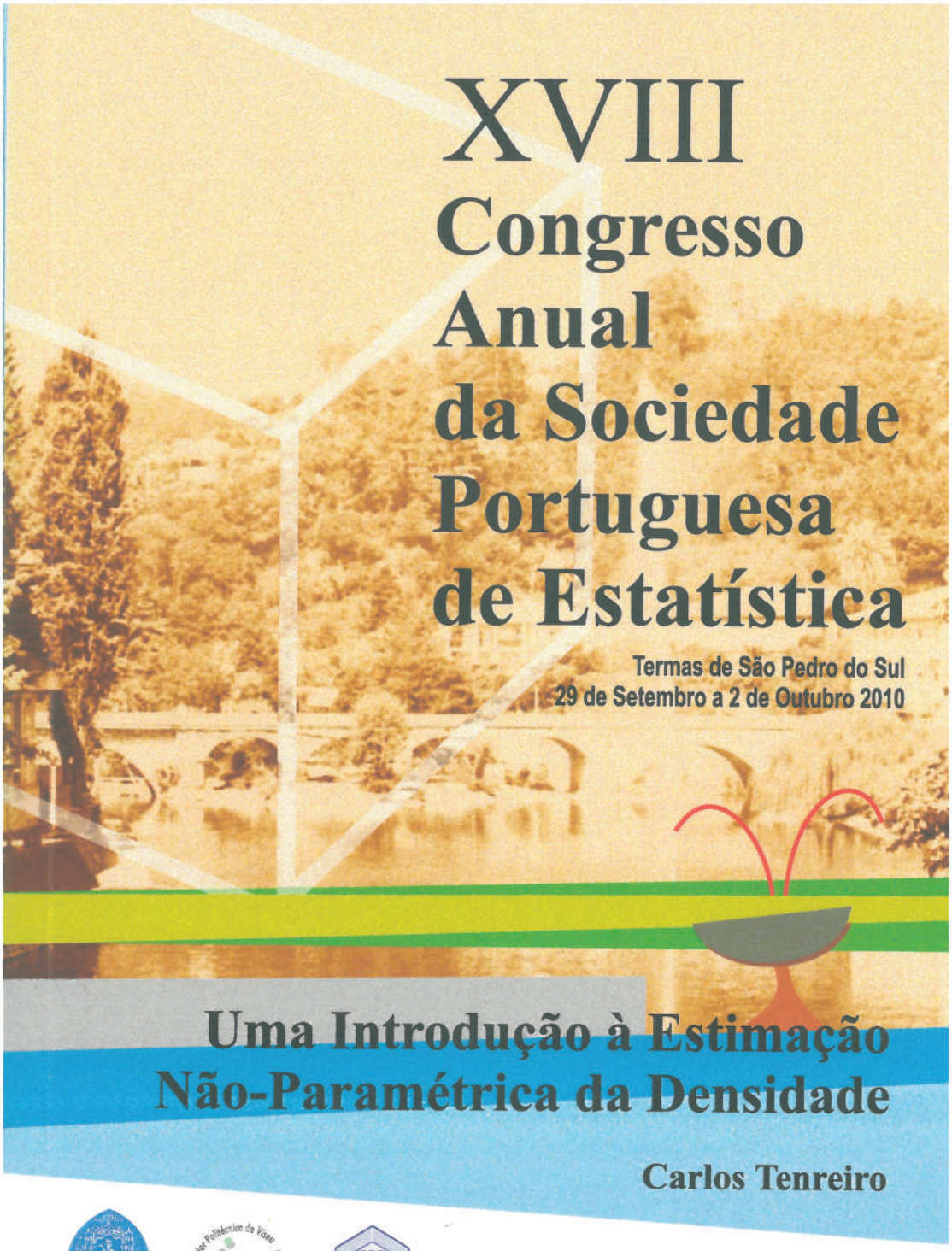

XVIII Congresso Anual da Sociedade Portuguesa de Estatística

Termas de São Pedro do Sul
29 de Setembro a 2 de Outubro 2010

Uma Introdução à Estimação Não-Paramétrica da Densidade

Carlos Tenreiro

Instituto Superior Politécnico de Viseu
SPE
SOCIEDADE PORTUGUESA DE ESTATÍSTICA

Carlos Tenreiro

Uma introdução à estimação não-paramétrica da densidade

Coimbra, 2010

Prefácio

Decorria o ano de 1956 e nos *Annals of Mathematical Statistics* surgia um artigo de Murray Rosenblatt onde era proposta uma família de estimadores da densidade de probabilidade subjacente às observações realizadas que ficariam conhecidos na literatura como estimadores do núcleo (*kernel estimates*). Apesar de outros autores, como Fix e Hodges (1951) ou Akaike (1954), terem previamente considerado estimadores do mesmo tipo no contexto da análise discriminante não-paramétrica, é sem dúvida o trabalho seminal de Rosenblatt que dá origem a toda uma vasta literatura sobre estimação não-paramétrica de funções, sejam elas densidades de probabilidade, derivadas da densidade, funções de distribuição, funções de regressão e suas derivadas, entre outras.

De entre os múltiplos estimadores não-paramétricos que os utilizadores têm presentemente à sua disposição (ver Prakasa Rao, 1983; Devroye e Györfi, 1985; Silverman, 1986; Bosq e Lecoutre, 1987; Härdle, 1990, 1991; Thompson e Tapia, 1990; Scott, 1992; Wand e Jones, 1995; Simonoff, 1996; Bowman e Azzalini, 1997; Fan e Gijbels, 1997; Härdle et al., 2004; Wasserman, 2006), o estimador do núcleo é, sem dúvida, o mais popular. Algumas das razões que podem ser avançadas para justificar tal facto, são seguramente a simplicidade da sua definição, a sua versatilidade e as suas boas propriedades teóricas e práticas. Se centrarmos a nossa atenção no caso da estimação não-paramétrica da densidade de probabilidade, tópico que estudamos neste mini-curso, a popularidade do estimador do núcleo só é

superada pela do estimador clássico do histograma que usamos em cursos introdutórios de Estatística e que é em muitos *softwares* estatísticos, o único estimador da densidade disponibilizado ao utilizador.

Por estes motivos, mas também por razões didácticas que se tornarão claras no decorrer deste mini-curso, decidimos organizar o presente texto em torno destes dois estimadores da densidade. Depois de no Capítulo 1 nos dedicarmos a questões de índole geral sobre a estimação não-paramétrica da densidade de probabilidade, nos Capítulos 2 e 3 estudamos os estimadores do histograma e do núcleo, respectivamente. Apesar de centrarmos a nossa atenção no estudo do caso unidimensional, não deixaremos de abordar a generalização dos dois estimadores ao contexto multivariado.

Tomando como ponto de partida os assuntos que expomos neste texto, teremos ainda oportunidade de abordar durante o curso, mesmo que de forma breve, outros tópicos aqui não incluídos como são os casos da estimação da densidade sob condições de dependência, dos testes de ajustamento baseados no estimador do núcleo da densidade ou da estimação pelo método do núcleo de outros parâmetros funcionais de interesse.

À Comissão Organizadora do XVIII Congresso Anual da Sociedade Portuguesa de Estatística agradeço o convite que simpaticamente me formulou para leccionar este mini-curso. Devo um agradecimento especial ao meu colega Paulo Eduardo Oliveira pelo apoio e estímulo sempre demonstrados.

Carlos Tenreiro

Coimbra, Julho de 2010

Índice

1 Estimação não-paramétrica da densidade **1**

1.1 Modelos paramétricos e não-paramétricos 1

1.2 Medidas da qualidade dum estimador 2

1.3 O estimador da janela móvel 4

1.4 Não existência de estimadores cêntricos 5

1.5 O papel do parâmetro de suavização 6

2 O estimador do histograma **9**

2.1 Definição do estimador 9

2.2 Propriedades locais de convergência 10

2.2.1 Viés 10

2.2.2 Variância 13

2.2.3 Erro quadrático médio 14

2.2.4 Convergência quase certa 16

2.3 Convergência L_∞ 17

2.4 Convergência L_1 20

2.5 Convergência em média quadrática integrada 21

2.6 Escolha assintoticamente óptima de h_n 24

2.7 Influência da origem da partição 29

2.8 Escolha prática de h_n 30

2.8.1 Métodos de utilização simples 31

2.8.2 O método de validação cruzada 33
2.8.3 Aplicação a um conjunto de dados reais 34
2.9 O polígono de frequências 36
2.10 Histogramas multivariados 39

3 O estimador do núcleo 43
3.1 Definição do estimador 43
3.2 Propriedades locais de convergência 49
3.2.1 Viés 50
3.2.2 Variância 52
3.2.3 Erro quadrático médio 54
3.2.4 Convergência quase certa 55
3.3 Convergência L_∞ 56
3.4 Convergência L_1 59
3.5 Convergência em média quadrática integrada 60
3.6 Escolha assintoticamente óptima de h_n 61
3.7 A escolha do núcleo 67
3.8 Núcleos de ordem superior. Redução de viés 73
3.9 Escolha prática de h_n 79
3.9.1 Métodos de utilização simples 79
3.9.2 Método de validação cruzada baseado no EQI 81
3.9.3 Estimação de funcionais da densidade 83
3.9.4 Outros métodos de validação cruzada 86
3.9.5 Métodos *plug-in* 89
3.9.6 Aplicação a um conjunto de dados reais 94
3.10 O estimador automático do núcleo 94
3.11 Estimação em pontos fronteiros 96
3.12 Estimador multivariado do núcleo 101

Bibliografia 105

Índice Remissivo 121

1

Estimação não-paramétrica da densidade

Modelos paramétricos e não-paramétricos. Medidas da qualidade dum estimador. O estimador da janela móvel. Não existência de estimadores cêntricos da densidade. O papel do parâmetro de suavização.

1.1 Modelos paramétricos e não-paramétricos

A partir de um conjunto de observações que interpretamos como realizações independentes de uma variável real X com distribuição absolutamente contínua com densidade de probabilidade f, desconhecida, isto é,

$$\mathrm{P}(a < X < b) = \int_a^b f(x)\,dx$$

para todo $-\infty < a < b < +\infty$, interessamo-nos neste curso pela estimação não-paramétrica da densidade f.

Contrariamente a uma abordagem paramétrica em que assumimos que f pertence a uma família paramétrica de distribuições, como a normal, a lognormal ou a gama, e em que estimamos os parâmetros desconhecidos usando, por exemplo, o estimador da máxima verosimilhança, numa abordagem não-paramétrica não é assumida qualquer forma funcional para f. A ideia de base da inferência não-paramétrica é a de usar as observações realizadas para inferir sobre f impondo condições o menos restritivas possível

sobre a distribuição de probabilidade subjacentes às mesmas.

Formalizando um pouco mais, podemos dizer que um modelo para a estimação de f consiste na introdução de uma restrição da forma

$$f \in \mathscr{D}, \tag{1.1.1}$$

onde $\mathscr{D}$ é uma família de densidades de probabilidade. Quando $\mathscr{D}$ pode ser indexada por um número finito de números reais, dizemos que o modelo é paramétrico. Caso contrário, quando a família $\mathscr{D}$ é demasiado vasta para ser indexada por um número finito de números reais, o modelo diz-se não-paramétrico (ver Bosq e Lecoutre, 1987; Ferraty e Vieu, 2006).

1.2 Medidas da qualidade dum estimador

Um estimador f_n da densidade de probabilidade desconhecida f, depende das observações realizadas e, para cada $x \in \mathbb{R}$, $f_n(x)$ é uma variável aleatória real fazendo sentido calcular a sua média e a sua variância. Várias medidas podem ser usadas para avaliar a discrepância entre f_n e a verdadeira densidade f. Quando consideramos a estimação num ponto x, uma medida natural da qualidade de $f_n(x)$ é o erro quadrático médio definido por

$$\mathrm{EQM}(f_n(x)) = \mathrm{E}\big\{f_n(x) - f(x)\big\}^2,$$

que admite a decomposição

$$\begin{aligned}\mathrm{EQM}(f_n(x)) &= \mathrm{E}\big\{f_n(x) - \mathrm{E}f_n(x)\big\}^2 + \big\{\mathrm{E}f_n(x) - f(x)\big\}^2 \\ &= \mathrm{Var}f_n(x) + \mathrm{Vi\acute{e}s}f_n(x)^2. \end{aligned} \tag{1.2.1}$$

A primeira e mais usada forma de medir a qualidade global de f_n como estimador de f é o erro quadrático médio integrado definido por

$$\mathrm{EQMI}(f_n) = \mathrm{E}\int \{f_n(x) - f(x)\}^2 dx \tag{1.2.2}$$

(cf. Rosenblatt, 1956; Parzen, 1962; Watson e Leadbetter, 1963). Atendendo à não negatividade da função integranda, o erro quadrático médio

integrado não é mais do que o integral, extendido a $\mathbb{R}$, do erro quadrático médio $\mathrm{EQM}(f_n(x))$. Tal como acontecia para este, o erro quadrático médio integrado é a soma de duas parcelas que resultam do variância e do viés do estimador:

$$\mathrm{EQMI}(f_n) = \mathrm{IVAR}(f_n) + \mathrm{IVIES}(f_n), \tag{1.2.3}$$

onde

$$\mathrm{IVAR}(f_n) = \int \mathrm{Var} f_n(x)\, dx \tag{1.2.4}$$

e

$$\mathrm{IVIES}(f_n) = \int \mathrm{Vi\acute{e}s} f_n(x)^2 dx. \tag{1.2.5}$$

Sempre que $\mathrm{EQMI}(f_n) \to 0$, quando $n \to \infty$, diremos que f_n converge para f em média quadrática integrada. Atendendo à desigualdade de Markov, tal implica que

$$||f_n - f||_2 \xrightarrow{p} 0,$$

onde

$$||f_n - f||_2 = \left(\int \{f_n(x) - f(x)\}^2 dx \right)^{1/2}$$

é a distância L_2 entre f_n e f, e $\xrightarrow{p}$ denota a convergência em probabilidade. Outras medidas que usaremos como medidas da discrepância entre f_n e f são, por exemplo, a distância L_1 e a distância uniforme ou L_∞ definidas, respectivamente, por

$$||f_n - f||_1 = \int |f_n(x) - f(x)| dx$$

e

$$||f_n - f||_\infty = \sup_{-\infty < x < +\infty} |f_n(x) - f(x)|.$$

Sobre estas e outras medidas de discrepância entre f_n e f ver Devroye (1987). Os modos de convergência em probabilidade ($\xrightarrow{p}$) e quase certa ($\xrightarrow{qc}$) serão por nós usados no estudo das medidas de discrepância anteriores. Como o leitor poderá verificar a partir das referências bibliográficas que vamos indicando, os resultados que apresentamos relativos à convergência quase certa são, na sua quase totalidade, também válidos para a convergência quase completa.

1.3 O estimador da janela móvel

Neste mini-curso vamos dar particular relevo a dois estimadores da densidade que definiremos nos parágrafos §2.1 e §3.1. Um deles é o estimador clássico do histograma cujas origens são incertas mas cuja designação é atribuída a Karl Pearson. O outro é o estimador do núcleo da densidade introduzido por Rosenblatt (1956), sendo por ele apresentado como uma extensão de um estimador da densidade previamente considerado por Fix e Hodges (1951) e Akaike (1954) no contexto da análise discriminante não-paramétrica. Esse estimador, designado habitualmente por estimador da janela móvel, é definido, para $x \in \mathbb{R}$, pela razão incremental

$$f_n(x) = \frac{F_n(x+h_n) - F_n(x-h_n)}{2h_n}, \tag{1.3.1}$$

onde (h_n) é uma sucessão de números reais estritamente positivos destinada a convergir para zero quando o tamanho da amostra aumenta, a que chamamos habitualmente janela, e

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x), \tag{1.3.2}$$

é a função de distribuição empírica associada à amostra $X_1, \ldots, X_n$ de cópias independentes da variável X.

Vamos deter-nos por breves momentos neste estimador. Possuindo $2nh_nf_n(x)$ uma distribuição binomial de parâmetros n e $F(x+h_n)-F(x-h_n)$, onde F é a função de distribuição com densidade f,

$$F(u) = \int_{]-\infty,u]} f(x)dx,$$

podemos com facilidade concluir que o viés e a variância de $f_n(x)$ são dados por

$$\text{Viés} f_n(x) = q_n(x) - f(x)$$

e

$$\text{Var} f_n(x) = \frac{1}{2nh_n}\, q_n(x)(1 - 2h_nq_n(x)),$$

onde

$$q_n(x) = \frac{F(x+h_n) - F(x-h_n)}{2h_n}.$$

Estas expressões permitem-nos, num primeiro momento, pôr em evidência as características mencionadas em §1.1 acerca do modelo estatístico associado ao problema de estimação anterior. Sendo f contínua em x, concluímos que se forem verificadas as condições $h_n \to 0$ e $nh_n \to +\infty$, então

$$\mathrm{EQM}(f_n(x)) \to 0,$$

ou seja, $f_n(x)$ converge para $f(x)$ em média quadrática. Reparemos que neste caso a família $\mathscr{D}$ em (1.1.1) é constituída pelas densidades de probabilidade contínuas no ponto x. Se f é de quadrado integrável, é ainda possível provar que nas condições anteriores sobre (h_n) se tem

$$\mathrm{EQMI}(f_n) \to 0,$$

ou seja, f_n converge para f em média quadrática integrada. Neste caso, à família $\mathscr{D}$ pertencem as densidades de probabilidade de quadrado integrável. Num e noutro caso, a família $\mathscr{D}$ é demasiado vasta para poder ser indexada por um número finito números reais, sendo não-paramétrico o modelo estatístico definido pela condição $f \in \mathscr{D}$.

1.4 Não existência de estimadores cêntricos

A expressão que obtivemos para o viés do estimador da janela móvel põe em evidência uma outra característica que, como veremos a seguir, não é específica deste estimador. Estamos a referir-nos ao facto de $f_n(x)$ não ser um estimador cêntrico de $f(x)$. Seguindo de perto a elegante demonstração apresentada por Rosenblatt (1956), mostramos a seguir que, sob certas condições gerais, não existem estimadores cêntricos da densidade. Representaremos por $\mathscr{C}$ a família das densidades de probabilidade contínuas em $\mathbb{R}$.

Teorema 1.4.1. *Dado $n \in \mathbb{N}$, não existe um estimador não-negativo $\gamma_n(\cdot; X_1, \ldots, X_n)$ de f tal que*

$$\mathrm{E}\gamma_n(x; X_1, \ldots, X_n) = f(x), \tag{1.4.2}$$

para todo o $f \in \mathscr{C}$ e $x \in \mathbb{R}$.

Dem: Sem perda de generalidade podemos assumir que $\gamma_n(\cdot; x_1, \ldots, x_n)$ é uma função simétrica de $x_1, \ldots, x_n$ (isto é, invariante para permutações das variáveis $x_1, \ldots, x_n$), uma vez que $\gamma_n^*(x; x_1, \ldots, x_n) = \frac{1}{n!}\sum \gamma_n(x; x_{i_1}, \ldots, x_{i_n})$, onde o somatório é extendido a todas as permutações $(i_1, \ldots, i_n)$ de $\{1, \ldots, n\}$, é uma função simétrica de $x_1, \ldots, x_n$ que também satisfaz (1.4.2).

Suponhamos por absurdo que $\gamma_n(\cdot; X_1, \ldots, X_n) \geq 0$ satisfaz (1.4.2) para todo o $f \in \mathscr{C}$ e $x \in \mathbb{R}$. Usando o teorema de Fubini, para $a < b$ temos

$$\begin{aligned} \mathrm{E}\int_a^b \gamma_n(x; X_1, \ldots, X_n)dx &= \int_a^b \mathrm{E}\gamma_n(x; X_1, \ldots, X_n)dx \\ &= \int_a^b f(x)dx = F(b) - F(a), \end{aligned}$$

e portanto $\int_a^b \gamma_n(x; X_1, \ldots, X_n)dx$ é um estimador cêntrico e simétrico de $F(b) - F(a)$. Como consequência do facto do vector das estatísticas ordinais ser uma estatística exaustiva e completa para o modelo estatístico em questão (cf. Lehmann, 1959, p. 133), $F_n(b) - F_n(a)$ é o único estimador cêntrico e simétrico de $F(b) - F(a)$ o que permite concluir que

$$F_n(b) - F_n(a) = \int_a^b \gamma_n(x; X_1, \ldots, X_n)dx,$$

para todo o $a < b$ e quase todo o $X_1, \ldots, X_n$. F_n seria assim uma função absolutamente contínua, o que não é verdade.

□

1.5 O papel do parâmetro de suavização

Outra característica importante revelada pelas expressões obtidas para o viés e para a variância do estimador da janela móvel, prende-se com o papel

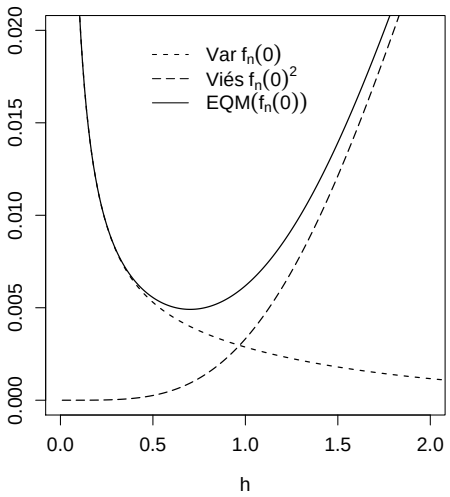

Figura 1.5.1: $\mathrm{Var} f_n(x)$, $\mathrm{Vies} f_n(x)^2$ *e* $\mathrm{EQM}(f_n(x))$ *em função de* $h_n = h$ *para a distribuição normal standard no ponto* $x = 0$ *com* $n = 100$.

desempenhado pela janela h_n, a que chamamos também parâmetro de suavização (*smoothing parameter*) e que na prática é escolhida pelo utilizador. Tomando $x = 0$, f a densidade normal standard, e usando as expressões deduzidas em §1.3, representamos na Figura 1.5.1 o comportamento da variância e do quadrado do viés de $f_n(0)$ em função da escolha da janela h_n. Verificamos que valores demasiado pequenos de h_n conduzem a um estimador com viés reduzido mas com grande variabilidade, enquanto que valores grandes de h_n reduzem a variabilidade do estimador mas aumentam o seu viés. Como se ilustra na Figura 1.5.2, no primeiro caso as estimativas produzidas apresentam grandes irregularidades fazendo surgir características artificiais de multimodalidade, enquanto que no segundo caso, tais características, a existirem, poderão não ser reveladas pelo estimador.

Uma escolha adequada da janela h_n passará necessariamente por estabelecer um compromisso entre estas duas situações. Este será um ponto fundamental do estudo que desenvolveremos nos próximos capítulos sobre dois dos mais populares estimadores da densidades existentes na literatura: o estimador do histograma e o estimador do núcleo. Salvo indicação em

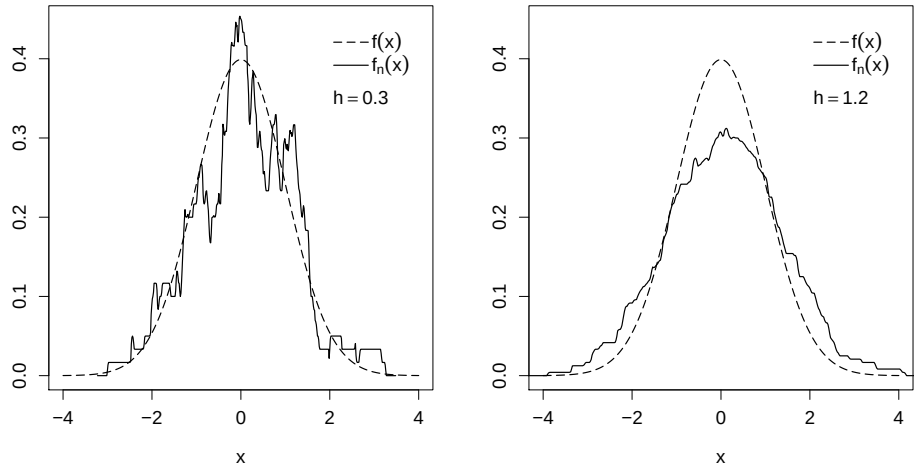

Figura 1.5.2: *Estimativas do estimador da janela móvel com* $h_n = 0.3$ *e* $h_n = 1.2$*, para uma amostra de tamanho* $n = 100$ *da distribuição normal standard.*

contrário, assumiremos que $X_1, \ldots, X_n$ são variáveis aleatórias reais independentes e identicamente distribuídas, possuindo uma distribuição absolutamente contínua com função de distribuição F e densidade de probabilidade f, desconhecidas.

Os gráficos anteriores, bem como outros gráficos que apresentamos ao longo do texto foram executados com recurso ao *software* R (R Development Core Team, 2009) e à *package* KernSmooth (S original by Matt Wand. R port by Brian Ripley, 2009).

2

O estimador do histograma

Definição do estimador. Propriedades locais de convergência. Desenvolvimentos assintóticos para o viés e variância num ponto. Propriedades globais de convergência. Desenvolvimento assintótico para o erro quadrático médio integrado. Escolha assintoticamente óptima de h_n. Influência da origem no comportamento do histograma. Escolha prática de h_n. O polígono de frequências. Histogramas multivariados.

2.1 Definição do estimador

Para $n \in \mathbb{N}$ defina-se uma partição Δ_n de $\mathbb{R}$ a partir de uma origem $a_{n,0}$ por

$$\Delta_{n,j} =]a_{n,j}, a_{n,j+1}],$$

onde

$$a_{n,j+1} = a_{n,j} + h_n,\ j \in \mathbb{Z},$$

e

$$h_n > 0.$$

Cada um dos intervalos $\Delta_{n,j}$ é habitualmente denominado classe. Ao conjunto dos pontos a_{nj}, $j \in \mathbb{Z}$, chamaremos malha da partição Δ_n.

O histograma associado à partição anterior é definido, para $x \in \Delta_{n,j}$, por

$$\hat{f}_n(x) = \frac{F_n(a_{n,j+1}) - F_n(a_{n,j})}{h_n} \tag{2.1.1}$$

$$= \frac{1}{nh_n} \sum_{i=1}^{n} I(X_i \in \Delta_{n,j}), \tag{2.1.2}$$

onde F_n é a função de distribuição empírica definida por (1.3.2).

Para todo o $n \in \mathbb{N}$, $\hat{f}_n$ é uma densidade de probabilidade. Claramente $\hat{f}_n(x) \geq 0$ para todo o $x \in \mathbb{R}$ e

$$\begin{aligned} \int \hat{f}_n(x)dx &= \sum_{j\in\mathbb{Z}} \int_{\Delta_{n,j}} \hat{f}_n(x)dx \\ &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j\in\mathbb{Z}} I(X_i \in \Delta_{n,j}) = 1. \end{aligned}$$

Independentemente da densidade f ser, ou não, contínua, o histograma é uma função escalonada, estando o conjunto dos seus pontos de descontinuidade contido na malha da partição. Apesar do histograma ser, por vezes, demasiado sensível à escolha da origem da partição, em particular quando o tamanho da amostra não é grande, a escolha do parâmetro h_n a que chamaremos janela, é de fundamental importância, dependendo dela as propriedades, quer assintóticas quer a distância finita, do histograma como estimador da densidade de probabilidade.

2.2 Propriedades locais de convergência

Os primeiros estudos sobre a qualidade do histograma como estimador da densidade de probabilidade são, de acordo com Cencov (1962) e Nadaraya (1989), efectuados por Glivenko (1934) e Smirnov (1950, 1951). Os resultados que apresentamos neste parágrafo sobre comportamento assintótico do viés, variância e erro quadrático médio do histograma, surgem nos trabalhos de Révész (1972), Scott (1979) e Freedman e Diaconis (1981).

2.2.1 Viés

Sabemos já que não existe um estimador não negativo da densidade de probabilidade que seja cêntrico para toda a densidade f pertencente a um

conjunto que inclua as densidade contínuas. No entanto, se $h_n \to 0$, quando $n \to +\infty$, mostramos a seguir que o histograma é assintoticamente cêntrico (ou assintoticamente sem viés) no conjunto $\mathscr{C}$ das densidades contínuas em $\mathbb{R}$.

Proposição 2.2.1. *Seja f contínua em $x \in \mathbb{R}$. Se $h_n \to 0$ então*

$$\mathrm{E}\hat{f}_n(x) \to f(x).$$

Dem: Para $n \in \mathbb{N}$, representemos por $\Delta_{n,j_x} =]a_{n,j_x}, a_{n,j_x} + h_n]$ o elemento da partição Δ_n que contém x. De (2.1.2) temos

$$\mathrm{E}\hat{f}_n(x) = \frac{1}{h_n} \int_{\Delta_{n,jx}} f(y)dy, \tag{2.2.2}$$

e também

$$|\mathrm{E}\hat{f}_n(x) - f(x)| \leq \frac{1}{h_n} \int_{\Delta_{n,jx}} |f(y) - f(x)|dy.$$

Pela continuidade de f em x, e atendendo a que $h_n \to 0$, basta agora notar que para $\epsilon > 0$, qualquer, e para n suficientemente grande, vale a desigualdade $|f(y) - f(x)| < \epsilon$, sempre que $y \in \Delta_{n,j_x}$.

□

A ordem de convergência para zero do viés depende de propriedades locais da densidade desconhecida f. Se além da continuidade no ponto x, f verificar uma condição de Lipschitz de ordem α, para $\alpha \in]0, 1]$, numa vizinhança de $x \in \mathbb{R}$, isto é, se existirem $\delta > 0$ e $M > 0$ (possivelmente dependentes de x) tais que

$$|f(y) - f(z)| \leq M|y - z|^\alpha,$$

para todo o $y, z \in]x - \delta, x + \delta[$, escrevemos neste caso $f \in \mathscr{L}_x(\alpha)$, da demonstração anterior concluímos que para n suficientemente grande

$$|\mathrm{Viés}\hat{f}_n(x)| \leq Mh_n^\alpha,$$

e portanto

$$\mathrm{Viés}\hat{f}_n(x) = O\left(h_n^\alpha\right). \tag{2.2.3}$$

Em particular, sendo f' limitada numa vizinhança de x temos

$$\text{Viés}\hat{f}_n(x) = O\left(h_n\right).$$

Como veremos a seguir, mesmo para densidades mais regulares a ordem de convergência anterior não pode ser melhorada. Para demonstrarmos este facto, vamos lançar mão da fórmula de Taylor que recordamos a seguir (cf. Lima, 1995, pp. 222–5, 262–3). Este é sem dúvida um instrumento fundamental para obter desenvolvimentos assintóticos e será por nós usado com frequência.

Teorema 2.2.4 (Fórmula de Taylor)**.** *Seja g uma função real de variável real admitindo derivada até à ordem $p-1$ numa vizinhança do ponto $x \in \mathbb{R}$. Então, sendo u_n uma qualquer sucessão de números reais convergente para zero temos*

$$g(x+u_n) = \sum_{\ell=0}^{p-1} \frac{u_n^\ell}{\ell!}\, g^{(\ell)}(x) + R_n(x),$$

onde

$$R_n(x) = o(u_n^{p-1}). \qquad \textit{[resto infinitesimal]}$$

Além disso:

a) Se g admite derivada de ordem p numa vizinhança do ponto x, então para $c_n(x) \in]x, x+u_n[$ temos

$$R_n(x) = \frac{u_n^p}{p!}\, g^{(p)}(c_n(x)), \qquad \textit{[resto de Lagrange]}$$

b) Se g admite derivada de ordem p contínua numa vizinhança do ponto x temos

$$R_n(x) = u_n^p \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!}\, g^{(p)}(x+tu_n)dt. \qquad \textit{[resto integral]}$$

No resultado seguinte assumimos que f' é de Lipschitz de ordem α no ponto x. Esta condição é em particular satisfeita com $\alpha = 1$ se f possuir derivada de segunda ordem numa vizinhança de x.

Teorema 2.2.5. *Para $x \in \mathbb{R}$, se $f' \in \mathscr{L}_x(\alpha)$ e $h_n \to 0$ então*

$$\mathrm{E}\hat{f}_n(x) - f(x) = \frac{h_n}{2} f'(x) \left(1 - 2(x - a_{n,j_x})/h_n\right) + O\left(h_n^{\alpha+1}\right),$$

onde $]a_{n,j_x}, a_{n,j_x} + h_n]$ é o elemento da partição Δ_n que contém x.

Dem: Atendendo a que da fórmula de Taylor com resto integral se tem

$$f(y) = f(x) + (y - x) \int_0^1 f'(x + t(y - x))dt,$$

de (2.2.2) concluímos que

$$\mathrm{E}\hat{f}_n(x) - f(x) = \frac{f'(x)}{h_n} \int_{a_{n,j_x}}^{a_{n,j_x}+h_n} (y - x)dy$$
$$+ \frac{1}{h_n} \int_{a_{n,j_x}}^{a_{n,j_x}+h_n} (y - x) \int_0^1 \left(f'(x + t(y - x)) - f'(x)\right) dtdy,$$

onde

$$\int_{a_{n,j_x}}^{a_{n,j_x}+h_n} (y - x)dy = \frac{h_n^2}{2} \left(1 - 2(x - a_{n,j_x})/h_n\right)$$

e

$$\left| \int_{a_{n,j_x}}^{a_{n,j_x}+h_n} (y - x) \int_0^1 \left(f'(x + t(y - x)) - f'(x)\right) dtdy \right| \leq M\, h_n^{\alpha+2},$$

para n suficientemente grande.

□

Reparemos que nas condições anteriores o viés é proporcional a h_n uma vez que $|1 - 2(x - a_{n,j_x})/h_n| \leq 1$. No entanto, se o ponto x é ponto médio de uma das classes das sucessivas partições, isto é, se $x = a_{n,j_x} + h_n/2$, o viés é de ordem $h_n^{\alpha+1}$ e não apenas de ordem h_n.

2.2.2 Variância

Tendo em conta (2.1.2) e (2.2.2), a variância de $\hat{f}_n(x)$ é dada por

$$\begin{aligned} \mathrm{Var}\hat{f}_n(x) &= \frac{1}{nh_n^2} \int_{\Delta_{n,j_x}} f(y)dy \left(1 - \int_{\Delta_{n,j_x}} f(y)dy\right) \\ &= \frac{1}{nh_n} \mathrm{E}\hat{f}_n(x)(1 - h_n \mathrm{E}\hat{f}_n(x)), \end{aligned} \tag{2.2.6}$$

onde Δ_{n,j_x} é o elemento da partição Δ_n que contém x. O resultado seguinte é consequência da Proposição 2.2.1.

Proposição 2.2.7. *Seja f contínua em $x \in \mathbb{R}$. Se $h_n \to 0$ então*

$$nh_n \operatorname{Var} \hat{f}_n(x) \to f(x).$$

Em particular, se $nh_n \to +\infty$ concluímos que

$$\operatorname{Var} \hat{f}_n(x) \to 0,$$

sendo a ordem de convergência para zero da variância do histograma dada por

$$\operatorname{Var} \hat{f}_n(x) = O\left((nh_n)^{-1}\right).$$

Com condições adicionais de regularidade sobre f podemos ainda obter o desenvolvimento seguinte. Reparemos, no entanto, que tais condições são mais fracas que as utilizadas no desenvolvimento assintótico do viés.

Teorema 2.2.8. *Seja f com derivada f' limitada numa vizinhança de $x \in \mathbb{R}$. Se $h_n \to 0$ então*

$$\operatorname{Var} \hat{f}_n(x) = \frac{1}{nh_n} f(x) + O(n^{-1}).$$

Dem: Consequência das igualdades (2.2.6) e $\mathrm{E}\hat{f}_n(x) = f(x) + O(h_n)$. □

2.2.3 Erro quadrático médio

O resultado seguinte é agora uma consequência da decomposição (1.2.1) do erro quadrático médio no ponto x como soma de dois termos provenientes da variância e do viés do estimador.

Teorema 2.2.9. *Seja f contínua em $x \in \mathbb{R}$. Se $h_n \to 0$ e $nh_n \to +\infty$ então*

$$\mathrm{EQM}(\hat{f}_n(x)) \to 0.$$

Sob condições adicionais de regularidade sobre f podemos ainda obter, como consequência imediata dos Teoremas 2.2.5 e 2.2.8, o desenvolvimento seguinte.

Teorema 2.2.10. *Para $x \in \mathbb{R}$, se f é derivável com $f' \in \mathscr{L}_x(\alpha)$ então,*

$$\begin{aligned}\mathrm{EQM}(\hat{f}_n(x)) = \frac{1}{nh_n} f(x) + \frac{h_n^2}{4} f'(x)^2 \left(1 - 2(x - a_{n,j_x})/h_n\right)^2 \\ + O(n^{-1}) + O\left(h_n^{2+\alpha}\right).\end{aligned}$$

Relativamente à ordem de convergência para zero do erro quadrático médio no ponto x, podemos concluir da discussão tida nos dois parágrafos anteriores que esta depende das condições de regularidade assumidas sobre f. Em particular, se $f \in \mathscr{L}_x(\alpha)$ com $\alpha \in]0, 1]$, então

$$\mathrm{EQM}(\hat{f}_n(x)) = O\left((nh_n)^{-1} + h_n^{2\alpha}\right).$$

Assim, a maior ordem de convergência é atingida tomando $h_n = c\, n^{-1/(1+2\alpha)}$, com $c > 0$, obtendo-se neste caso

$$\mathrm{EQM}(\hat{f}_n(x)) = O\big(n^{-2\alpha/(1+2\alpha)}\big).$$

No caso limite $\alpha = 1$, e tomando

$$h_n = c\, n^{-1/3}, \text{ com } c > 0,$$

a ordem de convergência em probabilidade para o estimador do histograma é dada por

$$\hat{f}_n(x) - f(x) = O_p\big(n^{-1/3}\big).$$

As decomposições anteriores põem em evidência o papel determinante que h_n tem no comportamento assintótico do estimador. A janela deverá convergir para zero para que o viés do estimador seja pequeno mas não poderá convergir muito rapidamente sob pena de obtermos um estimador com grande variabilidade. Dito de outra maneira, se a janela h_n é demasiado pequena o histograma será muito irregular, enquanto que se h_n é demasiado grande o histograma será demasiado suave. Uma escolha adequada de h_n passará necessariamente por estabelecer um compromisso entre estas duas situações. Este assunto será estudado mais à frente.

2.2.4 Convergência quase certa

A derivação de resultados sobre a convergência pontual quase certa do histograma, é conseguida a partir das chamadas desigualdades exponenciais, também conhecidas por desigualdades de tipo Bernstein. O resultado seguinte, consequência do Teorema 3 de Hoeffding (1963), dá-nos um exemplo de uma tal desigualdade.

Lema 2.2.11. *Sejam $Z_1, \ldots, Z_n$ variáveis aleatórias reais independentes e identicamente distribuídas com $\mathrm{E}(Z_i) = 0$, $|Z_i| \leq \alpha$ e $\mathrm{E}(Z_i)^2 \leq \beta^2$. Então, para $0 < \epsilon < \beta^2/\alpha$, vale a desigualdade*

$$\mathrm{P}\left(\Big| \sum_{i=1}^{n} Z_i \Big| \geq n\epsilon \right) \leq 2 \exp\left(-\frac{3n\epsilon^2}{8\beta^2} \right).$$

No resultado seguinte estabelecemos condições suficientes sobre a janela para a convergência pontual quase certa do histograma. Tais condições são em particular satisfeitas quando $h_n \to 0$ e $nh_n/\log n \to +\infty$.

Teorema 2.2.12. *Seja f contínua em $x \in \mathbb{R}$. Se $h_n \to 0$ e se*

$$\sum_{n=1}^{\infty} \exp(-\gamma n h_n) < \infty, \text{ para todo o } \gamma > 0,$$

então

$$\hat{f}_n(x) \xrightarrow{qc} f(x).$$

Dem: Atendendo à Proposição 2.2.1 basta mostrar que $\hat{f}_n(x) - \mathrm{E}\hat{f}_n(x) \xrightarrow{qc} 0$. Ora,

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

com

$$Z_i = \frac{1}{h_n} \big(I(X_i \in \Delta_{n,j_x}) - \mathrm{P}(X_i \in \Delta_{n,j_x}) \big),$$

onde $|Z_i| \leq 1/h_n$ e para n suficientemente grande, $\mathrm{E}(Z_i)^2 \leq C/h_n$, com $C > 0$. Pelo Lema 2.2.11 concluímos que para todo o $0 < \epsilon < C$, se tem

$$\mathrm{P}\big(|\hat{f}_n(x) - \mathrm{E}\hat{f}_n(x)| > \epsilon \big) \leq 2 \exp\big(-3nh_n\epsilon^2/(8C) \big). \tag{2.2.13}$$

Para concluir basta agora notar que por hipótese o segundo membro da desigualdade anterior é o termo geral de uma série convergente.

□

Ordem de convergência

A ordem de convergência quase certa que a seguir apresentamos é consequência imediata de (2.2.3) e da desigualdade (2.2.13).

Teorema 2.2.14. *Para $x \in \mathbb{R}$, seja $f \in \mathscr{L}_x(\alpha)$. Se $h_n \to 0$ e $nh_n/\log n \to +\infty$ então*

$$\hat{f}_n(x) - f(x) = O\bigg(\Big(\frac{\log n}{nh_n}\Big)^{1/2} + h_n^\alpha \bigg).$$

A maior ordem de convergência quase certa que podemos obter a partir do resultado anterior passa por tomar $h_n = c\,(\log n/n)^{1/(1+2\alpha)}$, com $c > 0$, obtendo-se neste caso

$$\hat{f}_n(x) - f(x) = O\bigg(\Big(\frac{\log n}{n}\Big)^{\alpha/(1+2\alpha)} \bigg).$$

No caso em que $\alpha = 1$ obtemos a ordem de convergência

$$\hat{f}_n(x) - f(x) = O\bigg(\Big(\frac{\log n}{n}\Big)^{1/3} \bigg),$$

que é ligeiramente inferior à ordem de convergência em probabilidade que obtivemos em §2.2.3 sob condições análogas sobre f.

2.3 Convergência L_∞

Estudamos neste parágrafo a convergência uniforme forte do estimador do histograma. Pretendemos determinar condições em h_n e f que assegurem a convergência $||\hat{f}_n - f||_\infty \xrightarrow{qc} 0$, onde $||\hat{f}_n - f||_\infty$ é a distância uniforme entre $\hat{f}_n$ e f. Atendendo à desigualdade

$$||\hat{f}_n - f||_\infty \leq ||\hat{f}_n - \mathrm{E}\hat{f}_n||_\infty + ||\mathrm{E}\hat{f}_n - f||_\infty, \tag{2.3.1}$$

um tal estudo pode ser decomposto na análise dos dois termos anteriores. Um termo aleatório $||\hat{f}_n - \mathrm{E}\hat{f}_n||_\infty$ e um termo determinista, $||\mathrm{E}\hat{f}_n - f||_\infty$, relativo ao viés do estimador.

Comecemos por reparar que não assumindo condições globais de regularidade sobre f não é possível garantir a convergência uniforme de $\hat{f}_n$ para f. Com efeito, se f for tal que $f(x_k) = 1$, para alguma sucessão (x_k) com $x_k \to +\infty$, temos $||\hat{f}_n - f||_\infty = 1$, uma vez que $\hat{f}_n$ é nulo fora de um intervalo limitado. Assumiremos no que se segue que f é uniformemente contínua em $\mathbb{R}$, isto é, para todo o $\epsilon > 0$, existe $\delta > 0$ tal que

$$|f(x) - f(y)| < \epsilon,$$

para todos os números reais x, y que satisfazem

$$|x - y| < \delta.$$

Se, por exemplo, f satisfaz a condição (global) de Lipschitz de ordem α,

$$|f(x) - f(y)| \leq M|x - y|^\alpha,$$

para todo o $x, y \in \mathbb{R}$, onde $M > 0$ e $\alpha \in]0, 1]$, então f é uniformemente contínua. Por maioria de razão é também uniformemente contínua uma função diferenciável em $\mathbb{R}$ com derivada limitada. Representaremos por $\mathscr{U}$ o conjunto das densidades de probabilidade uniformemente contínuas em $\mathbb{R}$.

Voltando à desigualdade (2.3.1), reparemos que da demonstração da Proposição 2.2.1 se deduz a convergência para zero do termo de viés, para todo o $f \in \mathscr{U}$, sempre que $h_n \to 0$. Basta assim estudar a convergência do termo aleatório que faremos à custa da desigualdade seguinte que decorre de (2.1.1):

$$||\hat{f}_n - \mathrm{E}\hat{f}_n||_\infty \leq \frac{2}{h_n} ||F_n - F||_\infty. \tag{2.3.2}$$

Podemos assim apresentar uma primeira condição suficiente para a convergência uniforme quase certa do histograma.

Teorema 2.3.3. *Se* $f \in \mathscr{U}$, $h_n \to 0$ *e* $nh_n^2/\log\log n \to +\infty$ *então*

$$||\hat{f}_n - f||_\infty \xrightarrow{qc} 0.$$

Dem: Consequência das desigualdades (2.3.1) e (2.3.2) e da lei do logaritmo iterado para a função de distribuição empírica (ver van der Vaart, 1998, p. 268),

$$\limsup \sqrt{\frac{2n}{\log\log n}}\, ||F_n - F||_\infty \leq 1 \, q.c. \tag{2.3.4}$$

□

Observação 2.3.5. Um conjunto alternativo de condições suficientes sobre h_n para a convergência uniforme forte do histograma pode ser obtido a partir da desigualdade exponencial seguinte válida para todo o $t > 0$ e $n \in \mathbb{N}$,

$$\mathrm{P}\big(||F_n - F||_\infty \geq t\big) \leq 2\exp(-2t^2 n)$$

(cf. Dvoretzky et al., 1956; Massart, 1990).

As condições anteriores sobre h_n podem ser ainda enfraquecidas. Uma análise mais fina do termo do termo aleatório $||\hat{f}_n - \mathrm{E}\hat{f}_n||_\infty$ permite a Geffroy (1974) estabelecer que, na presença na condição $h_n \to 0$, a condição $nh_n/\log n \to +\infty$ é necessária e suficiente para a convergência uniforme quase certa do estimador do histograma para toda a densidade de probabilidade uniformemente contínua. Além disso, os modos de convergência quase certa e em probabilidade são equivalentes. Outros resultados sobre a convergência uniforme forte foram obtidos por Révész (1972) (que também obtém ordens de convergência) e Bertrand-Retali (1974). É o resultado de Geffroy que enunciamos a seguir. A sua demonstração pode ser encontrada em Bosq e Lecoutre (1987, pp. 140–144).

Teorema 2.3.6. *Se* $h_n \to 0$ *então as seguintes proposições são equivalentes:*

i) $nh_n/\log n \to +\infty$*;*
ii) $\forall\, f \in \mathscr{U} \quad ||\hat{f}_n - f||_\infty \xrightarrow{p} 0$*;*
iii) $\forall\, f \in \mathscr{U} \quad ||\hat{f}_n - f||_\infty \xrightarrow{qc} 0.$

2.4 Convergência L_1

O Teorema de Glick (1974), que enunciamos a seguir, estabelece uma ligação interessante entre a convergência pontual e a convergência L_1 de um estimador próprio da densidade. Ver Devroye e Györfi (1985, pp. 10–11), para uma demonstração deste resultado.

Teorema 2.4.1 (de Glick)**.** *Seja $\gamma_n(\cdot) = \gamma_n(\cdot; X_1, \ldots, X_n)$ uma sucessão de estimadores da densidade com $\gamma_n \geq 0$ e $\int \gamma_n(x)\,dx = 1$. Se $\gamma_n(x) \xrightarrow{p} f(x)$ (resp. $\xrightarrow{qc}$) para quase todo o ponto $x \in \mathbb{R}$ (isto é, a menos de um conjunto de medida de Lebesgue nula), então $||\gamma_n - f||_1 \xrightarrow{p} 0$ (resp. $\xrightarrow{qc}$).*

Sendo f contínua em quase todo o ponto $x \in \mathbb{R}$ e tendo em conta os resultados estabelecidos nos parágrafos anteriores, podemos obter, via Teorema de Glick, condições suficientes para a convergência L_1, em probabilidade e quase certamente, do estimador histograma. Com efeito, se a janela h_n satisfaz as condições do Teorema 2.2.9 temos

$$||f_n - f||_1 \xrightarrow{p} 0,$$

e se h_n satisfaz as condições do Teorema 2.2.12, concluímos que

$$||f_n - f||_1 \xrightarrow{qc} 0.$$

Abou-Jaoudé (1976a,b) mostra que é possível fazer melhor estabelecendo que as condições $h_n \to 0$ e $nh_n \to +\infty$ são necessárias e suficientes para a convergência L_1 do estimador do histograma. Além disso, os modos de convergência quase certa e em probabilidade são equivalentes. Contrariamente ao resultado correspondente para a convergência L_∞, reparemos que não é feita sobre f qualquer hipótese de regularidade. Denotaremos por $\mathscr{F}$ a família das densidades de probabilidade sobre $\mathbb{R}$.

Teorema 2.4.2. *As proposições seguintes são equivalentes:*

i) $h_n \to 0$, $nh_n \to +\infty$;

ii) $\forall\, f \in \mathscr{F} \quad ||\hat{f}_n - f||_1 \xrightarrow{p} 0$;

iii) $\forall\, f \in \mathscr{F} \quad ||\hat{f}_n - f||_1 \xrightarrow{qc} 0$.

Ver Bosq e Lecoutre (1987, pp. 145–150) ou Devroye e Györfi (1985, pp. 19–23), para a demonstração deste resultado.

2.5 Convergência em média quadrática integrada

Estabelecemos nesta secção condições suficientes para a convergência em média quadrática integrada do estimador do histograma Começamos por obter um desenvolvimento exacto para o erro quadrático médio integrado de $\hat{f}_n$ a partir de desenvolvimentos exactos de IVAR($\hat{f}_n$) e IVIES($\hat{f}_n$) definidos por (1.2.4) e (1.2.5), respectivamente. Como veremos, cada uma destas quantidades pode ser expressa de forma simples em termos das probabilidade de cada um dos intervalos da partição que denotaremos por $p_{n,j}$, para $n \in \mathbb{N}$ e $j \in \mathbb{Z}$:

$$p_{n,j} = \int_{\Delta_{n,j}} f(y)\, dy.$$

Sendo g uma função de quadrado integrável, usaremos frequentemente a notação

$$R(g) = \int g(x)^2\, dx.$$

Teorema 2.5.1. *Se f é de quadrado integrável então*

$$\mathrm{IVAR}(\hat{f}_n) = \frac{1}{nh_n}\left(1 - \sum_{j\in\mathbb{Z}} p_{n,j}^2\right),$$

$$\mathrm{IVIES}(\hat{f}_n) = R(f) - \frac{1}{h_n}\sum_{j\in\mathbb{Z}} p_{n,j}^2$$

e

$$\mathrm{EQMI}(\hat{f}_n) = \frac{1}{nh_n} - \frac{n+1}{nh_n}\sum_{j\in\mathbb{Z}} p_{n,j}^2 + R(f).$$

Dem: Comecemos por notar que $\mathrm{E}\hat{f}_n(\cdot)$ é de quadrado integrável uma vez que para $x \in \Delta_{n,j}$,

$$(\mathrm{E}\hat{f}_n(x))^2 = \left(\frac{1}{h_n}\int_{\Delta_{n,j}} f(y)\, dy\right)^2 \leq \frac{1}{h_n^2}\int_{\Delta_{n,j}} f(y)^2\, dy,$$

e

$$\int_{\Delta n,j} (\mathrm{E}\hat{f}_n(x))^2 dx \leq \frac{1}{h_n} \int_{\Delta_{n,j}} f(y)^2 \, dy$$

onde

$$\sum_{j\in\mathbb{Z}} \int_{\Delta_{n,j}} f(y)^2 \, dy = \int f(y)^2 dy < \infty.$$

Além disso,

$$\int (\mathrm{E}\hat{f}_n(x))^2 dx = \sum_{j\in\mathbb{Z}} \int_{\Delta_{n,j}} \left(\frac{1}{h_n} \int_{\Delta_{n,j}} f(y) \, dy \right)^2 dx = \frac{1}{h_n} \sum_{j\in\mathbb{Z}} p_{n,j}^2. \quad (2.5.2)$$

A expressão apresentada para $\mathrm{IVAR}(\hat{f}_n)$ é agora consequência imediata da igualdade (2.2.6) uma vez que $\int \mathrm{E}\hat{f}_n(x)dx = 1$. Para provar a segunda igualdade basta ter em conta que $\int f(x)\mathrm{E}\hat{f}_n(x)dx = \int (\mathrm{E}\hat{f}_n(x))^2 dx$ e usar a igualdade

$$\mathrm{IVIES}(\hat{f}_n) = R(f) - \int (\mathrm{E}\hat{f}_n(x))^2 dx. \quad (2.5.3)$$

□

Para estabelecer a convergência em média quadrática de $\hat{f}_n$ necessitamos de assumir algumas condições globais de regularidade sobre a densidade desconhecida f. Em particular assumiremos que f é absolutamente contínua em $\mathbb{R}$. Recordemos que uma função real g definida em $\mathbb{R}$ é absolutamente contínua em $\mathbb{R}$ sempre que para todo o $\epsilon > 0$ exista $\delta > 0$ tal que

$$\sum_{i=1}^{n} |g(b_i) - g(a_i)| < \epsilon,$$

para todo o conjunto finito de números reais com $-\infty < a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_n < b_n < \infty$ que satisfazem $\sum_{i=1}^{n}(b_i - a_i) < \delta$ (sobre funções absolutamente contínuas, ver Cohn, 1980, pp. 146–148 e Bogachev, 2007, pp. 337–340).

Uma função absolutamente contínua em $\mathbb{R}$ é claramente uniformemente contínua e uma função diferenciável com derivada limitada é necessariamente absolutamente contínua. Uma caracterização importante das funções absolutamente contínuas é o de serem integral indefinido da sua própria

derivada (que se pode provar existir em quase todo o ponto de $\mathbb{R}$ e ser integrável num intervalo finito), isto é,

$$g(x) = g(a) + \int_a^x g'(t)\,dt,$$

para algum $a \in \mathbb{R}$. Atendendo ao teorema fundamental do cálculo, a caracterização anterior permite concluir que uma função g com derivada contínua é necessariamente absolutamente contínua.

A noção de função absolutamente contínua será para nós particularmente útil por permitir obter aproximações para somas de Riemann dessa função estendidas a uma partição de $\mathbb{R}$ de acordo com o resultado seguinte devido a Freedman e Diaconis (1981):

Lema 2.5.4. *Sejam g absolutamente contínua em $\mathbb{R}$, g, g' integráveis em $\mathbb{R}$ e $]a_j, a_{j+1}]$, $j \in \mathbb{Z}$, uma partição de $\mathbb{R}$. Sejam ainda x_j escolhido em $[a_j, a_{j+1}]$, para $j \in \mathbb{Z}$, e $h = \sup_{j\in\mathbb{Z}}(a_{j+1} - a_j)$. Então*

$$\left| \sum_{j\in\mathbb{Z}} g(x_j)(a_{j+1} - a_j) - \int g(x)\,dx \right| \leq h \int |g'(x)|dx.$$

Estamos agora em condições de estabelecer a convergência em média quadrática integrada do estimador do histograma. Uma versão mais fina deste resultado é devida a Lecoutre (cf. Bosq e Lecoutre, 1987, pp. 139–140).

Teorema 2.5.5. *Sejam f absolutamente contínua e f e f' de quadrado integrável. Se $h_n \to 0$ e $nh_n \to +\infty$ então $\hat{f}_n$ converge para f em média quadrática integrada.*

Dem: Tendo em conta a igualdade (1.2.3), basta mostrar que IVAR($\hat{f}_n$) e IVIES($\hat{f}_n$) convergem para zero com $1/n$. Para estabelecer a convergência para zero de IVIES($\hat{f}_n$), começamos por usar o teorema do valor médio para integrais (cf. Lima, 1995, p. 260) para obter

$$p_{n,j} = \int_{\Delta_{n,j}} f(y)dy = f(x_{n,j})h_n,$$

para algum $x_{n,j} \in]a_{nj}, a_{n,j+1}[$. Usando agora o Lema 2.5.4 e a integrabilidade do produto ff' temos

$$\frac{1}{h_n}\sum_{j\in\mathbb{Z}} p_{n,j}^2 = \sum_{j\in\mathbb{Z}} f(x_{nj})^2 h_n \to R(f),$$

o que, pelo Teorema 2.5.1, permite concluir que $\mathrm{IVIES}(\hat{f}_n) \to 0$. A convergência para zero da parcela $\mathrm{IVAR}(\hat{f}_n)$, resulta imediatamente do Teorema 2.5.1 pois como $0 < \sum_{j\in\mathbb{Z}} p_{n,j}^2 \leq 1$, vale a desigualdade $\mathrm{IVAR}(\hat{f}_n) \leq 1/(nh_n)$.

□

2.6 Escolha assintoticamente óptima de h_n

No resultado seguinte, obtido por Freedman e Diaconis (1981) sob condições menos restritivas sobre f, apresentamos um desenvolvimento assintótico para $\mathrm{EQMI}(\hat{f}_n)$. Resultados do mesmo tipo são obtidos por Scott (1979) e Lecoutre (1985).

Teorema 2.6.1. *Suponhamos que f possui derivada de terceira ordem, com f'' absolutamente contínua e f, f', f'', f''' de quadrado integrável. Se $h_n \to 0$ então*

$$\mathrm{IVAR}(\hat{f}_n) = \frac{1}{nh_n} + O(n^{-1}),$$

$$\mathrm{IVIES}(\hat{f}_n) = \frac{h_n^2}{12} R(f') + O\left(h_n^3\right)$$

e

$$\mathrm{EQMI}(\hat{f}_n) = \frac{1}{nh_n} + \frac{h_n^2}{12} R(f') + O(n^{-1}) + O\left(h_n^3\right).$$

Dem: O desenvolvimento apresentado para $\mathrm{IVAR}(\hat{f}_n)$ é consequência imediata do Teorema 2.5.1 e da desigualdade

$$\frac{1}{h_n}\sum_{j\in\mathbb{Z}} p_{n,j}^2 \leq R(f),$$

que podemos deduzir de (2.5.2) e (2.5.3).

A análise do termo de viés é mais delicada. Para $y \in \Delta_{n,j}$ comecemos por considerar um desenvolvimento de Taylor de segunda ordem de $f(y)$ numa vizinhança do ponto médio $x^*_{n,j}$ do intervalo $\Delta_{n,j}$:

$$f(y) = f(x^*_{n,j}) + (y - x^*_{n,j}) f'(x^*_{n,j}) + \frac{1}{2}(y - x^*_{n,j})^2 f''(c_{n,j}(y)), \qquad (2.6.2)$$

onde $c_{n,j}(y) \in]y, x^*_{n,j}[$. Assim, para $x \in \Delta_{n,j}$, pelo teorema do valor médio para integrais (cf. Lima, 1995, p. 260) temos

$$\begin{aligned}
\mathrm{E}\hat{f}_n(x) &= \frac{1}{h_n} \int_{\Delta_{n,j}} f(y) dy \\
&= f(x^*_{n,j}) + f'(x^*_{n,j}) \frac{1}{h_n} \int_{\Delta_{n,j}} (y - x^*_{n,j}) dy \\
&\quad + \frac{1}{2h_n} \int_{\Delta_{n,j}} (y - x^*_{n,j})^2 f''(c_{n,j}(y)) dy \\
&= f(x^*_{n,j}) + \frac{1}{2h_n} f''(c^*_{n,j}) \int_{\Delta_{n,j}} (y - x^*_{n,j})^2 dy \\
&= f(x^*_{n,j}) + \frac{1}{24} h_n^2 f''(c^*_{n,j}),
\end{aligned}$$

com $c^*_{n,j} \in [a_{n,j}, a_{n,j+1}]$ atendendo à continuidade da função $y \to f''(c_{n,j}(y))$ que podemos deduzir de (2.6.2).

Finalmente, concluímos que, para $x \in \Delta_{n,j}$,

$$\begin{aligned}
&\mathrm{E}\hat{f}_n(x) - f(x) \qquad (2.6.3) \\
&= -(x - x^*_{n,j}) f'(x^*_{n,j}) - \frac{1}{2}(x - x^*_{n,j})^2 f''(c_{n,j}(x)) + \frac{1}{24} h_n^2 f''(c^*_{n,j}).
\end{aligned}$$

Para ilustrar a técnica envolvida na análise das ordens de convergência das parcelas que se obtêm por desenvolvimento do quadrado de $\mathrm{E}\hat{f}_n(x) - f(x)$, consideremos a parcela relativa ao quadrado do primeiro termo de (2.6.3). Temos então

$$\begin{aligned}
\sum_{j \in \mathbb{Z}} \int_{\Delta_{n,j}} (x - x^*_{n,j})^2 f'(x^*_{n,j})^2 dx &= \sum_{j \in \mathbb{Z}} \frac{1}{12} h_n^3 f'(x^*_{n,j})^2 \\
&= \frac{h_n^2}{12} R(f') + O\left(h_n^3\right),
\end{aligned}$$

pois, pelo Lema 2.5.4,

$$\left| \sum_{j\in\mathbb{Z}} h_n \, f'(x^*_{n,j})^2 - R(f') \right| \leq 2h_n \int |f'(x) f''(x)| dx.$$

Procedendo de forma análoga para as restantes parcelas (e usando o teorema do valor médio para integrais quando necessário), encontramos parcelas de ordem inferior ou igual a h_n^3 e portanto

$$\text{IVIES}(\hat{f}_n) = \sum_{j\in\mathbb{Z}} \int_{\Delta_{n,j}} (\text{E}\hat{f}_n(x) - f(x))^2 dx = \frac{h_n^2}{12} R(f') + O\left(h_n^3\right).$$

□

Nas condições do teorema anterior, a maior ordem de convergência para zero do erro quadrático médio integrado é obtida quando tomamos

$$h_n = c\, n^{-1/3}, \text{ com } c > 0,$$

obtendo-se neste caso

$$\text{EQMI}(\hat{f}_n) = O\big(n^{-2/3}\big).$$

Janela assintoticamente óptima

Designemos por erro quadrático médio integrado assintótico, $\text{EQMIA}(h_n)$, a soma das duas parcelas mais significativas do desenvolvimento do $\text{EQMI}(\hat{f}_n)$ dado no Teorema 2.6.1:

$$\text{EQMIA}(h_n) = \frac{1}{nh_n} + \frac{h_n^2}{12} R(f').$$

A janela óptima no sentido da minimização do erro quadrático médio integrado assintótico é derivada no resultado seguinte.

Teorema 2.6.4. *Se f é tal que $R(f') > 0$, o valor de h_n que minimiza o* $\text{EQMIA}(h_n)$ *é dado por*

$$h_{\text{EQMIA}} = \big(6/R(f')\big)^{1/3} \, n^{-1/3}.$$

Além disso, nas condições do Teorema 2.6.1 temos

$$\mathrm{EQMI}(\hat{f}_n^{\circ}) = \left(3R(f')/32\right)^{1/3}\, n^{-2/3} + O(n^{-1}),$$

onde $\hat{f}_n^{\circ}$ *representa o estimador do histograma com janela* h_{EQMIA}.

Dem: Basta ter em conta que a função a minimizar é da forma

$$\psi(h) = Ah^{\alpha} + Bh^{-\beta},\ h > 0,$$

com $A, B, \alpha, \beta > 0$, e esta função atinge mínimo absoluto no ponto

$$h^* = (\beta B/(\alpha A))^{1/(\alpha+\beta)}$$

dado por

$$\psi(h^*) = (\alpha+\beta)\left((A/\beta)^{\beta}(B/\alpha)^{\alpha}\right)^{1/(\alpha+\beta)}.$$

□

Observemos que a função aleatória $\hat{f}_n^{\circ}$ definida à custa da janela assintoticamente óptima h_{EQMIA} não é efectivamente um estimador da densidade uma vez que a janela h_{EQMIA} depende da densidade que pretendemos estimar através da funcional $R(f')$. A função aleatória $\hat{f}_n^{\circ}$ é habitualmente designada por pseudo-estimador ou oráculo (ver Tsybakov, 2004, pp. 53–54).

A quantidade $R(f')$ pode ser interpretada como uma medida da variabilidade global de f. Quando $R(f')$ é pequeno, a densidade não apresenta grande variabilidade e a janela óptima anterior conduz a uma escolha grande de h_n. Quando a variabilidade de f é grande, a janela óptima é pequena tentando captar tais características de variabilidade. Reparemos também que quanto maior é a quantidade $R(f')$, mais difícil é estimar f através dum histograma.

EQMI *vs.* EQMIA

Na Figura 2.6.5 usamos o Teorema 2.5.1 para ilustrar o efeito da escolha da janela na variância e no viés do estimador do histograma, medidos por cada uma das parcelas IVAR e IVIES, no caso das distribuições normal e

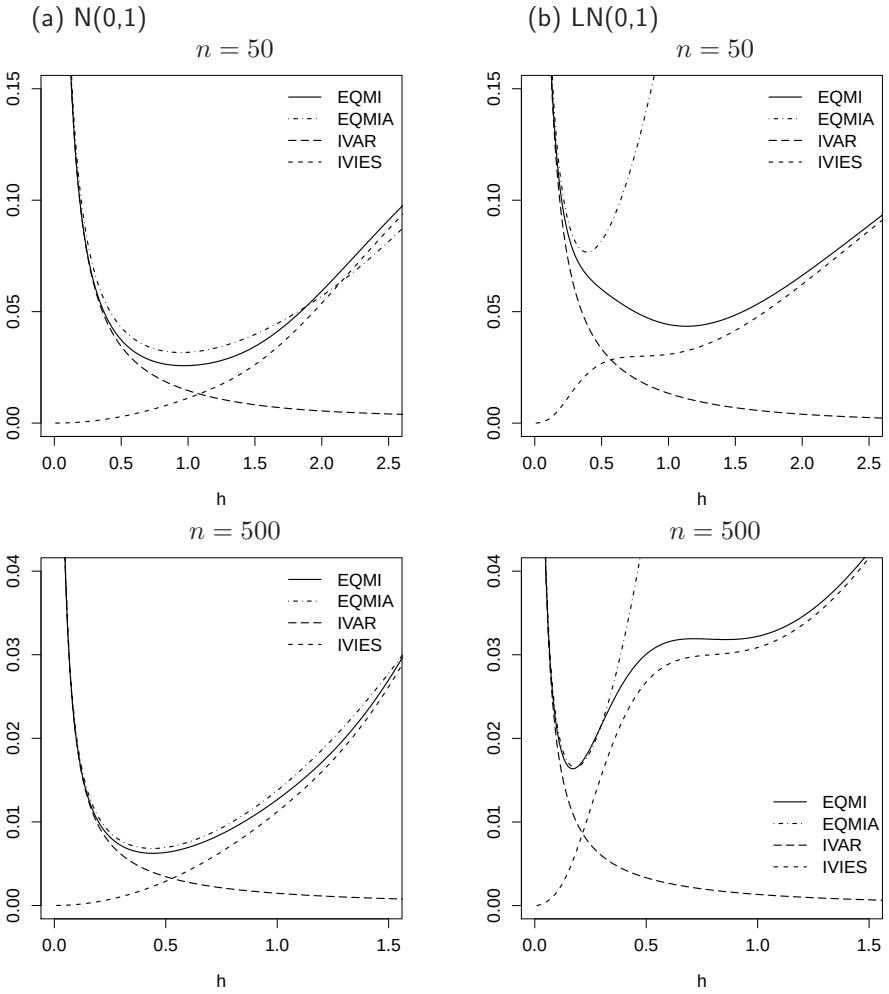

Figura 2.6.5: EQMI, EQMIA, IVAR *e* IVIES *em função de* $h_n = h$ *para as distribuições* $N(0,1)$ *e* $LN(0,1)$ *com* $n = 50$ *e* $n = 500$.

lognormal uma das parcelas IVAR e IVIES, no caso das distribuições normal e lognormal standard. Em todos os casos tomámos o valor $a_{n,0} = 0$ como origem das diversas partições. O carácter local e assintótico da aproximação do erro quadrático médio integrado dada no Teorema 2.6.1 é também bem visível no caso da distribuição lognormal. Em particular, para valores pequenos de n, vemos que a janela assintoticamente óptima é significativa-

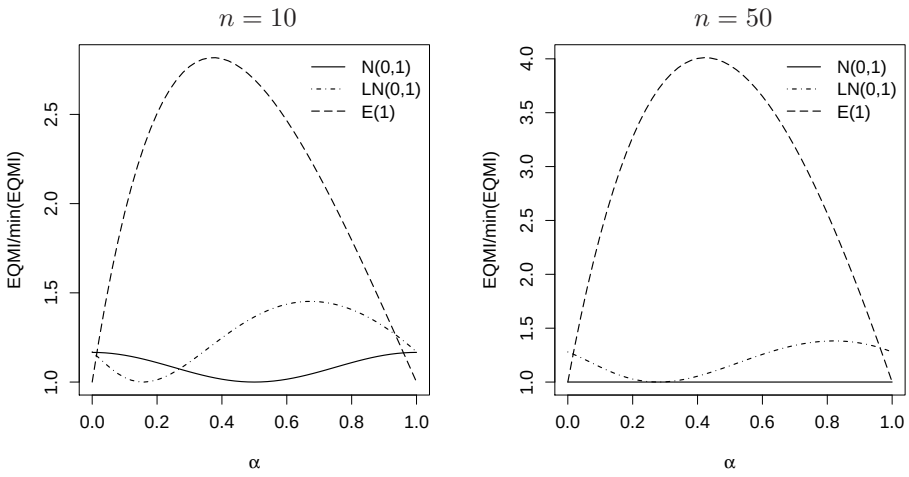

Figura 2.7.1: EQMI/ min(EQMI) *em função da origem* $a_{n,0} = \alpha h_{\text{EQMIA}}$ *da partição para as distribuições N(0,1), LN(0,1) e E(1) com* $n = 10$ *e* $n = 50$.

mente diferente da óptima. Essa discrepância é atenuada quando o tamanho da amostra aumenta.

2.7 Influência da origem da partição

Para uma densidade de probabilidade nas condições do Teorema 2.6.1 verificamos que a origem da partição não tem um papel determinante no comportamento assintótico do EQMI, sendo este dominado pelo comportamento da janela h_n. Atendendo ao caracter assintótico de tais conclusões, vamos neste parágrafo lançar mão do Teorema 2.5.1 para ilustrar a influência da escolha da origem da partição no comportamento do histograma. Para o efeito tomamos as distribuições normal e lognormal, já consideradas na secção anterior, e também a distribuição exponencial de parâmetro 1, e para um conjunto de partições com janela $h_n = h_{\text{EQMIA}}$ e origem $a_{n,0} = \alpha h_n$, para $\alpha \in [0, 1[$, calculamos o EQMI$(\hat{f}_n)$ respectivo que representamos, em função de α, na Figura 2.7.1.

No caso da distribuição normal, e como seria de esperar atendendo ao Teorema 2.5.1, verificamos que o melhor desempenho ocorre no caso em que o centro de simetria da distribuição é o ponto médio de uma das classes da partição. O histograma não é, para as distribuições normal e lognormal, muito sensível à escolha da origem da partição em particular quando o tamanho da amostra aumenta. O mesmo não se passa com a distribuição exponencial o que pode ser explicado pelo facto da densidade exponencial ser descontínua na origem.

Para $\alpha \in]0,1[$, consideremos a partição definida a partir da janela h_n com origem no ponto $a_{n,0} = \alpha h_n$. Para $x \in](\alpha-1)h_n, \alpha h_n]$, temos

$$\mathrm{E}\hat{f}_n(x) = \frac{1}{h_n}\int_0^{\alpha h_n} e^{-t}dt = \frac{1-e^{-\alpha h_n}}{h_n},$$

não sendo o histograma assintoticamente cêntrico para $x = 0$.

Além disso, temos

$$\begin{aligned}
&\int_{(\alpha-1)h_n}^{\alpha h_n} \mathrm{Viés}\hat{f}_n(x)^2 dx \\
&= \int_{(\alpha-1)h_n}^{0} \left(\frac{1-e^{-\alpha h_n}}{h_n} - 0\right)^2 dx + \int_0^{\alpha h_n} \left(\frac{1-e^{-\alpha h_n}}{h_n} - e^{-x}\right)^2 dx \\
&= \alpha(1-\alpha)h_n - \alpha^2(1-\alpha)h_n^2 + O(h_n^3),
\end{aligned}$$

e portanto

$$\mathrm{IVIES}(\hat{f}_n) = \alpha(\alpha-1)h_n + O(h_n^2),$$

ordem esta que é inferior à ordem que se obtém quando tomamos $a_{n,0} = 0$ para origem das diversas partições.

2.8 Escolha prática de h_n

A escolha da janela h_n é uma questão essencial na implementação prática do estimador do histograma. Uma das mais conhecidas regras práticas para a escolha do tamanho da janela num histograma foi proposta por Sturges (1926):

$$\hat{h}_{\mathrm{ST}} = \frac{\text{amplitude amostral}}{1+\log_2 n}.$$

Apesar do seu valor histórico, esta regra não é baseada num critério que tenha em conta a qualidade do histograma como estimador da densidade desconhecida f. Tendo em conta o que vimos atrás, esta janela conduzirá a um estimador com pouca variabilidade mas com viés elevado que produzirá estimativas demasiadamente suaves (*oversmoothing*).

De forma breve, apresentamos nos parágrafos seguintes alguns métodos para a escolha automática da janela. Outros métodos para escolha da janela foram considerados por Scott e Terrel (1987) e Wand (1997).

2.8.1 Métodos de utilização simples

Como vimos atrás, para uma vasta classe de densidades conhecemos a janela óptima no sentido da minimização do erro quadrático médio integrado assintótico:

$$h_{\text{EQMIA}} = \left(6/R(f')\right)^{1/3} n^{-1/3}.$$

No entanto, esta janela não pode ser directamente usada para escolher h_n uma vez que sendo f desconhecida, é também desconhecida a quantidade $R(f')$.

Os procedimentos seguintes para a escolha da janela são baseados na ideia de estimar de forma paramétrica a quantidade desconhecida que intervém na expressão de h_{EQMIA}. Este método geral para a escolha da janela, geralmente designado por método das distribuições de referência, foi primeiramente sugerido por Deheuvels (1977a) no contexto dos estimadores do núcleo da densidade que estudaremos no próximo capítulo.

Janela com distribuição de referência normal

Scott (1979) propõe que se utilize a regra anterior onde $R(f')$ é calculada tomando para f a densidade da distribuição $N(0, \sigma^2)$ ($R(f')$ não depende do parâmetro de posição). Neste caso,

$$R(f') = \frac{1}{4\sqrt{\pi}\,\sigma^3}$$

e

$$h_{\text{EQMIA}} = \left(24\sqrt{\pi}\right)^{1/3} \sigma\, n^{-1/3} \approx 3.49\, \sigma\, n^{-1/3}.$$

Estimando o parâmetro de escala σ pelo desvio-padrão empírico corrigido $\hat{s} = \sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)}$, obtemos a janela com distribuição de referência normal:

$$\hat{h}_{\text{NR}} = 3.49\, \hat{s}\, n^{-1/3}.$$

Havendo informação preliminar sobre a forma da distribuição, para distribuições fortemente assimétricas ou com caudas pesadas, Scott (1979) sugere escolhas da janela baseadas em outras distribuições de referência como a lognormal ou a distribuição de Student.

Regra de Silverman

Atendendo a que no caso duma distribuição normal com função de distribuição F se tem

$$\sigma = \frac{F^{-1}(3/4) - F^{-1}(1/4)}{F^{-1}_{N(0,1)}(3/4) - F^{-1}_{N(0,1)}(1/4)},$$

o desvio-padrão σ pode ser estimado por $\hat{\text{AIQ}}/1.349$, onde $\hat{\text{AIQ}}$ é a amplitude inter-quartil empírica $\hat{\text{AIQ}} = X_{[0.75n]} - X_{[0.25n]}$. Silverman (1986) sugere combinar os dois estimadores dando origem à janela

$$\hat{h}_{\text{SIL}} = \min(3.49\, \hat{s}, 2.59\, \hat{\text{AIQ}})\, n^{-1/3}.$$

Evocando razões de robustez, Deheuvels (1977a) advogava já a estimação do parâmetro de escala a partir de estimadores baseados nas estatísticas ordinais.

O princípio da distribuição mais suave

No sentido de evitar a escolha de uma janela demasiado pequena que pode conduzir a uma representação gráfica com características artificiais de multimodalidade, Terrell e Scott (1985) e Terrel (1990) advogam que a janela h_n deveria ser escolhida de modo a ser a maior possível que seja compatível

com a escala das observações. Atendendo à forma assintoticamente óptima para a janela dada no Teorema 2.6.4 e tomando o desvio-padrão σ como parâmetro de escala, o princípio anterior conduz a tomar como distribuição de referência a densidade f com variância σ^2 que torna mínima a quantidade $R(f')$. Podendo esta quantidade ser interpretada como uma medida da variabilidade global de f, a distribuição de referência é, no sentido anterior, a densidade mais suave possível, razão pela qual o critério anterior é designado por princípio da distribuição mais suave (*oversmoothing principle*). Terrel (1990) prova que a densidade com variância σ^2 que é solução do problema anterior é dada por g_δ onde $g(x) = \frac{15}{16}(1-x^2)^2 I(|x| \leq 1)$ e $\delta = \sqrt{7}\sigma$. Como $R(g') = 15/7$ obtemos a janela

$$\hat{h}_{\mathrm{OS}} = 3.73\, \hat{s}\, n^{-1/3}.$$

2.8.2 O método de validação cruzada

Não é de esperar que os métodos de utilização simples descritos no parágrafo anterior, produzam janelas de boa qualidade quando a distribuição subjacente aos dados é 'muito diferente' da distribuição usada para distribuição de referência. Rudemo (1982) e Bowman (1984) tomam o erro quadrático integrado, $\mathrm{EQI} = ||\hat{f}_n - f||_2^2$, como medida da discrepância entre $\hat{f}_n$ e a densidade desconhecida f, e propõem um método simples para a escolha da janela que ficaria conhecido como método de validação cruzada baseado no EQI (*least-square cross-validation*). A ideia é assim a de escolher a janela $h = h_n$ que minimiza um estimador cêntrico do erro quadrático integrado

$$\begin{aligned} \mathrm{EQI}(h) &= \int \{\hat{f}_n(x) - f(x)\}^2 dx \\ &= R(\hat{f}_n) - 2\int \hat{f}_n(x) f(x)\, dx + R(f). \end{aligned}$$

Não dependendo $R(f)$ de h, esta parcela não tem qualquer influência na minimização de $\mathrm{EQI}(h)$. Relativamente à segunda parcela, Rudemo nota que ela pode ser interpretada como a esperança matemática $\mathrm{E}(\hat{f}_n(X))$, onde $X_1, \ldots, X_n$ são considerados fixos. Para aproximar a quantidade anterior

a partir das observações $X_1, \ldots, X_n$, consideram-se n sub-amostras de tamanho $n-1$ cada uma delas obtida a partir da amostra original suprimindo uma observação. Denotando por $\hat{f}_{n,-i}$ o histograma associado à sub-amostra $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$, será natural propor para aproximar $\mathrm{E}(\hat{f}_n(X))$ a estatística

$$\frac{1}{n}\sum_{i=1}^{n} \hat{f}_{n,-i}(X_i).$$

Reparemos que o facto de não se incluir em $\hat{f}_{n,-i}$ a observação X_i, faz com que a estatística anterior e $\int \hat{f}_n(x) f(x)\, dx$ tenham a mesma esperança matemática.

Assim,

$$\mathrm{E}\left(R(\hat{f}_n) - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{n,-i}(X_i) \right) = \mathrm{EQMI}(\hat{f}_n) - R(f).$$

Na prática toma-se para janela o valor $\hat{h}_{\mathrm{CV}}$ que minimiza

$$\mathrm{CV}(h) = R(\hat{f}_n) - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{n,-i}(X_i) = \frac{2}{(n-1)h} - \frac{n+1}{n^2(n-1)h}\sum_{j\in\mathbb{Z}} \nu_{n,j}^2$$

onde $\nu_{n,j} = \sum_{i=1}^{n} I(X_i \in \Delta_{n,j})$. Reparemos que para cada h a partição que define o histograma é alterada sendo necessário recalcular $\nu_{n,j}$.

Sob condições pouco restritivas sobre o f, Stone (1985) mostra que $\hat{h}_{\mathrm{CV}}$ é assintoticamente óptima no sentido em que

$$\frac{\mathrm{EQI}(\hat{h}_{\mathrm{CV}})}{\mathrm{EQI}(\hat{h}_{\mathrm{EQI}})} \xrightarrow{qc} 1,$$

onde

$$\hat{h}_{\mathrm{EQI}} = \mathrm{argmin}_{h>0}\mathrm{EQI}(h).$$

2.8.3 Aplicação a um conjunto de dados reais

O conjunto de dados que consideramos neste parágrafo é relativo a duas variáveis observadas aquando de 272 erupções sucessivas do Old Faithful

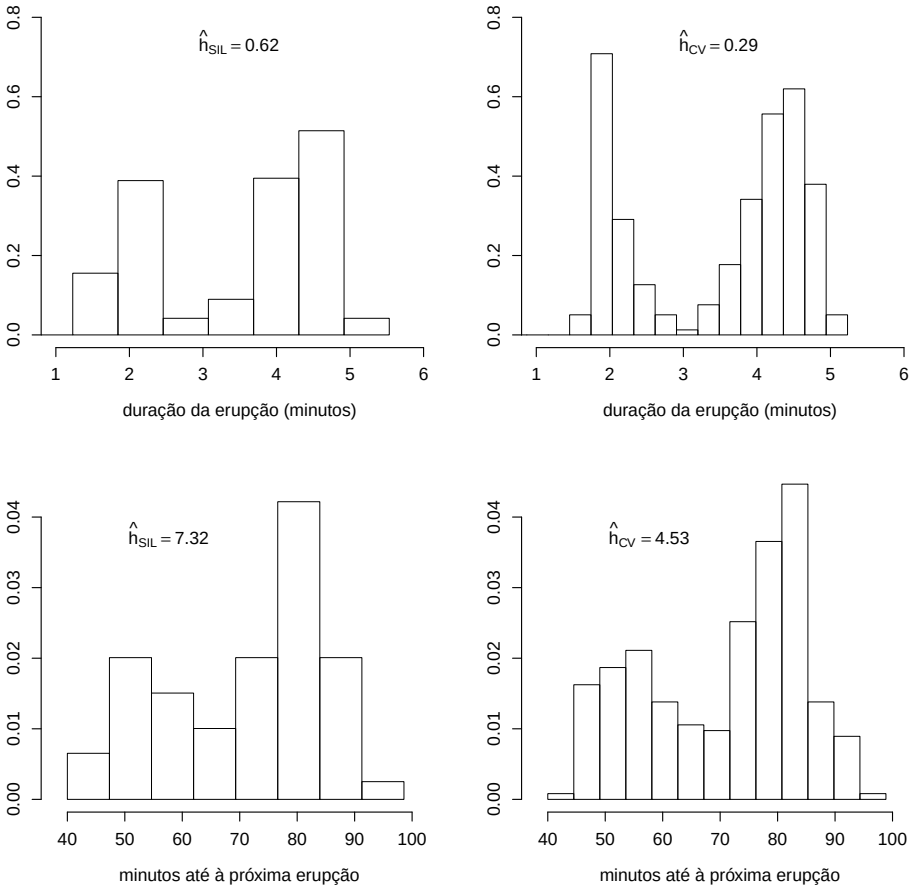

Figura 2.8.1: *Histogramas obtidos a partir das janelas* $\hat{h}_{\mathrm{SIL}}$ *e* $\hat{h}_{\mathrm{CV}}$.

Geyser no Yellowstone National Park nos Estados Unidos da América. A primeira variável é relativa à duração de cada uma das erupções, enquanto que a segunda variável diz respeito ao período de tempo que medeia entre cada uma das erupções e a erupção seguinte. Ambas as variáveis são registadas em minutos. Existem várias versões deste conjunto de dados analisadas por diversos autores como, por exemplo, Silverman (1986) ou Härdle (1991). A versão que aqui usamos está disponível na documentação do R (*faithful*).

Os histogramas da Figura 2.8.1 relativos aos conjuntos de dados descritos, são obtidos tomando para origem da partição $a_{n,0} = 0$ e $a_{n,0} = 40$, respectivamente, tendo as duas janelas usadas sido obtidas pela regra de Silverman e pelo método de validação cruzada. Os histogramas põem em evidência o carácter marcadamente bimodal de ambas as distribuições, mas as janelas $\hat{h}_{\mathrm{SIL}}$ são significativamente maiores que as janelas $\hat{h}_{\mathrm{CV}}$. Os valores aparentemente demasiado grandes das janelas obtidas pela regra Silverman podem ser explicados pelo facto das distribuições subjacentes possuírem características marcadas de não normalidade. Tal facto pode ter levado a uma subestimação das quantidades $R(f')$ o que conduziu a janelas demasiado grandes.

2.9 O polígono de frequências

O facto já discutido em §2.2 da ordem de convergência do viés do estimador do histograma não poder ser superior a h_n, mesmo quando a densidade subjacente apresenta derivadas de segunda ordem contínuas, é uma característica negativa deste estimador. A classe de estimadores que estudaremos no próximo capítulo permite, como veremos, contornar esta dificuldade. É, no entanto, interessante verificar que um outro estimador bem nosso conhecido possui um viés e uma variabilidade inferiores aos do histograma. Trata-se do polígono de frequências definido por

$$\hat{g}_n(x) = \left(\frac{1}{2} - \frac{x - a_{nj}}{h_n}\right) \hat{f}_n\Big(a_{n,j} - \frac{h_n}{2}\Big) + \left(\frac{1}{2} + \frac{x - a_{nj}}{h_n}\right) \hat{f}_n\Big(a_{n,j} + \frac{h_n}{2}\Big),$$

para $x \in]a_{n,j} - h_n/2, a_{n,j} + h_n/2]$ com $j \in \mathbb{Z}$. O estudo das propriedades teóricas deste estimador é iniciado por Scott (1985) (ver também Scott, 1992).

Viés e variância

A observação que fizemos a seguir ao Teorema 2.2.5, acerca do viés do estimador do histograma nos pontos médios da partição, explica o facto

do polígono de frequências ter um viés inferior ao histograma. Com efeito, atendendo ao Teorema 2.2.5, sabemos que se f' é de Lipschitz (de ordem 1) numa vizinhança de x então

$$\mathrm{E}\hat{f}_n\Big(a_{n,j} \pm \frac{h_n}{2}\Big) = f\Big(a_{n,j} \pm \frac{h_n}{2}\Big) + O(h_n^2)$$

onde

$$\begin{aligned} f\Big(a_{n,j} \pm \frac{h_n}{2}\Big) &= f(x) \pm h_n\Big(\frac{1}{2} \mp \frac{x - a_{nj}}{h_n}\Big) f'(x) + O(h_n^2) \\ &= f(x) + O(h_n^2). \end{aligned}$$

O viés do polígono de frequências é assim da ordem de h_n^2 e não da ordem de h_n, como acontece com o histograma:

$$\mathrm{E}\hat{g}_n(x) = f(x) + O(h_n^2).$$

Podemos ser ainda mais precisos. Se admitirmos que f possui derivada de segunda ordem contínua numa vizinhança de x, vale o desenvolvimento

$$\mathrm{E}\hat{g}_n(x) = f(x) + \left(\frac{1}{3} - \Big(\frac{x - a_{nj}}{h_n}\Big)^2\right) \frac{h_n^2}{2} f''(x) + o(h_n^2).$$

Usando agora (2.2.6), concluiríamos de forma análoga que a variância do polígono de frequências é dada por

$$\mathrm{Var}\hat{g}_n(x) = \left(\frac{1}{2} + 2\Big(\frac{x - a_{nj}}{h_n}\Big)^2\right) \frac{1}{nh_n} f(x) + O(n^{-1}),$$

sendo o seu termo mais representativo inferior ao do estimador do histograma.

Ordem de convergência óptima

Tomando $h_n = c\, n^{-1/5}$, com $c > 0$, concluímos das expressões anteriores que se f' é de Lipschitz (de ordem 1) numa vizinhança de x então

$$\mathrm{EQM}(\hat{g}_n(x)) = O\big(n^{-4/5}\big),$$

o que supera a ordem $n^{-2/3}$ obtida para o estimador do histograma.

Uma pergunta natural é a de saber se é ainda possível melhorar a ordem de convergência anterior. Existirá algum estimador da densidade cujo erro quadrático médio convirja para zero com uma ordem superior a $n^{-4/5}$? Para que uma tal questão tenha interesse, é essencial que a ordem de convergência anterior seja relativa a um conjunto fixo (e razoavelmente vasto) de densidades de probabilidade. Com efeito, tal acontece relativamente ao conjunto $\mathscr{C}_{2\alpha}$ de densidades f com derivada absolutamente contínua e cuja derivada de segunda ordem satisfaz $||f''||_\infty \leq \alpha$. Adaptando as técnicas que temos vindo a desenvolver podemos provar que

$$\limsup_{n\to\infty} \sup_{f\in\mathscr{C}_{2\alpha}} n^{4/5}\,\mathrm{EQM}(\hat{g}_n(x)) < \infty.$$

O resultado que enunciamos a seguir, devido a Farrel (1972), garante que a ordem de convergência anterior é a melhor possível. Farrel estabelece que não existe um estimador da densidade cujo erro quadrático médio no ponto x convirja para zero, uniformemente no conjunto $\mathscr{C}_{2\alpha}$, com uma ordem superior a $n^{-4/5}$.

Para $k \in \mathbb{N}$ e $\alpha > 0$, denotamos por $\mathscr{C}_{k\alpha}$ o conjunto das densidades de probabilidade f com derivada de ordem $k-1$ absolutamente contínua e cuja derivada de ordem k, que existe em quase todo o ponto, satisfaz $||f^{(k)}||_\infty \leq \alpha$.

Teorema 2.9.1. *Para $k \in \mathbb{N}$ e $\alpha > 0$ se $\gamma_n(\cdot) = \gamma_n(\cdot; X_1, \ldots, X_n)$ é tal que*

$$\limsup_{n\to\infty} \sup_{f\in\mathscr{C}_{k\alpha}} a_n^{-2}\,\mathrm{EQM}(\gamma_n(x)) < \infty$$

com $a_n \to 0$, então

$$\liminf_{n\to\infty} n^{-2k/(2k+1)} a_n^2 > 0.$$

Do resultado anterior decorre também que qualquer estimador da densidade possui um erro quadrático médio cuja ordem de convergência para zero, uniformemente em $\mathscr{C}_{k\alpha}$, não pode ser superior a $n^{-2k/(2k+1)}$. Não sendo esta ordem de convergência atingida pelo polígono de frequências

quando $k > 2$, fica em aberto a questão da existência de estimadores da densidade que atinjam tal ordem de convergência. Voltaremos a ela no próximo capítulo.

EQMI e janela assintoticamente óptima

O resultado seguinte de Scott (1985), descreve o comportamento assintótico do EQMI associado ao polígono de frequências. Por comparação com o correspondente resultado obtido em §2.5, mais uma vez se ilustra, agora usando uma medida global da qualidade do estimador, a superioridade assintótica do polígono de frequências relativamente ao histograma.

Teorema 2.9.2. *Suponhamos que f possui derivada de terceira ordem, com f'' absolutamente contínua e f, f', f'', f''' de quadrado integrável. Se $h_n \to 0$ então*

$$\mathrm{EQMI}(\hat{g}_n) = \frac{2}{3nh_n} + \frac{49}{2880}\, h_n^4 R(f'') + O(n^{-1}) + O(h_n^5).$$

A janela óptima no sentido da minimização da soma das duas parcelas mais significativas do desenvolvimento anterior é assim dada por

$$h_{\mathrm{EQMIA}} = 2\left(15/(49R(f''))\right)^{1/5}\, n^{-1/5}$$

a que corresponde o erro quadrático médio integrado

$$\mathrm{EQMI}(\hat{g}_n^\circ) = \frac{5}{12}\left(49R(f'')/15\right)^{1/5}\, n^{-4/5} + O(n^{-1}), \tag{2.9.3}$$

onde $\hat{g}_n^\circ$ o estimador do polígono de frequências com janela h_{EQMIA}.

2.10 Histogramas multivariados

Quando as observações $X_1, \ldots, X_n$ são vectores aleatórios independentes e identicamente distribuídos com valores em $\mathbb{R}^d$ e densidade de probabilidade f, os resultados que apresentámos ao longo deste capítulo admitem versões

multidimensionais para o estimador do histograma definido a partir da partição

$$\Delta_{n,j_1,\ldots,j_d} = \prod_{i=1}^{d} \,]\, a^i_{n,j_i}, a^i_{n,j_i+1}],\ j_1,\ldots,j_d \in \mathbb{Z},$$

com origem $(a^1_{n,0},\ldots,a^d_{n,0})$, onde $a^i_{n,j+1} = a^i_{n,j} + h_i$, $j \in \mathbb{Z}$, e $h_i = h_{n,i} > 0$. Para $x \in \Delta_{n,j_1,\ldots,j_d}$, o estimador do histograma é definido por

$$\hat{f}_n(x) = \frac{1}{nh_1\ldots h_d} \sum_{i=1}^{n} I(X_i \in \Delta_{n,j_1,\ldots,j_d}).$$

Uma forma simplificada do estimador do histograma pode ainda ser obtida assumindo que $h_i = h$ para todo o $i = 1,\ldots,d$. Neste caso o estimador toma a forma

$$\hat{f}_n(x) = \frac{1}{nh^d} \sum_{i=1}^{n} I(X_i \in \Delta_{n,j_1,\ldots,j_d}),$$

para $x \in \Delta_{n,j_1,\ldots,j_d}$.

A escolha da janela

Tal como no caso real, sob condições de regularidade sobre f e assumindo que $h_i \to 0$ e $nh_1\ldots h_d \to +\infty$, podemos obter um desenvolvimento assintótico para o erro quadrático médio integrado de $\hat{f}_n$ (cf. Scott, 1992; Lecoutre, 1985) dando origem ao EQMIA definido por

$$\text{EQMIA} = \frac{1}{nh_1\ldots h_d} + \frac{1}{12}\sum_{i=1}^{d} h_i^2 R(f_i),$$

onde f_i denota a derivada parcial de f relativamente à sua i-ésima variável. Da minimização do EQMIA relativamente a cada uma das janelas h_k resulta a seguinte expressão para a janela assintoticamente óptima

$$h_{k,\text{EQMIA}} = R(f_k)^{-1/2}\left(6\prod_{i=1}^{d} R(f_i)^{1/2}\right)^{1/(d+2)} n^{-1/(d+2)}.$$

d	Tamanhos equivalentes da amostra		
1	10	50	100
2	39	333	838
3	172	2510	7967
4	838	20943	83776
5	4446	190058	957834

Tabela 2.10.2: *Tamanhos equivalentes da amostra, em função da dimensão d, para a distribuição normal multivariada standard segundo o critério de Epanechnikov.*

Para uma escolha assintoticamente óptima das janelas $h_1, \ldots, h_d$, obtemos

$$\text{EQMIA} = (1 + d/2)\, 6^{-d/(d+2)} \left(\prod_{i=1}^{d} R(f_i) \right)^{1/(d+2)} n^{-2/(d+2)}. \qquad (2.10.1)$$

Tomando como distribuição de referência a distribuição normal $N(0, \Sigma)$, com $\Sigma = diag(\sigma_1^2, \ldots, \sigma_d^2)$, temos

$$h_{k,\text{EQMIA}} = 2 \cdot 3^{1/(d+2)} \pi^{d/(2d+4)} \sigma_k \, n^{-1/(d+2)},$$

expressão esta que pode ser usada para propor uma escolha prática da janela.

A maldição da dimensão

A expressão (2.10.1) põe em evidência um dos problemas da estimação da densidade num contexto multivariado, habitualmente designado por *maldição da dimensão* (*curse of dimensionality*): a order de convergência do EQMIA decresce drasticamente à medida que a dimensão d aumenta, sendo bastante inferior à ordem de convergência paramétrica n^{-1}.

A gravidade deste problema pode ser claramente ilustrada adoptando o procedimento proposto por Epanechnikov (1969) para comparar o desempenho de estimadores da densidade em função da dimensão d dos dados (outros critérios são considerados em Scott e Wand, 1991). Para a distribuição

normal multivariada standard $N(0, I)$, apresentamos na Tabela 2.10.2 os tamanhos das amostras que permitem obter, para várias dimensões d, um mesmo valor da medida normalizada da qualidade do estimador dada por

$$\text{EQMIA}/R(f) = (1 + d/2)(\pi/3)^{d/(2+d)} n^{-2/(2+d)}.$$

Sobre a utilização do estimador do histograma em contextos multidimensionais ver também Scott e Sain (2005) e Scott (2004).

3

O estimador do núcleo

Definição do estimador. O estimador do núcleo como soma de funções elementares. Propriedades locais de convergência. Desenvolvimentos assintóticos para o viés e variância num ponto. Propriedades globais de convergência. Desenvolvimento assintótico para o erro quadrático médio integrado. Escolha assintoticamente óptima de h_n e de K. Núcleos de ordem superior e redução do viés. Escolha prática de h_n. Estimador automático do núcleo. Estimação em pontos fronteiros. Estimador multivariado do núcleo.

3.1 Definição do estimador

Dada uma função real de variável real integrável K, com

$$\int K(x)\,dx = 1,$$

e uma sucessão (h_n) de números reais estritamente positivos que admitimos satisfazer

$$h_n \to 0,\ n \to +\infty,$$

o estimador do núcleo da densidade desconhecida f baseado na amostra $X_1, \ldots, X_n$ é definido, para $x \in \mathbb{R}$, por

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)$$

(cf. Rosenblatt, 1956; Parzen, 1962). A função K é designada por núcleo e a h_n chamamos janela. Quando K é uma densidade de probabilidade, habitualmente tomada simétrica relativamente à origem, f_n é também uma densidade de probabilidade. O estimador da janela móvel, considerado nos trabalhos pioneiros de Fix e Hodges (1951), Akaike (1954) e Rosenblatt (1956) (ver §1.3), é o estimador associado ao núcleo uniforme $K(u) = \frac{1}{2}I(|u| \leq 1)$. O núcleo normal $K(u) = (2\pi)^{-1/2}\exp(-u^2/2)$, é um dos mais utilizados na prática.

O estimador do núcleo como soma de funções elementares

Reescrevendo o estimador do núcleo na forma

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x - X_i),$$

onde

$$K_h(x) = K(x/h)/h,$$

para $x \in \mathbb{R}$ e $h > 0$, podemos de forma simples perceber a contribuição de cada observação X_i para f_n. À observação X_i o estimador do núcleo associa a função

$$\varphi_{X_i}(x) = \frac{1}{n}\, K_{h_n}(x - X_i),$$

que, caso o núcleo K seja simétrico, é simétrica relativamente ao ponto $u = X_i$ com integral

$$\int \varphi_{X_i}(x)dx = \frac{1}{n},$$

e que irá atribuir a ponderação $\varphi_{X_i}(x)$ aos pontos x que estão numa vizinhança de X_i. Apesar de algumas das propriedades que deduziremos serem válidas para núcleos não necessariamente simétricos, o facto de não haver à partida uma razão plausível para tratarmos de forma diferenciada os pontos à esquerda de X_i dos pontos à sua direita, leva a que se tome habitualmente para K uma função simétrica (a este propósito ver Tiago de Oliveira, 1963, p. 34). Como veremos mais à frente, outras razões há para procedermos desta forma.

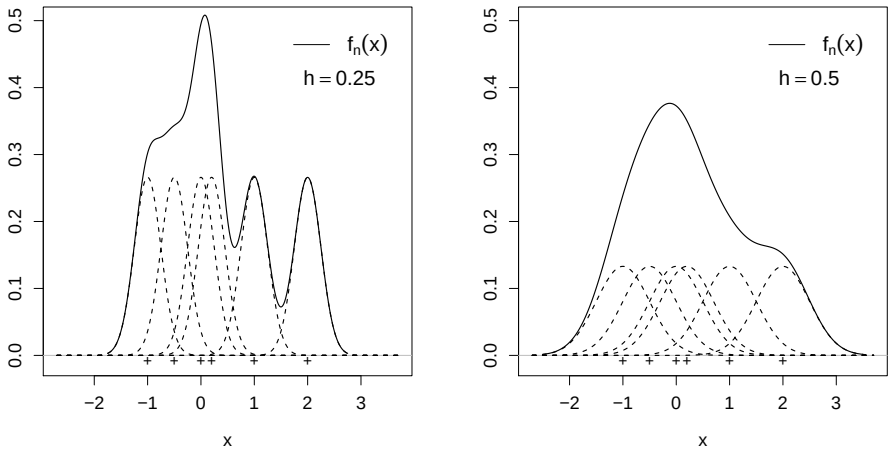

Figura 3.1.1: *Estimador do núcleo como soma de funções elementares. Núcleo normal e janelas* $h = 0.25, 0.5$.

No caso de tomarmos para K o núcleo uniforme sobre o intervalo $[-1, 1]$, todos os pontos do intervalos $[X_i - h_n, X_i + h_n]$ recebem uma ponderação igual a $0.5/n$. Aos restantes pontos é atribuída uma ponderação nula. No caso do núcleo normal, todos os pontos de $\mathbb{R}$ recebem uma ponderação positiva devido à observação X_i. A maior ponderação é atribuída ao ponto $x = X_i$, diminuindo esta à medida que o ponto x se afasta, por excesso ou por defeito, da observação X_i. No caso do núcleo normal o efeito anterior é ilustrado na Figura 3.1.1 onde o estimador do núcleo é construído a partir das observações -1,-0.5,0,0.2,1,2, com $h = 0.25$ e $h = 0.5$.

Exemplos de núcleos

Na Tabela 3.1.2 listamos algumas densidades de probabilidade que podemos usar como núcleos K (ver Figura 3.1.4). Como já referimos, o núcleo uniforme foi primeiramente considerado por Akaike (1954) e Rosenblatt (1956). Os núcleos triangular, normal e de Laplace são considerados em Parzen (1962). O núcleo polinomial de Epanechnikov é primeiramente considerado em Bartlett (1963) mas é com o trabalho de Epanechnikov (1969)

Núcleo	$K(u)$
Uniforme	$\frac{1}{2}I(\|u\| \leq 1)$
Triangular	$(1-\|u\|)I(\|u\| \leq 1)$
Epanechnikov	$\frac{3}{4}(1-u^2)I(\|u\| \leq 1)$
Biweight	$\frac{15}{16}(1-u^2)^2 I(\|u\| \leq 1)$
Triweight	$\frac{35}{32}(1-u^2)^3 I(\|u\| \leq 1)$
Normal	$(2\pi)^{-1/2}\exp(-u^2/2)$
Laplace	$\exp(-\|u\|)/2$
Cauchy de ordem 2	$2\pi^{-1}(1+u^2)^{-2}$

Tabela 3.1.2: *Exemplos de núcleos.*

que ganha notoriedade. Este núcleo, bem como os restantes núcleos polinomiais mencionados, pertence à família

$$K(u;p) = \frac{1}{2^{2p+1}B(p+1,p+1)}\,(1-u^2)^p\,I(|u| \leq 1),$$

onde B é a função beta. O núcleo de Cauchy de ordem 2 é um caso particular da família das densidades de Cauchy de ordem p, considerada em Deheuvels (1977a), definida por

$$K(u;p) = \frac{\Gamma(p)}{\sqrt{\pi}\,\Gamma(-1/2+p)}\,\frac{1}{(1+x^2)^p},$$

onde Γ é a função gama.

Os núcleos anteriores gozam de diferentes propriedades de regularidade, sendo estas propriedades transferidas para o estimador do núcleo. Assim, se K é contínua ou diferenciável, o estimador do núcleo baseado em K gozará dessas mesmas propriedades. Verificaremos mais à frente que, para além das propriedades de regularidade, o núcleo não assume, a nível da

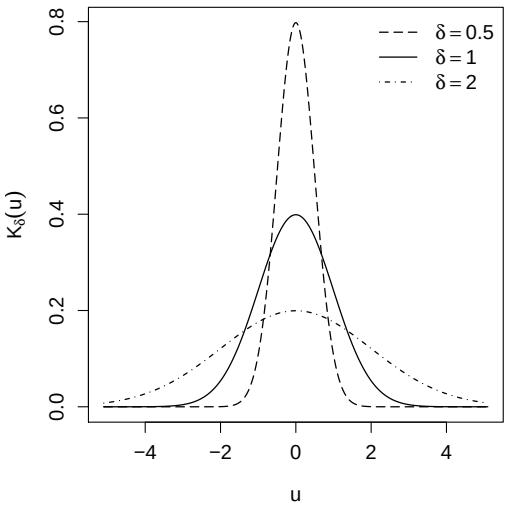

Figura 3.1.3: *Mudanças de escala K_δ com K o núcleo normal standard.*

qualidade do estimador, o papel primordial que a janela irá assumir. Tal como já acontecia no caso do histograma, as propriedades assintóticas e a distância finita do estimador do núcleo são, no essencial, determinadas pela escolha da janela.

Mudança de escala

Para um valor $\delta > 0$, a função definida a partir de K por

$$K_\delta(u) = K(u/\delta)/\delta,$$

é ainda um núcleo que dizemos ter sido obtido de K através da mudança de escala δ. Na Figura 3.1.3 representam-se as mudanças de escala $K_{1/2}$ e K_2 quando K é o núcleo normal. Reparemos que a escolha da escala para um núcleo está intimamente relacionada com a escolha da janela. O estimador que se obtém tomando para núcleo a função K e para janela a sucessão h_n coincide com o estimador de núcleo K_δ e janela h_n/δ.

Uma forma de melhor compararmos estimadores do núcleo construídos a partir de diferentes núcleos é assumir que para além de possuírem o mesmo

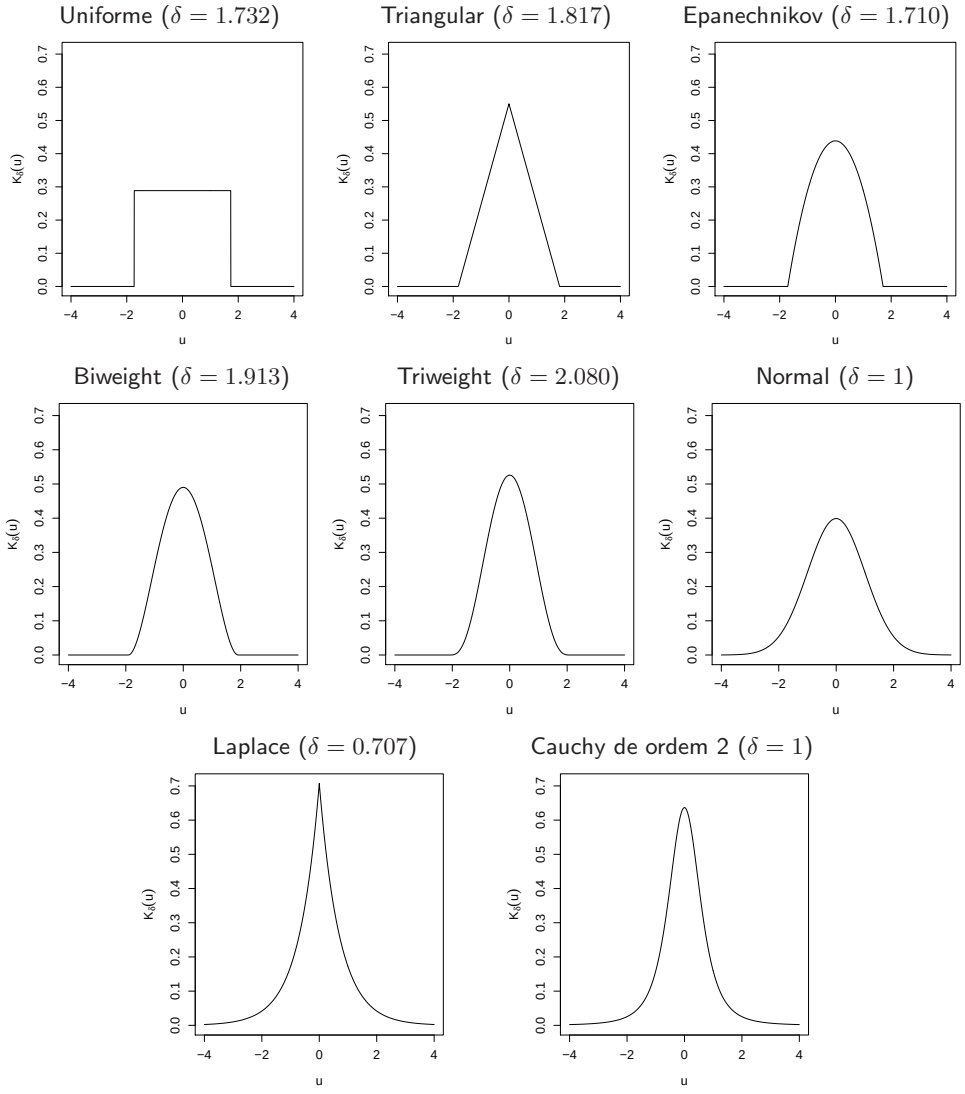

Figura 3.1.4: *Mudanças de escala K_δ para os núcleos K da Tabela 3.1.2.*

momento de primeira ordem $m_1(K) = \int uK(u)du$, que é habitualmente nulo devido à habitual simetria do núcleo relativamente à origem, também partilham de um mesmo momento de segunda ordem $m_2(K) = \int u^2K(u)du$ (que supomos existir). Para os núcleos K da Tabela 3.1.2, representamos

na Figura 3.1.4 os núcleos K_δ onde as mudanças de escala δ são escolhidas de modo que

$$m_2(K_\delta) = \delta^2 m_2(K) = 1.$$

Desta forma, todos os núcleos K_δ representados possuem a mesma variabilidade.

3.2 Propriedades locais de convergência

Iniciamos este parágrafo pelo estudo do comportamento assintótico do viés, variância e erro quadrático médio do estimador do núcleo. Os resultados que apresentamos são no essencial devidos a Rosenblatt (1956), no caso do estimador da janela móvel, e a Parzen (1962) (ver também Tiago de Oliveira, 1963, pp. 28–42). Neste estudo tem um papel de relevo o resultado seguinte devido a Bochner (1955).

Lema 3.2.1 (de Bochner)**.** *Sejam $\varphi : \mathbb{R} \to \mathbb{R}$ uma função integrável e limitada, $g : \mathbb{R} \to \mathbb{R}$ uma função contínua em $x \in \mathbb{R}$ e $h_n \to 0$.*

a) Se g é limitada em $\mathbb{R}$, ou, em alternativa, se g é integrável e φ é tal que $\lim_{|x|\to+\infty} x\varphi(x) = 0$, então

$$\varphi_{h_n} * g(x) = \int \varphi_{h_n}(x-y)g(y)dy \to g(x)\int \varphi(y)dy,$$

onde $*$ denota o produto de convolução.

b) Se g é limitada e uniformemente contínua, então a convergência anterior é uniforme em x.

Dem: No caso em que g limitada, as alíneas a) e b) são consequência da desigualdade seguinte válida para todo o $\delta > 0$ e $h_n > 0$:

$$\begin{aligned}\left|\varphi_{h_n} * g(x) - g(x)\int \varphi(y)dy\right| &= \left|\int \varphi(y)(g(x-yh_n) - g(x))dy\right| \\ &\leq \sup_{|z|\leq\delta} |g(x-z) - g(x)| \int |\varphi(y)|dy \\ &\quad + 2\sup_{x\in\mathbb{R}} |g(x)| \int_{|y|>\delta/h_n} |\varphi(y)|dy.\end{aligned}$$

Sendo g integrável e φ tal que $\lim_{|x|\to+\infty} x\varphi(x) = 0$, a convergência expressa em a) é consequência da desigualdade seguinte, que é válida para todo o $\delta > 0$ e $h_n > 0$:

$$\begin{aligned}\left|\varphi_{h_n} * g(x) - g(x)\int \varphi(y)dy\right| \\ \leq \sup_{|z|\leq\delta} |g(x-z) - g(x)| \int |\varphi(y)|dy \\ + \delta^{-1} \sup_{|z|>\delta/h_n} |z\varphi(z)| \int |g(y)|dy + |g(x)| \int_{|y|>\delta/h_n} |\varphi(y)|dy.\end{aligned}$$

□

Observação 3.2.2. Se g é uma função integrável e uniformemente contínua, g é necessariamente limitada, valendo a convergência $\varphi_{h_n} * g(x) \to g(x) \int \varphi(y)dy$, uniformemente em x.

Doravante designaremos por núcleo de Parzen–Rosenblatt qualquer núcleo limitado K que satisfaça a condição adicional

$$\lim_{|x|\to+\infty} xK(x) = 0.$$

Um núcleo limitado de suporte limitado é claramente de Parzen–Rosenblatt. Todos os núcleos da Tabela 3.1.2 são de Parzen–Rosenblatt.

3.2.1 Viés

Sendo K um núcleo limitado, o estimador do núcleo tem por média

$$\mathrm{E}f_n(x) = \int K_{h_n}(x-y)f(y)dy = K_{h_n} * f(x), \tag{3.2.3}$$

Como consequência imediata do lemma de Bochner concluímos que o estimador do núcleo é assintoticamente cêntrico.

Proposição 3.2.4. *Sejam K um núcleo de Parzen–Rosenblatt e f contínua em $x \in \mathbb{R}$. Se $h_n \to 0$ então*

$$\mathrm{E}f_n(x) \to f(x).$$

No caso do estimador do histograma $\hat{f}_n$, vimos que se $f \in \mathscr{L}_x(\alpha)$ então

$$\text{Viés}\hat{f}_n(x) = O(h_n^\alpha),$$

(cf. (2.2.3)) e que a ordem de convergência que se obtém para $\alpha = 1$ não pode ser melhorada mesmo que a densidade subjacente verifique condições adicionais de regularidade. Vejamos agora o que se passa com o estimador do núcleo. Começaremos por assumir que K tem suporte limitado. Atendendo a que

$$\begin{aligned} \text{Viés}f_n(x) &= K_{h_n} * f(x) - f(x) \\ &= \int K(z)(f(x - zh_n) - f(x))dz, \end{aligned} \tag{3.2.5}$$

tal como para o histograma concluímos que

$$\text{Viés}f_n(x) = O(h_n^\alpha),$$

sempre que $f \in \mathscr{L}_x(\alpha)$. No entanto, se assumirmos que a densidade f verifica $f' \in \mathscr{L}_x(\alpha)$, usando a fórmula de Taylor com resto integral obtemos o desenvolvimento

$$\begin{aligned} \text{Viés}f_n(x) &= -h_n \int yK(y) \int_0^1 f'(x - th_n y)\, dtdy \\ &= -h_n f'(x) \int yK(y)\, dy + O(h_n^{\alpha+1}), \end{aligned}$$

pois

$$\left| \int yK(y) \int_0^1 (f'(x - th_n y) - f'(x))\, dtdy \right| \leq h_n^\alpha \frac{M}{\alpha + 1} \int |y|^{\alpha+1} |K(y)|\, dy,$$

onde M é a constante de Lipschitz de f'.

Significa isto que se K for tal que $m_1(K) = \int yK(y)\, dy = 0$, o que acontece em particular quando K é simétrico relativamente à origem, o viés do estimador do núcleo é de ordem inferior ao do histograma.

Teorema 3.2.6. *Sejam K um núcleo limitado com $m_1(K) = 0$ e $\int z^2|K(z)|dz < \infty$, f com derivada de segunda ordem f'' limitada e contínua em $\mathbb{R}$ e $x \in \mathbb{R}$. Se $h_n \to 0$ então*

$$\mathrm{E}f_n(x) - f(x) = \frac{h_n^2}{2} m_2(K) f''(x) + o(h_n^2).$$

Dem: Atendendo a (3.2.5) e à condição $m_1(K) = 0$, da fórmula de Taylor de segunda ordem com resto integral e do teorema da convergência dominada obtemos

$$\begin{aligned} \mathrm{E}f_n(x) - f(x) &= h_n^2 \int z^2 K(z) \int_0^1 (1-t) f''(x - tzh_n) dt dz \qquad (3.2.7) \\ &= h_n^2 \left(\frac{1}{2} m_2(K) f''(x) + o(1) \right). \end{aligned}$$

□

Observação 3.2.8. Se o núcleo K é de suporte limitado, o resultado anterior mantém-se válido assumindo apenas que f possui derivada de segunda ordem contínua numa vizinhança de x.

3.2.2 Variância

Analisemos agora a variância do estimador do núcleo.

Proposição 3.2.9. *Sejam K um núcleo de Parzen–Rosenblatt e f contínua em $x \in \mathbb{R}$. Se $h_n \to 0$ então*

$$nh_n \mathrm{Var} f_n(x) \to R(K) f(x).$$

Dem: Para $x \in \mathbb{R}$ temos

$$\begin{aligned} \mathrm{Var} f_n(x) &= \frac{1}{n} \mathrm{Var}(K_{h_n}(x - X_1)) \\ &= \frac{1}{n} \left(\mathrm{E}K_{h_n}(x - X_1)^2 - (\mathrm{E}K_{h_n}(x - X_1))^2 \right) \\ &= \frac{1}{nh_n} \left(\int K(z)^2 f(x - zh_n) dz - h_n \left(\mathrm{E}f_n(x) \right)^2 \right). \qquad (3.2.10) \end{aligned}$$

Para concluir basta ter em conta que $\mathrm{E}f_n(x) = f(x) + o(1)$ e usar o lema de Bochner com $\varphi = K^2$.

□

Em particular, se $nh_n \to +\infty$ concluímos que

$$\mathrm{Var} f_n(x) \to 0,$$

sendo a ordem de convergência para zero da variância dada por

$$\mathrm{Var} f_n(x) = O\big((nh_n)^{-1}\big).$$

Esta é precisamente a ordem de convergência que encontrámos para a variância pontual do estimador do histograma. Sob condições adicionais de regularidade sobre f podemos ainda obter o desenvolvimento seguinte.

Teorema 3.2.11. *Sejam K um núcleo limitado com $\int |z|K(z)^2 dz < \infty$ e f com derivada f' limitada em $\mathbb{R}$. Para $x \in \mathbb{R}$, se $h_n \to 0$ então*

$$\mathrm{Var} f_n(x) = \frac{1}{nh_n} R(K)\, f(x) + O(n^{-1}).$$

Dem: Atendendo a (3.2.10), basta ter em conta que

$$\int K(z)^2 f(x - zh_n) dz = R(K)\, f(x) + \int K(z)^2 (f(x - zh_n) - f(x)) dz,$$

onde

$$\left| \int K(z)^2 (f(x - zh_n) - f(x)) dz \right| \leq h_n ||f'||_\infty \int |z| K(z)^2 dz,$$

com $||\varphi||_\infty = \sup_{x \in \mathbb{R}} |\varphi(x)|$, e que, de acordo com o lema de Bochner, $\mathrm{E}f_n(x) \to f(x)$, uma vez que f é limitada e contínua em $\mathbb{R}$.

□

Observação 3.2.12. Se o núcleo K é de suporte limitado, o resultado anterior mantém-se válido para f com derivada limitada numa vizinhança de x.

3.2.3 Erro quadrático médio

De forma análoga ao que fizemos para o histograma, analisamos agora o erro quadrático médio, $\mathrm{EQM}(f_n(x))$, do estimador do núcleo no ponto x. Tendo em conta a decomposição de $\mathrm{EQM}(f_n(x))$ como soma dum termo de variância e dum termo de quadrado do viés, os resultados seguintes são consequência imediata dos resultados dos parágrafos anteriores.

Teorema 3.2.13. *Sejam K um núcleo de Parzen-Rosenblatt e f contínua em $x \in \mathbb{R}$. Se $h_n \to 0$ e $nh_n \to +\infty$ então*

$$\mathrm{EQM}(f_n(x)) \to 0.$$

Teorema 3.2.14. *Sejam K um núcleo limitado com $\int zK(z)dz = 0$ e $\int z^2|K(z)|dz < \infty$, f com derivadas f' e f'' limitadas e contínuas em $\mathbb{R}$ e $x \in \mathbb{R}$. Se $h_n \to 0$ temos*

$$\mathrm{EQM}(f_n(x)) = \frac{1}{nh_n} R(K)\, f(x) + \frac{h_n^4}{4} m_2^2(K)\, f''(x)^2 + O(n^{-1}) + o(h_n^4).$$

Esta decomposição põe mais uma vez em evidência o papel determinante que h_n tem no comportamento assintótico do estimador. A janela deverá convergir para zero para que o viés do estimador seja pequeno mas não o poderá fazer violentamente sob pena de obtermos um estimador com grande variabilidade.

A ordem de convergência para zero do erro quadrático médio no ponto x depende, como vimos, da regularidade de f. Assumindo que K é de suporte limitado e que $f^{(k)} \in \mathscr{L}_x(\alpha)$, para $\alpha \in]0,1]$ e $k \in \{0,1\}$, então

$$\mathrm{EQM}(f_n(x)) = O\left((nh_n)^{-1} + h_n^{2k+2\alpha}\right).$$

A maior ordem de convergência ocorre quando $h_n = c\, n^{-1/(1+2k+2\alpha)}$, com $c > 0$, obtendo-se neste caso

$$f_n(x) - f(x) = O_p\left(n^{-(k+\alpha)/(1+2k+2\alpha)}\right).$$

Nas condições do teorema anterior temos $k+\alpha=2$ obtendo-se a ordem de convergência

$$f_n(x)-f(x)=O_p\big(n^{-2/5}\big),$$

quando

$$h_n=c\,n^{-1/5},\ \text{com } c>0.$$

3.2.4 Convergência quase certa

Os resultados seguintes sobre a convergência pontual quase certa do estimador do núcleo são em tudo análogos aos obtidas para o estimador do histograma. As diferenças residem nas ordens de convergência exibidas devido à diferente velocidade de convergência para zero do viés do estimador. A convergência pontual quase certa do estimador do núcleo é estudada por Van Ryzin (1969), Deheuvels (1974) e Devroye e Penrod (1984).

Teorema 3.2.15. *Sejam K um núcleo de Parzen-Rosenblatt e f contínua em $x\in\mathbb{R}$. Se $h_n\to 0$ e se*

$$\sum_{n=1}^{\infty}\exp(-\gamma n h_n)<\infty,\ \textit{para todo o } \gamma>0,$$

então

$$f_n(x)\xrightarrow{qc} f(x).$$

Dem: Atendendo à Proposição 3.2.4 basta mostrar que $f_n(x)-\mathrm{E}f_n(x)\xrightarrow{qc}0$. Comecemos por notar que

$$f_n(x)-\mathrm{E}f_n(x)=\frac{1}{n}\sum_{i=1}^{n}Z_i,$$

com

$$Z_i=K_{h_n}(x-X_i)-\mathrm{E}K_{h_n}(x-X_i),$$

onde $|Z_i|\le 2||K||_\infty/h_n$ e para n suficientemente grande, $\mathrm{E}(Z_i)^2\le C/h_n$, com $C>0$. Para concluir, basta notar que pelo Lema 2.2.11 se tem

$$\mathrm{P}(|f_n(x)-\mathrm{E}f_n(x)|>\epsilon)\le 2\exp\left(-3nh_n\epsilon^2/(8C)\right),$$

para todo o $0<\epsilon<C/(2||K||_\infty)$.

□

Ordem de convergência

Da demonstração anterior, decorre imediatamente o seguinte resultado.

Teorema 3.2.16. *Sejam K um núcleo limitado de suporte limitado e $f^{(k)} \in \mathscr{L}_x(\alpha)$ para algum $k \in \{0,1\}$ e $x \in \mathbb{R}$. Se $h_n \to 0$ e $nh_n/\log n \to +\infty$ então*

$$f_n(x) - f(x) = O\left(\left(\frac{\log n}{nh_n} \right)^{1/2} + h_n^{k+\alpha} \right).$$

A maior ordem de convergência quase certa que podemos obter a partir do resultado anterior passa por tomar $h_n = c\,(\log n/n)^{1/(1+2k+2\alpha)}$, com $c > 0$, obtendo-se neste caso

$$f_n(x) - f(x) = O\left(\left(\frac{\log n}{n} \right)^{(k+\alpha)/(1+2k+2\alpha)} \right).$$

No caso em que $k + \alpha = 2$, a escolha da janela $h_n = c\,(\log n/n)^{1/5}$, com $c > 0$, conduz a uma ordem de convergência quase certa ligeiramente inferior à ordem de convergência em probabilidade estabelecida atrás:

$$f_n(x) - f(x) = O\left(\left(\frac{\log n}{n} \right)^{2/5} \right).$$

3.3 Convergência L_∞

Atendendo à segunda parte do Lema 3.2.1, para todo o núcleo limitado K e toda a densidade $f \in \mathscr{U}$ sabemos que $||\mathrm{E}f_n - f||_\infty \to 0$ sempre que $h_n \to 0$. O estudo da convergência uniforme do estimador do núcleo, segundo um determinado modo de convergência estocástica, reduz-se assim ao estudo do termo aleatório $||f_n - \mathrm{E}f_n||_\infty$.

A hipótese de continuidade uniforme feita sobre f tem aqui um papel essencial. Para K e h_n nas condições do resultado seguinte é possível mostrar que a condição $f \in \mathscr{U}$ é uma condição necessária para a convergência uniforme quase certa do estimador do núcleo (cf. Schuster, 1969, 1970).

No caso do estimador da janela móvel definido por (1.3.1), a abordagem usada aquando do estimador do histograma pode ser directamente utilizada, decorrendo a convergência uniforme quase certa da desigualdade

$$||f_n - \mathrm{E}f_n||_\infty \leq \frac{1}{h_n} ||F_n - F||_\infty.$$

Para núcleos de variação limitada em $\mathbb{R}$, uma generalização da desigualdade anterior foi utilizada por Nadaraya (1965) para estabelecer a convergência uniforme quase certa do estimador do núcleo sob condições sobre h_n ligeiramente diferentes das que a seguir apresentamos. Recordemos que uma função g diz-se de variação limitada em $\mathbb{R}$ se existe $M > 0$ tal que

$$\sum_{i=1}^{n} |g(t_{i+1}) - g(t_i)| \leq M,$$

para todo o conjunto finito de números reais com $t_1 \leq t_2 \leq \cdots \leq t_{n+1}$. Ao supremo de todas as somas da forma anterior chamamos variação de g e representa-mo-lo por $V(g)$ (sobre funções de variação limitada ver Cohn, 1980, pp. 143–149, e Bogachev, 2007, pp. 332–336). Os núcleos apresentados na Tabela 3.1.2 são claramente de variação limitada em $\mathbb{R}$. Uma caracterização das funções de variação limitada é dada pelo facto delas poderem ser escritas como diferença de duas funções não-decrescentes. O motivo para a consideração de núcleos de variação limitada será claro da demonstração do resultado seguinte que estabelece condições suficientes para a convergência uniforme quase certa do estimador do núcleo.

Teorema 3.3.1. *Sejam K um núcleo de variação limitada e $f \in \mathscr{U}$. Se $h_n \to 0$ e $nh_n^2/\log\log n \to +\infty$ então*

$$||f_n - f||_\infty \xrightarrow{qc} 0.$$

Dem: Para $x \in \mathbb{R}$ temos

$$f_n(x) = \frac{1}{h_n} \int K\left(\frac{x-y}{h_n}\right) dF_n(y) = \frac{1}{h_n} \int \psi_{x,h_n}(y) dF_n(y)$$

onde $\psi_{x,h_n}(y) = K\big(\frac{x-y}{h_n}\big)$. Sendo K de variação limitada, o núcleo K é limitado e existem os limites $\lim_{u\to\pm\infty} K(u)$, que são necessariamente nulos

uma vez que K é integrável. Assim, atendendo a que ψ_{x,h_n} é de variação limitada com $\lim_{u\to\pm\infty}\psi_{x,h_n}(u)=0$, da fórmula de integração por partes para o integral de Legesgue–Stieltjes (ver por exemplo Cohn, 1980, p. 163), concluímos que com probabilidade 1 se tem

$$f_n(x) = -\frac{1}{h_n}\int F_n(y)d\Psi_{x,h_n}(y),$$

onde Ψ_{x,h_n} é a medida de Legesgue-Stiltjes associada a ψ_{x,h_n}. De forma análoga obtemos

$$\mathrm{E}f_n(x) = \frac{1}{h_n}\int \psi_{x,h_n}(y)dF(y) = -\frac{1}{h_n}\int F(y)d\Psi_{x,h_n}(y),$$

e portanto

$$\begin{aligned}
||f_n - \mathrm{E}f_n||_\infty &= \frac{1}{h_n}\sup_{x\in\mathbb{R}}\left|\int (F_n(y)-F(y))d\Psi_{x,h_n}(y)\right| \\
&\leq \frac{1}{h_n}\sup_{x\in\mathbb{R}}|\Psi_{x,h_n}|(\mathbb{R})||F_n-F||_\infty \\
&\leq \frac{1}{h_n}\sup_{x\in\mathbb{R}}V(\psi_{x,h_n})||F_n-F||_\infty \\
&\leq \frac{1}{h_n}V(K)||F_n-F||_\infty.
\end{aligned}$$

O resultado é agora uma consequência imediata da lei do logaritmo iterado (2.3.4).

□

Tal como para o estimador do histograma, uma análise mais fina permite obter a convergência uniforme quase certa do estimador do núcleo sob condições mais fracas sobre a janela. Neste sentido apontam os resultados de Révész (1972), Deheuvels (1974), Silverman (1978) e Bertrand-Retali (1978). No que se segue, limitar-nos-emos a enunciar o resultado de Bertrand-Retali (1978) que estabelece que na presença da condição $h_n \to 0$ a condição $nh_n/\log n \to +\infty$, além de suficiente, é também necessária para a convergência uniforme quase certa do estimador do núcleo para toda a densidade uniformemente contínua. A sua demonstração pode ser encontrada em Bosq e Lecoutre (1987, pp. 65–71).

Diremos que um núcleo limitado K é de Bertrand-Retali–Geffroy se o conjunto do seus pontos descontinuidade tem medida de Lebesgue nula e se a aplicação $x \to \sup\{|K(t)| : ||t-x|| \leq 1\}$ é integrável. Um núcleo limitado de suporte limitado e contínuo em quase todo o ponto, é de Bertrand-Retali–Geffroy. O mesmo acontece com um núcleo de variação limitada.

Teorema 3.3.2. *Se K é um núcleo de Bertrand-Retali–Geffroy e se $h_n \to 0$ então as seguintes proposições são equivalentes:*

i) $nh_n/\log n \to +\infty$;

ii) $\forall f \in \mathscr{U}\ ||f_n - f||_\infty \xrightarrow{p} 0$;

iii) $\forall f \in \mathscr{U}\ ||f_n - f||_\infty \xrightarrow{qc} 0$.

3.4 Convergência L_1

Tal como para o estimador do histograma, é também Abou-Jaoudé (1977) que inicia o estudo da convergência L_1 do estimador do núcleo mas o resultado de convergência que obtém surge apenas publicado na sua tese de doutoramento, o que possivelmente levou a que o mesmo não tenha sido referido nos trabalhos subsequentes de Devroye e Wagner (1979) e Devroye (1983). É o resultado de Devroye (1983) que enunciamos a seguir.

Teorema 3.4.1. *Se $K \in \mathscr{F}$ então as seguintes proposições são equivalentes:*

i) $h_n \to 0$, $nh_n \to +\infty$;

ii) $\forall f \in \mathscr{F} \quad ||f_n - f||_1 \xrightarrow{p} 0$;

iii) $\forall f \in \mathscr{F} \quad ||f_n - f||_1 \xrightarrow{qc} 0$;

iv) $\exists f \in \mathscr{F} \quad ||f_n - f||_1 \xrightarrow{p} 0$.

Além disso, para um qualquer núcleo K, i) implica iii).

Reparemos que sendo o núcleo uma densidade de probabilidade, ou o estimador do núcleo converge para todo o $f \in \mathscr{F}$, ou então não converge

para nenhum $f \in \mathscr{F}$ (no sentido L_1 e segundo os modos estocásticos de convergência considerados). Não existem portanto casos intermédios, sendo esta a principal novidade do resultado de Devroye (1983) relativamente ao de Abou-Jaoudé (1977). Veja-se Devroye e Györfi (1985, pp. 12–19) para a demonstração do resultado anterior.

3.5 Convergência em média quadrática integrada

Neste parágrafo estabelecemos a convergência em média quadrática integrada do estimador do núcleo. Começamos por obter um desenvolvimento exacto para o erro quadrático médio integrado definido por (1.2.2).

Teorema 3.5.1. *Sejam K e f de quadrado integrável. Para $h_n > 0$ temos*

$$\mathrm{IVAR}(f_n) = \frac{1}{nh_n} R(K) - \frac{1}{n} \int \left(\bar{K}_{h_n} * K_{h_n}\right) * f(x) f(x)\, dx,$$

$$\mathrm{IVIES}(f_n) = \int \left(\bar{K}_{h_n} * K_{h_n} - 2K_{h_n}\right) * f(x) f(x)\, dx + R(f)$$

e

$$\mathrm{EQMI}(f_n) = \frac{1}{nh_n} R(K) + \int (\tfrac{n-1}{n} \bar{K}_{h_n} * K_{h_n} - 2K_{h_n}) * f(x) f(x) dx + R(f),$$

onde $\bar{K}(u) = K(-u)$, para $u \in \mathbb{R}$.

Dem: De (3.2.10) e (3.2.3) temos, respectivamente,

$$\mathrm{IVAR}(f_n) = \frac{1}{nh_n} R(K) - \frac{1}{n} \int K_{h_n} * f(x)^2 dx \tag{3.5.2}$$

e

$$\begin{aligned} \mathrm{IVIES}(f_n) &= \int \{K_{h_n} * f(x) - f(x)\}^2 dx \\ &= \int K_{h_n} * f(x)^2 dx - 2 \int K_{h_n} * f(x) f(x) dx + R(f). \end{aligned}$$

Para concluir basta ter em conta (1.2.3) e a igualdade

$$\begin{aligned}\int K_{h_n} * f(x)^2 dx &= \iiint \bar{K}_{h_n}(z-y-u)K_{h_n}(u)f(y)f(z)\,dudydz \\ &= \iint \bar{K}_{h_n} * K_{h_n}(z-y)f(y)f(z)\,dydz \\ &= \int \bar{K}_{h_n} * K_{h_n} * f(z)f(z)\,dz.\end{aligned}$$

□

Estamos agora em condições de mostrar que, tal como para o estimador do histograma, as condições $h_n \to 0$ e $nh_n \to +\infty$ são suficiente para a convergência média quadrática integrada do estimador do núcleo.

Teorema 3.5.3. *Sejam K e f de quadrado integrável. Se $h_n \to 0$ e $nh_n \to +\infty$ então f_n converge para f em média quadrática integrada.*

Dem: Atendendo ao teorema anterior basta mostrar que

$$\int \varphi_{h_n} * f(x)f(x)\,dx \to R(f)\int \varphi(z)\,dz, \tag{3.5.4}$$

para toda a função integrável φ. Tal é de facto verdade uma vez que

$$\begin{aligned}\int \varphi_{h_n} * f(x)f(x)\,dx &= \iint \varphi_{h_n}(x-y)f(y)f(x)\,dxdy \\ &= \iint \varphi(z)\bar{f}(zh_n - x)f(x)\,dxdz \\ &= \int \varphi(z)\bar{f} * f(zh_n)\,dz,\end{aligned}$$

com $\bar{f}(x) = f(-x)$, bastando agora usar o teorema da convergência dominada de Lebesgue e o facto de $\bar{f} * f$ ser limitada e contínua na origem visto ser o produto de convolução de funções de quadrado integrável.

□

3.6 Escolha assintoticamente óptima de h_n

Começamos este parágrafo apresentando um desenvolvimento assintótico para o erro quadrático médio integrado que terá um papel relevante na

discussão da escolha óptima da janela e do núcleo que faremos a seguir. Resultados do mesmo género foram obtidos por Rosenblatt (1956, 1971), Epanechnikov (1969) e Nadaraya (1974).

Teorema 3.6.1. *Sejam K um núcleo de quadrado integrável com $m_1(K) = 0$ e $\int z^2|K(z)|dz < \infty$, e f com derivada de segunda ordem contínua e f e f'' de quadrado integrável. Se $h_n \to 0$ então*

$$\mathrm{IVAR}(f_n) = \frac{1}{nh_n} R(K) + O(n^{-1}),$$

$$\mathrm{IVIES}(f_n) = \frac{h_n^4}{4} m_2^2(K)\, R(f'') + o\left(h_n^4\right)$$

e

$$\mathrm{EQMI}(f_n) = \frac{1}{nh_n} R(K) + \frac{h_n^4}{4} m_2^2(K)\, R(f'') + O(n^{-1}) + o\left(h_n^4\right).$$

Dem: O desenvolvimento apresentado para $\mathrm{IVAR}(f_n)$ é consequência imediata do Teorema 3.5.1 e da convergência (3.5.4) com $\varphi = \bar{K} * K$. Para obter o desenvolvimento assintótico de $\mathrm{IVIES}(f_n)$, vamos usar (3.2.7) e o facto de f possuir derivada de segunda ordem contínua de quadrado integrável. Nestas condições, pelo teorema da convergência dominada temos

$$\begin{aligned}
\mathrm{IVIES}(f_n) &= h_n^4 \int\int\int\int_0^1\int_0^1 u^2v^2K(u)K(v) \\
&\qquad \times (1-s)(1-t)f''(x-suh_n)f''(x-tvh_n)dsdtdudvdx \\
&= h_n^4 \int\int\int\int_0^1\int_0^1 u^2v^2K(u)K(v) \\
&\qquad \times (1-s)(1-t)\bar{f}'' * f''((su-tv)h_n)dsdtdudv \\
&= \frac{h_n^4}{4} m_2^2(K) R(f'') + o(h_n^4).
\end{aligned}$$

O desenvolvimento $\mathrm{EQMI}(f_n)$ decorre agora da decomposição (1.2.3). □

Nas condições do teorema anterior, a maior ordem de convergência para zero do erro quadrático médio integrado é

$$\mathrm{EQMI}(f_n) = O(n^{-4/5})$$

e é obtida quando tomamos

$$h_n = c\, n^{-1/5}, \text{ com } c > 0.$$

Janela assintoticamente óptima

Definindo o erro quadrático médio integrado assintótico como a soma das duas parcelas mais significativas do desenvolvimento do EQMI dado no Teorema 3.6.1,

$$\text{EQMIA}(K, h_n) = \frac{1}{nh_n}\, R(K) + \frac{h_n^4}{4}\, m_2^2(K)\, R(f''),$$

a janela óptima no sentido da minimização do erro quadrático médio integrado assintótico é obtida no resultado seguinte.

Teorema 3.6.2. *Se f é tal que $R(f'') > 0$, o valor de h_n que minimiza o* $\text{EQMIA}(K, h_n)$ *é dado por*

$$h_{\text{EQMIA}} = c_K R(f'')^{-1/5}\, n^{-1/5},$$

onde

$$c_K = \big(\, R(K)/m_2^2(K)\big)^{1/5}.$$

Além disso, nas condições do Teorema 3.6.1 temos

$$\text{EQMI}(f_n^\circ) = \frac{5}{4}\, R(K)^{4/5} m_2(K)^{2/5}\, R(f'')^{1/5}\, n^{-4/5} + o(n^{-4/5}),$$

onde f_n° representa o estimador do núcleo com janela h_{EQMIA}.

Interpretando a quantidade $R(f'')$ como uma medida da variabilidade global de f, reparemos que quanto maior é $R(f'')$, mais difícil é estimar f através dum estimador do núcleo.

Eficiência relativamente ao polígono de frequências

Tendo em conta (2.9.3) e definindo o erro quadrático médio integrado assintótico associado ao polígono de frequências com janela assintoticamente óptima por

$K(u)$	$(\Psi(K))^{5/4}$	$K(u)$	$(\Psi(K))^{5/4}$
Uniforme	1.180	Triangular	1.251
Epanechnikov	1.269	*Biweight*	1.261
Triweight	1.252	Normal	1.207
Laplace	0.963	Cauchy de ordem 2	0.856

Tabela 3.6.3: *Eficiência do estimador com núcleo K relativamente ao polígono de frequências.*

$$\mathrm{EQMIA}(h_{\mathrm{EQMIA}}) = \frac{5}{12}\left(49R(f'')/15\right)^{1/5}\, n^{-4/5},$$

do resultado anterior concluímos ser válida a igualdade

$$\mathrm{EQMIA}(h_{\mathrm{EQMIA}}) = \Psi(K)\,\mathrm{EQMIA}(K, h_{\mathrm{EQMIA}}),$$

onde

$$\Psi(K) = \frac{1}{3}\left(\frac{49}{15R(K)^4\, m_2^2(K)}\right)^{1/5}.$$

Na Tabela 3.6.3 apresentamos os valores de $\Psi(K)^{5/4}$ para os núcleos que temos vindo a considerar. Este valor traduz a eficiência do estimador com núcleo K e escolha assintoticamente óptima da janela, relativamente ao polígono de frequências com janela assintoticamente óptima. Com efeito, $\Psi(K)^{5/4}$ não é mais do que o quociente n/n_K, onde n_K é o tamanho da amostra que é necessário usar no estimador com núcleo K para obter um EQMIA igual ao do polígono de frequências baseado numa amostra de tamanho n.

A eficiência obtida para os vários núcleos revela que nem sempre o estimador do núcleo é mais eficiente que o polígono de frequências. No entanto, para os seis primeiros núcleos da tabela o estimador do núcleo é claramente mais eficiente que o polígono de frequências. Reparemos também que para estes núcleos não há grandes diferenças entre os diversos estimadores do

núcleo. No entanto, o estimador com núcleo de Epanechnikov apresenta maior eficiência entre os estimadores considerados.

EQMI *vs.* EQMIA

Tomando para núcleo do estimador o núcleo normal, as quantidades IVAR, IVIES e EQMIA, são facilmente calculáveis quando f é uma mistura de densidades normais. Com efeito, se f toma a forma

$$f(x) = \sum_{\ell=1}^{k} w_\ell \phi_{\sigma_\ell}(x - \mu_\ell),$$

onde $\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$, $\phi_\sigma(x) = \sigma^{-1}\phi(x/\sigma)$, $w_1 \ldots, w_k$ são números reais positivos de soma unitária, $\mu_\ell \in \mathbb{R}$ e $\sigma_\ell > 0$, podemos escrever

$$\mathrm{IVAR}(f_n) = \frac{1}{2\sqrt{\pi}} \frac{1}{nh_n} - \frac{1}{n}\,\mathrm{U}(h_n, 0, 2),$$

$$\mathrm{IVIES}(f_n) = \mathrm{U}(0, 0, 0) - 2\mathrm{U}(h_n, 0, 1) + \mathrm{U}(h_n, 0, 2),$$

e

$$\mathrm{EQMIA}(f_n) = \frac{1}{2\sqrt{\pi}} \frac{1}{nh_n} + \frac{h_n^4}{4}\,\mathrm{U}(0, 2, 0),$$

onde

$$\mathrm{U}(h, s, q) = \sum_{\ell,\ell'=1}^{k} w_\ell w_{\ell'} \phi^{(2s)}_{\sigma_{\ell\ell' q}}(\mu_\ell - \mu_{\ell'}),$$

com $\sigma_{\ell\ell' q} = (\sigma_\ell^2 + \sigma_{\ell'}^2 + qh^2)^{1/2}$ e $\phi_\sigma^{(s)}(x) = \sigma^{-(s+1)}\phi^{(s)}(x/\sigma)$ (cf. Marron e Wand, 1992).

As expressões anteriores permitem-nos ilustrar o efeito da escolha da janela na variância e no viés do estimador do núcleo, bem como as limitações da aproximação EQMIA de EQMI. O carácter local e assintótico de tal aproximação é bem retratado na Figura 3.6.4 para a distribuição

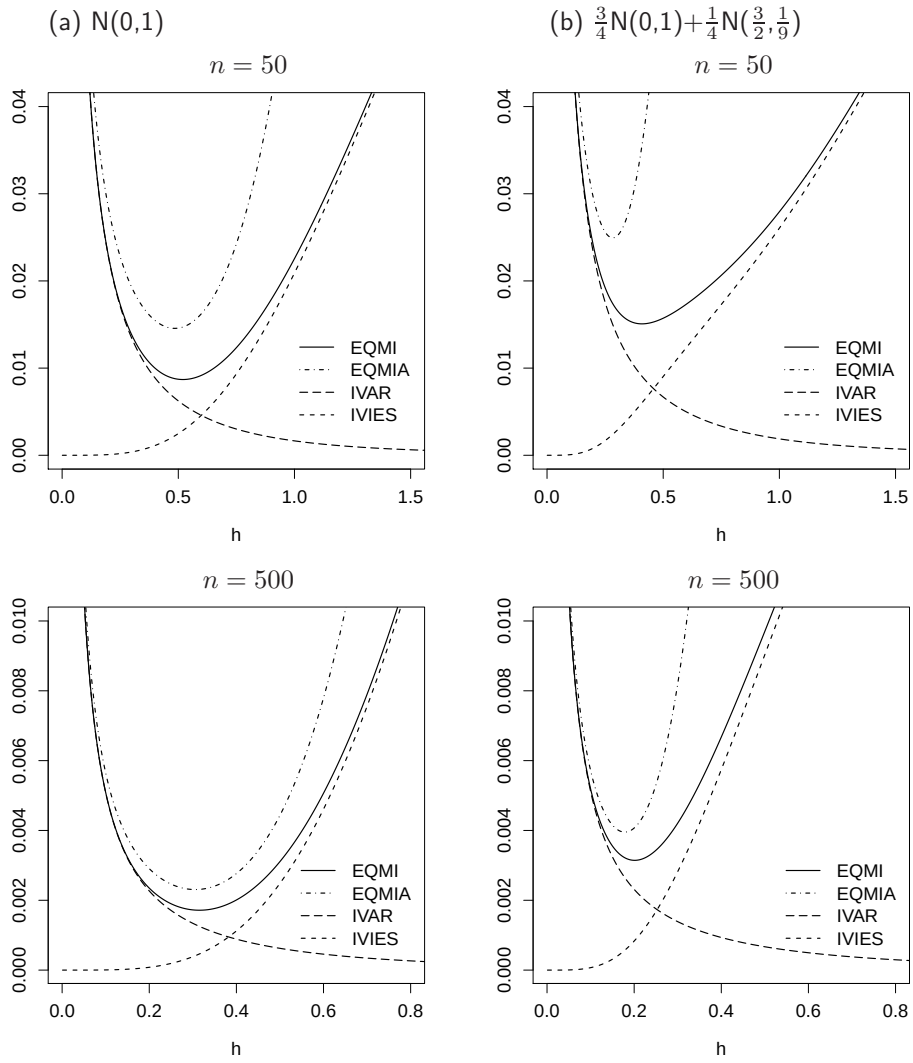

Figura 3.6.4: EQMI, EQMIA, IVAR *e* IVIES *em função de* $h_n = h$ *para as distribuições* $N(0,1)$ *e* $\frac{3}{4}N(0,1) + \frac{1}{4}N(\frac{3}{2}, \frac{1}{9})$ *com* $n = 50$ *e* $n = 500$.

normal standard e para a mistura de normais com densidade

$$f(x) = \frac{3}{4}\,\phi(x) + \frac{1}{4}\,\phi_{1/3}(x - 3/2). \tag{3.6.5}$$

Reparemos que as discrepâncias entre as janelas óptima (no sentido do

EQMI) e assintoticamente óptima, que são atenuadas com o aumento do tamanho da amostra, podem revelar-se significativas para amostras de pequena dimensão.

3.7 A escolha do núcleo

Quando no estimador com núcleo K usamos a janela assintoticamente óptima

$$h_{\mathrm{EQMIA}}(K) = \delta(K) R(f'')^{-1/5}\, n^{-1/5},$$

onde $\delta(K) = \big(R(K)/m_2^2(K)\big)^{1/5}$, obtemos para erro quadrático médio integrado assintótico deste 'estimador de oráculo' a quantidade

$$\mathrm{EQMIA}(K, h_{\mathrm{EQMIA}}) = \frac{5}{4}\, R(K)^{4/5} m_2(K)^{2/5}\, R(f'')^{1/5}\, n^{-4/5}, \qquad (3.7.1)$$

que depende de K através da funcional

$$\Phi(K) = R(K)^{4/5} m_2(K)^{2/5}.$$

Esta funcional é invariante para mudanças de escala. Com efeito, para $\delta > 0$,

$$m_2(K_\delta) = \frac{1}{\delta} \int u^2 K\Big(\frac{u}{\delta}\Big) du = \delta^2 m_2(K)$$

e

$$R(K_\delta) = \frac{1}{\delta^2} \int K^2\Big(\frac{u}{\delta}\Big) du = \frac{1}{\delta} R(K),$$

e portanto

$$\Phi(K_\delta) = \Phi(K).$$

Denotando por $\mathscr{N}$ a classe dos núcleos limitados, simétricos e não negativos K com $\int u^2 K(u) du < \infty$, a propriedade de invariância anterior leva-nos a colocar o problema da determinação do núcleo óptimo, no sentido da minimização da funcional Φ, nos seguintes termos: pretendemos minimizar

$$\int K(u)^2 du$$

sujeito às condições

$$K \in \mathscr{N} \text{ e } m_2(K) = a^2,$$

onde $a \neq 0$ é fixado à partida.

Num contexto distinto do da estimação da densidade, Hodges e Lehmann (1956) mostraram que

$$\Phi(E) \leq \Phi(K),$$

para todo o $K \in \mathscr{N}$, onde

$$E(x) = \frac{3}{4}\,(1 - x^2)\,I(|x| \leq 1).$$

Sendo a funcional Φ invariante para mudanças de escala, todo o núcleo que se obtém de E por mudança de escala é também óptimo. O núcleo anterior não é mais do que o núcleo de Epanechnikov incluído na Tabela 3.1.2, sendo assim denominado devido às suas propriedades de optimalidade no contexto da estimação da densidade terem sido descritas por Epanechnikov (1969). Refira-se no entanto que a optimalidade deste núcleo quadrático havia sido já mencionada em Bartlett (1963).

Eficiência de outros núcleos

De acordo com (3.7.1) os erros quadráticos médios integrados associados aos estimadores com núcleos K e E, onde E é o núcleo de Epanechnikov, e janelas assintoticamente óptimas, estão relacionados pela igualdade

$$\mathrm{EQMIA}(E, h_{\mathrm{EQMIA}}(E)) = \frac{\Phi(E)}{\Phi(K)}\,\mathrm{EQMIA}(K, h_{\mathrm{EQMIA}}(K)). \qquad (3.7.2)$$

Na Tabela 3.7.3 apresentamos os valores de $(\Phi(E)/\Phi(K))^{5/4}$ para os núcleos que temos vindo a considerar. Este quociente traduz a eficiência do estimador com núcleo K relativamente ao estimador com núcleo óptimo E, uma vez, que de acordo com (3.7.1) ele não é mais do que o quociente n/n_K, onde n_K é o tamanho da amostra que é necessário usar no estimador com núcleo K para obter um EQMIA igual ao do estimador com núcleo de

$K(u)$	$(\Phi(E)/\Phi(K))^{5/4}$	$K(u)$	$(\Phi(E)/\Phi(K))^{5/4}$
Uniforme	0.930	Triangular	0.986
Epanechnikov	1.000	*Biweight*	0.994
Triweight	0.987	Normal	0.951
Laplace	0.759	Cauchy de ord. 2	0.674

Tabela 3.7.3: *Eficiência do estimador com núcleo K relativamente ao estimador com núcleo de Epanechnikov.*

Epanechnikov baseado numa amostra de tamanho n. A eficiência obtida para os vários núcleos revela que quando para janela do estimador tomamos a janela assintoticamente óptima, a escolha do núcleo não é uma questão essencial do ponto de vista do EQMIA. Núcleos distintos do óptimo podem apresentar eficiência quase óptima. Dos núcleos considerados apenas os núcleos de Laplace e de Cauchy de ordem 2 revelam uma pequena eficiência relativamente ao núcleo de Epanechnikov.

Núcleos equivalentes

Sendo a igualdade (3.7.2) válida para o estimador de oráculo com janela assintoticamente óptima, dela não se pode concluir, sem mais, que a escolha do núcleo não tenha influência no desempenho do estimador uma vez que uma tal janela é, na prática, desconhecida do utilizador. Sendo E o núcleo de Epanechnikov e h_n a janela seleccionada para este estimador, os erros quadráticos médios integrados assintóticos EQMIA(E, h_n) e EQMIA(K, h_n) não são, em geral, directamente comparáveis, mesmo que o núcleo K seja um núcleo com eficiência próxima da unidade. Este facto é ilustrado na Figura 3.7.4 (a).

Uma primeira forma de atenuar o efeito da escolha do núcleo é a de substituir o núcleo K pela mudança de escala K_{δ_1} de modo que os núcleos E e K_{δ_1} tenham a mesma variabilidade. A escala δ_1 é assim determinada

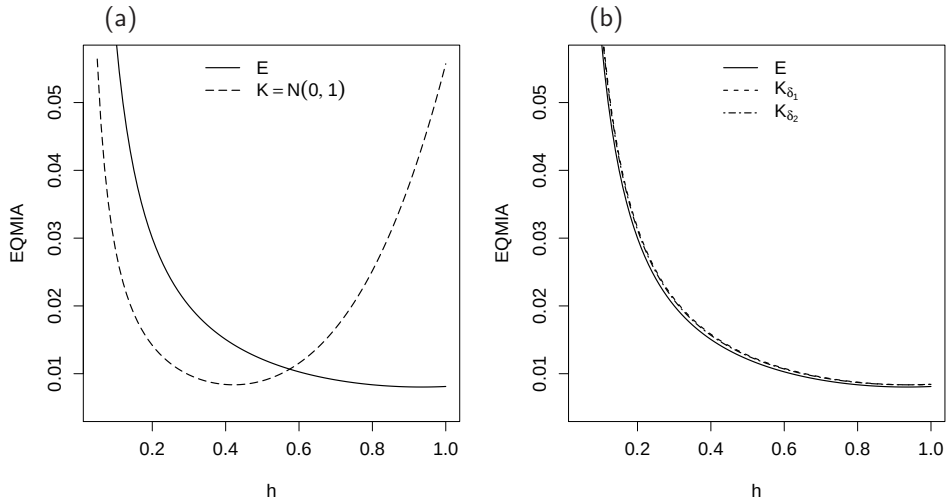

Figura 3.7.4: EQMIA *em função de h, para a distribuição* $N(0,1)$ *e* $n=100$.

pela condição

$$m_2(K_{\delta_1}) = m_2(E),$$

ou seja,

$$\delta_1 = \left(\frac{m_2(E)}{m_2(K)}\right)^{1/2}.$$

Se K for tal que o quociente

$$\Lambda_{K,E} = \Phi(E)/\Phi(K)$$

seja próximo da unidade, o que acontece com todos os núcleos mencionados na Tabela 3.7.3 com excepção dos dois últimos, concluímos que os erros quadráticos médios integrados assintóticos associados a ambos os estimadores são aproximadamente iguais:

$$\mathrm{EQMIA}(K_{\delta_1}, h_n) = \frac{1}{nh_n}\left(\frac{1}{\Lambda_{K,E}^{5/4}} - 1\right) + \mathrm{EQMIA}(E, h_n).$$

Uma outra maneira de abordar esta questão foi proposta por Marron e

$K(u)$	$\delta(K)$	$K(u)$	$\delta(K)$
Uniforme	1.351	Triangular	1.888
Epanechnikov	1.719	*Biweight*	2.036
Triweight	2.312	Normal	0.776

Tabela 3.7.5: *Escala* $\delta(K)$.

Nolan (1989), que sugerem a utilização da mudança de escala K_{δ_2} com

$$\delta_2 = \frac{\delta(K)}{\delta(E)}$$

onde

$$\delta(K) = \left(R(K)/m_2^2(K) \right)^{1/5}.$$

Na Tabela 3.7.5 indicam-se os valores das escalas $\delta(K)$ para os diferentes núcleos aí considerados. Com esta mudança de escala, obtemos:

$$\text{EQMIA}(K_{\delta_2}, h_n) = \frac{1}{\Lambda_{K,E}} \text{EQMIA}(E, h_n).$$

Com uma ou com outra das mudanças de escala anteriores, será de esperar que, se K for escolhido de modo que o quociente $\Lambda_{K,E}$ seja próximo da unidade, a eficiência do estimador com núcleo E e janela h_n seja próxima da do estimador com núcleo K_δ, com $\delta = \delta_1$ ou $\delta = \delta_2$, e janela h_n. Na Figura 3.7.4 ilustra-se o efeito destas duas mudanças de escala quando K é o núcleo normal standard e f é a densidade normal standard.

A partir da igualdade anterior decorre também que

$$\text{EQMIA}(K_{\delta(K)}, h_n) = \frac{1}{\Lambda_{K,E}} \text{EQMIA}(E_{\delta(E)}, h_n),$$

para todo o h_n. Podemos assim concluir que se em cada um dos estimadores usarmos as mudanças de escala $E_{\delta(E)}$ e $K_{\delta(K)}$ com $\Lambda_{K,E} \approx 1$, a utilização da mesma janela em ambos os estimadores dá origem a estimadores do núcleo

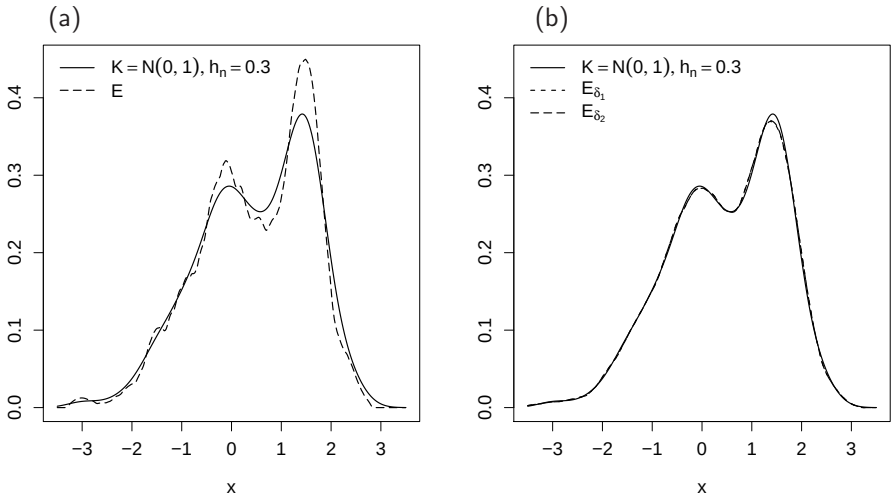

Figura 3.7.6: *Estimativas baseadas nos estimadores com núcleos normal standard e de Epanechnikov E, E_{δ_1} e E_{δ_2}, sendo a janela dada por $h_n = 0.3$.*

com semelhantes EQMIA. Os núcleos da forma $K_{\delta(K)}$ foram introduzidos por Marron e Nolan (1989) e são ditos núcleos canónicos.

As conclusões anteriores, que são naturalmente válidas para qualquer par de núcleos, são ilustradas a partir de uma amostra de tamanho 400 da mistura de normais (3.6.5). Vamos admitir que a janela $h_n = 0.3$ é escolhida para ser usada no estimador com núcleo normal standard e que a mesma é usada no estimador com núcleo de Epanechnikov. As estimativas obtidas por ambos os estimadores que apresentamos na Figura 3.7.6 (a) são perfeitamente distintas, ilustrando o impacto da escolha do núcleo. No entanto, se no estimador com núcleo de Epanechnikov usarmos a mudança de escala $\delta_2 = \delta(E)/\delta(K) = 1.719/0.776 = 2.215$ e a janela $h_n = 0.3$, o que é equivalente a usar no estimador com núcleo de Epanechnikov a janela $h_n = 2.215 \times 0.3 \approx 0.665$, constatamos da Figura 3.7.6 (b) que as estimativas produzidas por ambos os estimadores são agora praticamente idênticas. Resultados semelhantes são obtidos para a mudança de escala $\delta_1 = (m_2(K)/m_2(E))^{1/2} = 2.236$.

3.8 Núcleos de ordem superior. Redução de viés

Admitindo que f possui derivada de ordem $k \geq 2$ contínua e limitada em $\mathbb{R}$ e que o núcleo K é tal que $\int |u^k K(u)|du < \infty$, é fácil obter desenvolvimentos assintóticos para o viés do estimador do núcleo que generalizam os derivados em §3.2. Com efeito, usando a fórmula de Taylor podemos obter o desenvolvimento

$$\begin{aligned} \mathrm{E}f_n(x) - f(x) = &\sum_{\ell=1}^{k-1} (-1)^\ell \frac{h_n^\ell}{\ell!} m_\ell(K) f^{(\ell)}(x) \\ &+ h_n^k \iint_0^1 z^k K(z) \frac{(1-t)^{k-1}}{(k-1)!} f^{(k)}(x - tzh_n) dt dz, \end{aligned} \tag{3.8.1}$$

onde

$$m_\ell(K) = \int u^\ell K(u) du,$$

é o momento de ordem ℓ do núcleo K.

Quando tomamos para K um núcleo simétrico e não negativo, é necessariamente verdade que $m_1(K) = 0$ e $m_2(K) > 0$, o que leva a que a ordem de convergência para zero do viés do estimador seja precisamente h_n^2 (a menos de pontos x onde $f''(x) = 0$). Quando tomamos K satisfazendo $m_2(K) = 0$, o que é primeiramente sugerido por Bartlett (1963), a ordem h_n^2 pode ser melhorada. Maiores ordens de convergência podem ser obtidas se K for um núcleo de ordem k, isto é, se

$$m_0(K) = 1, m_j(K) = 0, \text{ para } j = 1, \ldots, k-1, \text{ e } m_k(K) \neq 0.$$

Habitualmente assumimos ainda que K é simétrico o que implica que a ordem k seja par. Uma densidade de probabilidade simétrica K com momento de segunda ordem finito é um núcleo de segunda ordem. Com excepção destes núcleos, os núcleos de ordem superior à segunda tomam necessariamente valores negativos, o que faz com que o estimador do núcleo nele baseado não seja mais uma densidade de probabilidade. Este é o preço a pagar por um estimador com viés de ordem superior a h_n^2 (sobre correcções para estimadores impróprios da densidade ver Glad et al., 2003).

Construção de núcleos de ordem k

Há vários métodos para construir núcleos de ordem superior à segunda (ver, por exemplo, Deheuvels, 1977a; Lejeune, 1985; Ruppert e Wand, 1992; Jones e Foster, 1993). Uma forma simples de construir um núcleo de ordem par $k > 2$ é tomar um polinómio simétrico de ordem $k-2$ definido no intervalo $[-1, 1]$, sendo os seus $k/2$ coeficientes determinados pelas condições $m_{2j}(K) = 0$ para $j = 0, 1, \ldots, k/2 - 1$. Por exemplo, se admitirmos que K é um polinómio simétrico de segunda ordem sobre o intervalo $[-1, 1]$,

$$K(x) = a + bx^2,$$

então K é de ordem 4 se

$$a + \frac{b}{3} = \frac{1}{2} \quad \text{e} \quad \frac{a}{3} + \frac{b}{5} = 0.$$

O núcleo de ordem 4 é assim dado por

$$K(x) = \frac{3}{8}(3 - 5x^2) I(|x| \leq 1).$$

Denotando por $K_{[k]}$ um núcleo simétrico e diferenciável de ordem k (k par), a construção de um núcleo de ordem $k+2$ pode ser feita a partir da equação

$$K_{[k+2]}(x) = \frac{k+1}{k} K_{[k]}(x) + \frac{1}{k} x K'_{[k]}(x).$$

Tomando $K_{[2]} = \phi$, com $\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$, $x \in \mathbb{R}$, os núcleos $K_{[k]}$ de ordem superior que se obtêm a partir de relação de recorrência anterior são os núcleos G_k definidos por

$$G_k(x) = \sum_{s=0}^{k/2-1} \frac{(-1)^s}{(2^s s!)} \phi^{(2s)}(x) \tag{3.8.3}$$

(ver Deheuvels, 1977a; Wand e Schucany, 1990).

Análise do viés

O viés, local e global, do estimador do núcleo quando K é um núcleo de ordem k é descrito no resultado seguinte.

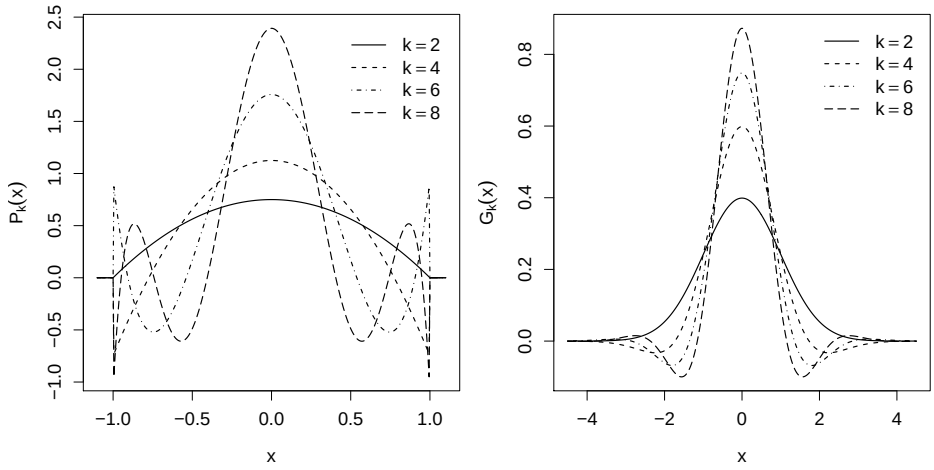

Figura 3.8.2: *Núcleos de ordem k polinomiais (P_k) e núcleos de ordem k baseados na densidade normal standard (G_k).*

Teorema 3.8.4. *Sejam K um núcleo simétrico de ordem k e f com derivada de ordem p contínua e limitada em $\mathbb{R}$. Para $x \in \mathbb{R}$, se $h_n \to 0$ então*

$$\mathrm{E}f_n(x) - f(x) = \frac{h_n^r}{r!}\, m_r(K) f^{(r)}(x) + o(h_n^r),$$

onde

$$r = \min\{k, p\}.$$

Além disso, se $f^{(r)}$ é de quadrado integrável temos

$$\mathrm{IVIES}(f_n) = \frac{h_n^{2r}}{(r!)^2}\, m_r^2(K) R(f^{(r)}) + o(h_n^{2r}).$$

Dem: Basta ter em conta que o desenvolvimento (3.8.1) é válido com $r = \min\{k, p\}$ no lugar de k e proceder como nas demonstrações dos Teoremas 3.2.6 e 3.6.1.

□

Ordem de convergência óptima

Conjugando o resultado anterior com a expressão da variância do estimador apresentada no Teorema 3.2.11, concluímos que

$$\mathrm{EQM}(f_n(x)) = \frac{1}{nh_n}\, R(K)\, f(x) + \frac{h_n^{2k}}{(k!)^2}\, m_k^2(K) f^{(k)}(x)^2 + o(h_n^{2k}),$$

sempre que f admita derivada de ordem k contínua e limitada em $\mathbb{R}$. A maior ordem de convergência é obtida tomando

$$h_n = c\, n^{-1/(2k+1)}, \text{ com } c > 0,$$

caso em que obtemos

$$\mathrm{EQM}(f_n(x)) = O\big(n^{-2k/(2k+1)}\big).$$

Sendo $\mathscr{C}_{k\alpha}$ o conjunto das densidades de probabilidade definido em §2.9, é ainda possível mostrar que a ordem de convergência $n^{-2k/(2k+1)}$ atingida pelo estimador do núcleo é uniforme relativamente a $\mathscr{C}_{k\alpha}$:

$$\limsup_{n\to\infty} \sup_{f\in\mathscr{C}_{k\alpha}} n^{2k/(2k+1)}\, \mathrm{EQM}(f_n(x)) < \infty.$$

Tendo em conta o Teorema 2.9.1, isto significa que a ordem de convergência $n^{-2k/(2k+1)}$ é a maior possível. Não existe um estimador da densidade cujo erro quadrático médio no ponto x convirja para zero, uniformemente em $\mathscr{C}_{k\alpha}$, com uma ordem de convergência superior a $n^{-2k/(2k+1)}$.

EQMI e janela assintoticamente óptima

Dos Teoremas 3.6.1 e 3.8.4 concluímos que o erro quadrático médio integrado admite o desenvolvimento

$$\mathrm{EQMI}(f_n) = \frac{1}{nh_n}\, R(K) + \frac{h_n^{2k}}{(k!)^2}\, m_k^2(K) R(f^{(k)}) + O(n^{-1}) + o\left(h_n^{2k}\right),$$

sendo a janela assintoticamente óptima, no sentido da minimização das duas parcelas mais representativas do desenvolvimento anterior, dada por

$$h_{\mathrm{EQMIA}} = \left(\frac{(k!)^2 R(K)}{2k\, m_k^2(K) R(f^{(k)})}\right)^{1/(2k+1)} n^{-1/(2k+1)}.$$

Para uma janela h_n da forma

$$h_n = c\, n^{-1/(2k+1)}, \text{ com } c > 0,$$

obtemos um erro quadrático médio integrado da ordem

$$\mathrm{EQMI}(f_n) = O\big(n^{-2k/(2k+1)}\big),$$

onde $2k/(2k+1)$ é uma função crescente de k que converge para 1, aproximando-se assim o EQMI da ordem de convergência n^{-1}. No entanto esta ordem de convergência não é atingida por qualquer estimador do núcleo com $K \in L_2$ e para qualquer densidade $f \in L_2$. Com efeito, de (3.5.2) temos

$$nh_n\mathrm{EQMI}(f_n) \geq R(K) - h_n \int K_{h_n} * f(x)^2 dx,$$

ou ainda,

$$\liminf nh_n\mathrm{EQMI} \geq R(K),$$

o que implica que

$$n\,\mathrm{EQMI}(f_n) \to +\infty.$$

Supernúcleos

Ao fixarmos a ordem k do núcleo, estamos naturalmente a limitar a ordem de convergência do viés do estimador do núcleo para densidades que possam admitir derivadas contínuas e limitadas para além da ordem k. Com efeito, do Teorema 3.8.4 sabemos que

$$\mathrm{E}f_n(x) - f(x) = O\big(h_n^{\min\{k,p\}}\big),$$

onde f admite derivada de ordem p contínua e limitada em $\mathbb{R}$. A utilização de um núcleo de ordem k não permite assim que o viés do estimador convirja para zero mais rapidamente que h_n^k. Esta limitação seria evitada se o núcleo K admitisse momentos nulos de todas as ordens. Neste caso,

$$\mathrm{E}f_n(x) - f(x) = O\big(h_n^p\big),$$

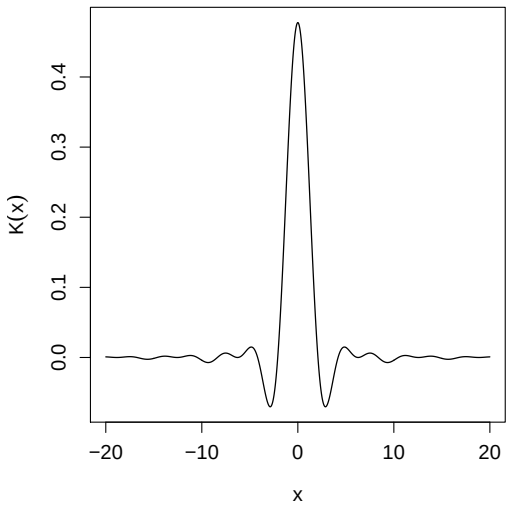

Figura 3.8.5: *Supernúcleo trapezoidal.*

ficando a ordem de convergência do viés unicamente dependente da regularidade de f. Os supernúcleos, que são caracterizados pelo facto da sua função característica ser constante numa vizinhança da origem, possuem a propriedade anterior (ver Devroye, 1987, Cap. 7; Chacón et al., 2007a). Um exemplo clássico de um supernúcleo é dado pelo núcleo trapezoidal (ver Figura 3.8.5)

$$K(x) = \frac{\cos(x) - \cos(2x)}{\pi\, x^2},$$

de função característica $\varphi_K(t) = I(0 \leq |t| < 1) + (2 - |t|)I(1 \leq |t| < 2)$.

Devido ao caracter assintótico dos resultados anteriores, não é de todo claro que para os tamanhos de amostras que ocorrem em situações práticas, a utilização de núcleos de ordem superior à segunda implique uma melhoria efectiva da qualidade do estimador. Acrescem a isto, problemas relacionados com a escolha prática da janela. Sobre esta e outras questões relativas à utilização de núcleos de ordem superior ver Marron e Wand (1992) e Jones e Signorini (1997).

3.9 Escolha prática de h_n

Dos dois parâmetros, núcleo e janela, que devem ser fixados pelo utilizador, vimos já que, limitando a nossa escolha à classe dos núcleos simétricos de segunda ordem, a escolha do núcleo não é de importância fundamental. A sua escolha é habitualmente pautada por questões relacionadas com propriedades de regularidade que pretendemos que as estimativas de densidade possuam e não por princípios de eficiência uma vez que, como vimos, núcleos diferentes do óptimo possuem uma eficiência, no sentido do erro quadrático médio integrado assintótico, próxima da unidade. Admitiremos no que se segue que K é uma densidade de probabilidade simétrica.

Tal como acontecia no caso do histograma, a escolha da janela h_n é uma questão crucial na implementação prática do estimador do núcleo. Nesta secção descrevemos alguns dos métodos automáticos que são utilizados na prática para efectuar esta escolha (ver também Wand e Jones, 1995; Simonoff, 1996; Sheather, 2004). Centraremos a nossa atenção na descrição das suas propriedades assintóticas que, muitas vezes, não reflectem de forma fidedigna o seu comportamento a distância finita. Por esta razão, são diversos os estudos de simulação que podemos encontrar na literatura com o objectivo de comparar os diversos métodos, não havendo nenhum método para a escolha da janela h_n que, relativamente a um conjunto vasto de densidades f, seja uniformemente melhor que os restantes. A título de exemplo refiram-se os estudos de Cao et al. (1994), Berlinet e Devroye (1994), Chiu (1996) e Jones et al. (1996b).

3.9.1 Métodos de utilização simples

Para uma vasta classe de densidades conhecemos, a menos da quantidade $R(f'')$, a janela óptima no sentido da minimização do erro quadrático médio integrado assintótico:

$$h_{\text{EQMIA}} = c_K R(f'')^{-1/5} n^{-1/5},$$

onde

$$c_K = \big(R(K)/m_2^2(K) \big)^{1/5}.$$

Janela com distribuição de referência normal

Deheuvels (1977a) e Deheuvels e Hominal (1980) propõem que se utilize a distribuição normal como distribuição de referência e que $R(f'')$ seja calculada tomando para f a densidade normal com média 0 e desvio-padrão σ. Neste caso,

$$R(f'') = \frac{3}{8\sqrt{\pi}\,\sigma^5}$$

e

$$h_{\text{EQMIA}} = c_K \left(\frac{8\sqrt{\pi}}{3} \right)^{1/5} \sigma\, n^{-1/5} \approx 1.36\, c_K\, \sigma\, n^{-1/5}.$$

Finalmente, estimando σ pelo desvio-padrão empírico corrigido $\hat{s}$ obtemos a janela com distribuição de referência normal:

$$\hat{h}_{\text{NR}} = 1.36\, c_K\, \hat{s}\, n^{-1/5}.$$

Uma janela alternativa pode ser obtida se como estimador de σ usarmos $\hat{\text{AIQ}}/1.349$, onde $\hat{\text{AIQ}}$ é a amplitude inter-quartil empírica (ver §2.8.1). Silverman (1986) sugere que se estime a escala da distribuição pelo estimador combinado

$$\hat{\sigma}_{\text{SIL}} = \min\big(\hat{s}, \hat{\text{AIQ}}/1.349\big), \tag{3.9.1}$$

dando origem à janela

$$\hat{h}_{\text{SIL}} = \min(1.36\,\hat{s}, 1.01\,\hat{\text{AIQ}})\, c_k\, n^{-1/5}.$$

No caso de tomarmos para núcleo K o núcleo normal, obtemos

$$\hat{h}_{\text{SIL}} = \min(1.06\,\hat{s}, 0.79\,\hat{\text{AIQ}})\, n^{-1/5}.$$

O princípio da distribuição mais suave

No contexto do estimador do núcleo e tomando para parâmetro de escala o desvio-padrão σ, o princípio da distribuição mais suave introduzido por Terrell e Scott (1985) e Terrel (1990), a que fizemos referência em §2.8.1, consiste em tomar como distribuição de referência a densidade f com variância σ^2 que torna mínima a quantidade $R(f'')$. Terrel (1990) prova que a densidade com variância σ^2 que é solução do problema anterior é dada por g_δ onde $g(x) = \frac{35}{32}(1-x^2)^3 I(|x| \leq 1)$ e $\delta = 3\sigma$. Como $R(g'') = 35$ obtemos

$$\hat{h}_{\mathrm{OS}} = 1.47\, c_K\, \hat{s}\, n^{-1/5}.$$

No caso de K ser o núcleo normal, temos

$$\hat{h}_{\mathrm{OS}} = 1.14\, \hat{s}\, n^{-1/5}.$$

Com excepção da família de densidades tomada para distribuição de referência, os métodos anteriores não produzem janelas automáticas assintoticamente equivalentes à janela assintoticamente óptima. Os métodos que descrevemos a seguir para a escolha automática da janela possuem essa propriedade para uma vasta família de densidades de probabilidade.

3.9.2 Método de validação cruzada baseado no EQI

Como já referimos a propósito do estimador do histograma, o método de validação cruzada baseado no erro quadrático integrado $\mathrm{EQI} = ||f_n - f||_2^2$ proposto por Rudemo (1982) e Bowman (1984) (*least-square cross-validation* ou *unbiased least-square cross-validation*) tem como objectivo escolher uma janela que minimize a função aleatória

$$\begin{aligned} \mathrm{EQI}(h) &= \int \{f_n(x) - f(x)\}^2 dx \\ &= R(f_n) - 2\int f_n(x) f(x)\, dx + R(f), \end{aligned}$$

onde, por simplicidade de escrita, escrevemos h em vez de h_n.

Procedendo por analogia ao que fizemos para o histograma, para $i = 1, \ldots, n$, vamos considerar o estimador do núcleo, $f_{n,-i}$, associado à subamostra $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$, isto é,

$$f_{n,-i}(x) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}} K_h(x - X_j),$$

e tomamos a estatística

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

para aproximar $\int f_n(x) f(x)\, dx$.

Na prática toma-se para janela o valor $\hat{h}_{\mathrm{CV}}$ que minimiza o estimador cêntrico de variância mínima de $\mathrm{EQMI}(f_n) - R(f)$ (cf. Teorema 3.5.1) dado por

$$\begin{aligned} \mathrm{CV}(h) &= R(f_n) - \frac{2}{n} \sum_{i=1}^{n} f_{n,-i}(X_i) \\ &= \frac{R(K)}{nh} + \frac{1}{n(n-1)} \sum_{i \neq j} (\tfrac{n-1}{n} K_h * K_h - 2K_h)(X_i - X_j). \end{aligned}$$

As primeiras propriedades teóricas da janela aleatória anterior foram estabelecidas por Hall (1983) e Stone (1984). Sob condições pouco restritivas sobre o núcleo e sobre f estes autores mostram que $\hat{h}_{\mathrm{CV}}$ é assintoticamente óptima no sentido em que vale a convergência

$$\frac{\mathrm{EQI}(\hat{h}_{\mathrm{CV}})}{\mathrm{EQI}(\hat{h}_{\mathrm{EQI}})} \xrightarrow{qc} 1,$$

onde $\hat{h}_{\mathrm{EQI}}$ é a janela aleatória óptima no sentido do EQI:

$$\hat{h}_{\mathrm{EQI}} = \mathrm{argmin}_{h>0} \mathrm{EQI}(h).$$

Além disso, vale a convergência

$$\frac{\hat{h}_{\mathrm{CV}}}{h_{\mathrm{EQMI}}} \xrightarrow{qc} 1,$$

(cf. Hall, 1983) onde h_{EQMI} é a janela óptima no sentido da minimização de EQMI, isto é,

$$h_{\text{EQMI}} = \text{argmin}_{h>0}\text{EQMI}(h)$$

(sobre a existência e o comportamento assintótico da janela h_{EQMI} ver Chacón et al., 2007b). Hall e Marron (1987a) precisam ainda a ordem de convergência com que o erro relativo $\hat{h}_{\text{CV}}/h_{\text{EQMI}} - 1$ converge em probabilidade para zero, estabelecendo que

$$n^{1/10}\left(\frac{\hat{h}_{\text{CV}}}{h_{\text{EQMI}}} - 1\right) \xrightarrow{d} N(0, \sigma^2_{\text{CV}}),$$

onde $\sigma^2_{\text{CV}} > 0$ depende de f e de K. Esta é efectivamente uma baixa ordem de convergência para o erro relativo de $\hat{h}_{\text{CV}}$ quando comparada com a ordem de convergência óptima $n^{-1/2}$ (cf. Hall e Marron, 1991). Esta baixa ordem de convergência reflecte-se na prática no facto da janela $\hat{h}_{\text{CV}}$ apresentar uma grande variabilidade amostral.

3.9.3 Estimação de funcionais da densidade

Alguns dos métodos automáticos para a escolha da janela do estimador do núcleo que apresentamos a seguir passam, de uma forma ou de outra, pela estimação da quantidade desconhecida $R(f'')$ que surge nas expressões do EQMIA ou da janela assintoticamente óptima h_{EQMIA}, ou mais geralmente, pela estimação duma funcional do tipo

$$\theta_r = R(f^{(r)}) = \int f^{(r)}(x)^2 dx$$

onde r é um inteiro não negativo e $f^{(r)}$ é de quadrado integrável. Neste parágrafo estudamos uma classe de estimadores do núcleo deste tipo de funcionais (ver Chiu, 1991, e Wu, 1995, para uma classe alternativa de estimadores baseados na função característica empírica).

Se f admite derivadas até à ordem $2r$ limitadas e de quadrado integrável, a funcional anterior pode ser escrita na forma

$$\theta_r = (-1)^r \int f^{(2r)}(x) f(x) dx = (-1)^r \,\text{E}(f^{(2r)}(X)),$$

o que motiva considerarmos o estimador de $R(f^{(r)})$ dado por

$$\hat{\theta}_r(g) = \frac{(-1)^r}{n^2} \sum_{i,j=1}^{n} L_g^{(2r)}(X_i - X_j), \tag{3.9.2}$$

onde assumimos que L é um núcleo simétrico com derivada até à ordem $2r$, $g = g_n > 0$ é a janela e $L_g^{(s)}$ representa a derivada de ordem s da função $L_g(x) = L(x/g)/g$, isto é, $L_g^{(s)}(x) = L^{(s)}(x/g)/g^{s+1}$ (cf. Hall e Marron, 1987b, 1991; Jones e Sheather, 1991).

No resultado seguinte apresentam-se desenvolvimentos assintóticos para o viés e para a variância de $\hat{\theta}_r(g)$.

Teorema 3.9.3. *Seja L um núcleo simétrico de ordem 2ν, $\nu \in \mathbb{N}$, com derivadas limitadas e de quadrado integrável até à ordem $2r$, tal que $\int |u|^{2\nu+1}|L(u)|du < \infty$. Se f possui derivadas limitadas, contínuas e de quadrado integrável até à ordem $2r + 2\nu + 1$ e $g = g_n \to 0$ então:*

$$\begin{aligned} a)\ \operatorname{Vi\acute{e}s} \hat{\theta}_r(g) &= \frac{(-1)^r L^{(2r)}(0)}{n g^{2r+1}} + g^{2\nu} \frac{(-1)^\nu m_{2\nu}(L)}{(2\nu)!}\, \theta_{r+\nu} \\ &\quad + O(n^{-1}) + O(g^{2\nu+1}). \end{aligned}$$

$$\begin{aligned} b)\ \operatorname{Var} \hat{\theta}_r(g) &= \frac{4}{n} \operatorname{Var} f^{(2r)}(X_1) + \frac{2}{n^2\, g^{4r+1}}\, R(L^{(2r)})\theta_0 \\ &\quad + O\left(n^{-1} g^{2\nu}\right) + o\left(n^{-2} g^{-4r-1}\right). \end{aligned}$$

Dem: Vamos limitar-nos aqui ao cálculo do viés do estimador. A variância pode ser obtida usando a estrutura de U-estatística de $\hat{\theta}_r(g)$ e os resultados clássicos de Hoeffding (1948) (ver também Wand e Jones, 1995, §3.5). O desenvolvimento apresentado para o viés de $\hat{\theta}_r(g)$ decorre da expressão

$$\begin{aligned} \mathrm{E}\, \hat{\theta}_r(g) &= \frac{(-1)^r L_g^{(2r)}(0)}{n} + (-1)^r \left(1 - n^{-1}\right) \mathrm{E} L_g^{(2r)}(X_1 - X_2) \\ &= \frac{(-1)^r L^{(2r)}(0)}{n g^{2r+1}} + (-1)^r \left(1 - n^{-1}\right) \iint L_g^{(2r)}(x - y) f(x) f(y) dx dy, \end{aligned}$$

onde

$$\int L_g^{(2r)}(x-y)f(x)dx$$
$$= \int L_g(x-y)f^{(2r)}(x)dx = \int L(u)f^{(2r)}(x-ug)dx$$

$$= f^{2r)}(y) + g^{2\nu}m_{2\nu}(L)f^{2r+2\nu)}(y)$$
$$- g^{2\nu+1}\int\int_0^1 u^{2\nu+1}L(u)\frac{(1-t)^{2\nu}}{(2\nu)!}f^{(2r+2\nu+1)}(y-tug)dtdu,$$

e

$$\iint L_g^{(2r)}(x-y)f(x)f(y)dxdy = (-1)^r\theta_r + g^{2\nu}\frac{(-1)^{r+\nu}m_{2\nu}(L)}{(2\nu)!}\,\theta_{r+\nu} + O(g^{2\nu+1}).$$

□

Concluímos assim que

$$\begin{aligned}\mathrm{EQM}(\hat{\theta}_r(g)) = \frac{4}{n}\,\mathrm{Var}f^{(2r)}(X_1) + \frac{2}{n^2\,g^{4r+1}}\,R(L^{(2r)})\theta_0 \\ + \left(\frac{(-1)^r L^{(2r)}(0)}{ng^{2r+1}} + g^{2\nu}\theta_{r+\nu}\frac{(-1)^\nu m_{2\nu}(L)}{(2\nu)!}\right)^2 \\ + o\left(n^{-2}g^{-4r-1}\right) + O\left(n^{-1}g^{2\nu-2r} + g^{4\nu+1}\right),\end{aligned}$$

e portanto, se $ng^{2r+1} \to +\infty$ temos

$$\hat{\theta}_r(g) \xrightarrow{p} \theta_r.$$

Se $(-1)^{r+\nu}L^{(2r)}(0)m_{2\nu}(L) < 0$, a janela assintoticamente óptima, no sentido da minimização dos termos mais representativos do EQM anterior, anula os termos dominantes do viés do estimador sendo dada por

$$g_{\mathrm{EQMA}} = \left(\frac{(-1)^{r+\nu+1}(2\nu)!L^{(2r)}(0)}{m_{2\nu}(L)\,\theta_{r+\nu}\,n}\right)^{1/(2r+2\nu+1)}.$$

Neste caso,

$$\hat{\theta}_r(g_{\mathrm{EQMA}}) - \theta_r = \begin{cases} O_p\left(n^{-(2\nu+1/2)/(2r+2\nu+1)}\right) & \text{se } \nu < r \\ O_p(n^{-1/2}) & \text{se } \nu \geq r, \end{cases}$$

Quando a janela g é escolhida da forma $g = c\, n^{-1/(2r+2\nu+1)}$, com $c > 0$, podemos concluir que

$$\hat{\theta}_r(g) - \theta_r = \begin{cases} O_p\left(n^{-2\nu/(2r+2\nu+1)}\right) & \text{se } \nu \leq r \\ O_p(n^{-1/2}) & \text{se } \nu > r, \end{cases}$$

o que ilustra bem o efeito de considerarmos núcleos L de ordem superior no estimador $\hat{\theta}_r(g)$: a ordem de convergência paramétrica $n^{-1/2}$ pode ser alcançada desde que L possua uma ordem superior a $2r$. No entanto, uma tal a ordem é obtida à custa de impormos condições de regularidade mais restritivas sobre a densidade f.

As condições impostas sobre o núcleo L são em particular satisfeitas quando tomamos $L = G_{2\nu}$, onde $G_{2\nu}$ é o núcleo de ordem 2ν definido a partir da densidade normal standard por (3.8.3). Neste caso, atendendo ao facto da função $(-1)^r L^{(2r)}$ ser definida positiva, podemos ainda concluir que o estimador $\hat{\theta}_r(g)$ preserva o sinal de $R(f^{(2r)})$.

No caso particular em que L é da forma $L = K * K$, onde K é um núcleo simétrico de ordem 2ν, com derivadas limitadas e integráveis até à ordem r, as condições do teorema anterior são satisfeitas e o estimador $\hat{\theta}_r(g)$ não é mais do que o estimador que se obtém substituindo em $R(f^{(r)})$ a densidade f pelo estimador do núcleo da densidade com núcleo K e janela g:

$$\begin{aligned} \hat{\theta}_r(g) &= \frac{(-1)^r}{n^2} \sum_{i,j=1}^{n} K_g^{(r)} * K_g^{(r)}(X_i - X_j) \\ &= \int \left(\frac{1}{n} \sum_{i=1} K_g^{(r)}(x - X_i) \right)^2 dx \\ &= R(f_n^{(r)}) \end{aligned}$$

(cf. Nadaraya, 1974).

3.9.4 Outros métodos de validação cruzada

No sentido de melhorar as propriedades assintóticas e a distância finita da janela $\hat{h}_{\mathrm{CV}}$, foram vários os métodos propostos na literatura para a escolha

automática da janela. Para uma breve descrição de muitos deles, ver, por exemplo, Cao et al. (1994), Jones et al. (1996a,b) e Chiu (1996). Neste parágrafo centramos a nossa atenção em alguns dos métodos de validação cruzada propostos.

Método de validação cruzada baseado no EQMIA

Um primeiro passo no sentido da diminuição da variabilidade da janela $\hat{h}_{\mathrm{CV}}$ é dado por Scott e Terrel (1987). Estes autores propõem um método de validação cruzada (*biased least-square cross-validation*) baseado na minimização do estimador do EQMIA que se obtém substituindo a quantidade desconhecida $R(f'')$ por um estimador baseado em $R(f_n'')$, onde f_n é o estimador do núcleo com núcleo K e janela h_n.

Sabemos do Teorema 3.9.3 que quando h_n é da ordem de $n^{-1/5}$, $R(f_n'')$ é um estimador assintoticamente enviesado de $R(f'')$:

$$\mathrm{E}(R(f_n'')) = R(f'') + \frac{R(K'')}{nh_n^5} + o(n^{-2/5}).$$

Este facto leva a que Scott e Terrel (1987) (ver também Deheuvels e Hominal, 1980) sugiram a utilização do estimador corrigido,

$$R(f_n'') - \frac{R(K'')}{nh_n^5} = \frac{1}{n^2}\sum_{i\neq j} K_{h_n}'' * K_{h_n}''(X_i - X_j).$$

Assim na prática toma-se para janela o valor $\hat{h}_{\mathrm{BCV}}$ que minimiza

$$\mathrm{BCV}(h) = \frac{1}{nh}\,R(K) + \frac{h^4}{4}\,m_2^2(K)\frac{1}{n^2}\sum_{i\neq j} K_h'' * K_h''(X_i - X_j).$$

A principal vantagem assintótica da janela $\hat{h}_{\mathrm{BCV}}$ relativamente a $\hat{h}_{\mathrm{CV}}$ é a sua menor variabilidade. Sob certas condições de regularidade sobre f, Scott e Terrel (1987) e Jones e Kappenman (1992), provam que

$$n^{1/10}\left(\frac{\hat{h}_{\mathrm{BCV}}}{h_{\mathrm{EQMI}}} - 1\right) \xrightarrow{d} N(0, \sigma^2_{\mathrm{BCV}}),$$

onde $\sigma^2_{\text{BCV}} > 0$ depende de f e de K mas $\sigma^2_{\text{CV}}/\sigma^2_{\text{BCV}}$ depende apenas do núcleo K. No caso de K ser o núcleo normal standard temos

$$\frac{\sigma^2_{\text{CV}}}{\sigma^2_{\text{BCV}}} \approx 15.7,$$

o que indica que devemos esperar uma menor variabilidade amostral de $\hat{h}_{\text{BCV}}$ relativamente a $\hat{h}_{\text{CV}}$. Apesar desta expectável menor variabilidade, a ordem de convergência obtida para o erro relativo de $\hat{h}_{\text{BCV}}$ é igual à da janela $\hat{h}_{\text{CV}}$.

Métodos de validação cruzada suave

Uma abordagem diferente é sugerida por Hall et al. (1992) e Jones et al. (1991), que partindo da aproximação do EQMI dada por

$$\frac{R(K)}{nh} + \int \{K_h * f(x) - f(x)\}^2 dx,$$

sugerem estimá-la por versões da função

$$\text{SCV}(h) = \frac{R(K)}{nh} + \int \{K_h * \hat{f}_L(x;g) - \hat{f}_L(x;g)\}^2 dx,$$

onde $\hat{f}_L(\cdot;g)$ é um estimador preliminar de f com núcleo L e janela g. Reparemos que, contrariamente aos métodos de validação cruzada anteriores, este método não é completamente automático uma vez que é necessário proceder a escolhas preliminares de L e g. Para escolhas adequadas da janela g e do núcleo L é possível provar que a janela $\hat{h}_{\text{SCV}}$ (*smoothed cross-validation bandwidth*) que minimiza $\text{SCV}(h)$ atinge a ordem de convergência óptima $n^{-1/2}$:

$$\frac{\hat{h}_{\text{SCV}}}{h_{\text{EQMI}}} - 1 = O_p(n^{-1/2}).$$

Na versão do método considerada por Jones et al. (1991), é interessante notar que a ordem de convergência anterior pode ser obtida sem ser necessário utilizar um núcleo L de ordem superior à segunda.

Método de validação cruzada de Chiu

A ordem de convergência anterior é também obtida pela janela gerada pelo método de validação cruzada proposto por Chiu (1991). Este autor usa uma representação de $\mathrm{CV}(h)$ em termos da função característica empírica $\tilde{\phi}(\lambda) = n^{-1}\sum_{j=1}^{n}\exp(i\lambda X_j)$ devida a Silverman (1986), e faz notar que a maior fonte de variação presente em $\mathrm{CV}(h)$ é principalmente devido às frequências elevadas λ de $\tilde{\phi}(\lambda)$. De forma a reduzir o impacto de tais frequências, é sugerida a modificação da função característica empírica para além de uma frequência de corte Λ o que dá origem à função de validação cruzada estabilizada

$$\mathrm{S}(h) = \frac{R(K)}{nh} + \frac{1}{2\pi}\int_{|\lambda|\leq\Lambda}\{|\tilde{\phi}(\lambda)|^2 - 1/n\}\{1 - k(h\lambda)\}\,d\lambda,$$

onde $k(t) = \int \exp(itx)K(x)dx$ é a função característica de K. A frequência de corte Λ depende das observações e é ela própria escolhida através de um procedimento de validação cruzada baseado no estimador de Fourier de f como descrito em Chiu (1992). Denotando por $\hat{h}_{\mathrm{S}}$ a janela que minimiza $\mathrm{S}(h)$, sob certas condições de regularidade em f vale a convergência

$$n^{1/2}\left(\frac{\hat{h}_{\mathrm{S}}}{h_{\mathrm{EQMI}}} - 1\right) \xrightarrow{d} N(0, \sigma^2(f)),$$

onde

$$\sigma^2(f) = \frac{4}{25}\left(\frac{\mathrm{E}(f^{(4)}(X_1))^2}{\mathrm{E}^2(f^{(4)}(X_1))} - 1\right). \tag{3.9.4}$$

Reparemos que não só a ordem de convergência é óptima como a variância assintótica anterior coincide com a menor constante que é possível obter, no sentido descrito por Fan e Marron (1992), para o momento de segunda ordem do erro relativo de uma qualquer janela automática.

3.9.5 Métodos *plug-in*

Os métodos *plug-in* para a escolha da janela remontam aos trabalhos de Woodroofe (1970), Nadaraya (1974) e Deheuvels e Hominal (1980), e são

baseados na ideia de substituir as quantidades desconhecidas que surgem na expressão da janela assintoticamente óptima por estimadores convergentes.

Métodos directos

Uma primeira classe de métodos, ditos métodos *plug-in* directos, usa a expressão da janela assintoticamente óptima dada por

$$h_{\text{EQMIA}} = c_K\, \theta_2^{-1/5}\, n^{-1/5},$$

onde $c_K = \big(R(K)/m_2^2(K)\big)^{1/5}$ e $\theta_2 = R(f'')$, e são baseado na janela aleatória que se obtém por substituição da funcional desconhecida θ_2 por um estimador convergente da mesma. Usando o estimador de θ_2 dado por (3.9.2), onde tomamos para L o núcleo de ordem 2ν baseado na densidade normal standard dado por (3.8.3), obtemos a janela

$$\hat{h}_{\text{PI}} = c_K\, \hat{\theta}_2(g)^{-1/5}\, n^{-1/5}.$$

Esta janela não é completamente automática pois depende da escolha de um novo parâmetro de regularização g. Atendendo aos resultados descritos em §3.9.3, a janela g pode ser escolhida com base na fórmula da janela assintoticamente óptima para o estimador $\hat{\theta}_2(g)$ dada por

$$g_2 = \left(\frac{(-1)^{\nu+1}(2\nu)! L^{(4)}(0)}{m_{2\nu}(L)\theta_{2+\nu}\, n} \right)^{1/(5+2\nu)},$$

onde $m_{2\nu}(L) = (-1)^{\nu+1}(2\nu)!/(2^\nu \nu!)$. A utilização desta fórmula levanta o mesmo problema que levantava a expressão de h_{EQMIA}, uma vez que g_2 depende da funcional desconhecida $\theta_{2+\nu}$. Podemos novamente estimar $\theta_{2+\nu}$ usando o estimador $\hat{\theta}_{2+\nu}(g)$, mas sabemos já que a janela assintoticamente óptima deste estimador depende da funcional $\theta_{2+2\nu}$, originando um processo multietápico que na sua etapa j passa pela estimação de $\theta_{2+(j-1)\nu}$ a partir do estimador $\hat{\theta}_{2+(j-1)\nu}(g)$ cuja janela assintoticamente óptima é dada por

$$g_{2+(j-1)\nu} = \left(\frac{(-1)^{j\nu+1}(2\nu)! L^{(4+2(j-1)\nu)}(0)}{m_{2\nu}(L)\theta_{2+j\nu}\, n} \right)^{1/(5+2j\nu)}. \tag{3.9.5}$$

Uma forma de evitar este ciclo vicioso é a de, na última das etapas anteriores, digamos ℓ, estimar a funcional $\theta_{2+\ell\nu}$ usando o método das distribuições de referência. A distribuição de referência mais usada na prática é a distribuição normal de média 0 e de desvio-padrão σ, para a qual a funcional θ_s é dada por

$$\theta_s^{\mathrm{NR}}(\sigma) = \frac{(2s)!}{(2\sigma)^{2s+1} s! \sqrt{\pi}},$$

podendo $\theta_{2+\ell\nu}$ ser estimada por $\theta_{2+\ell\nu}^{\mathrm{NR}}(\hat{\sigma})$, onde $\hat{\sigma}$ é um qualquer estimador de σ. A ideia agora é a de estimar sucessivamente as ℓ funcionais

$$\theta_{2+(\ell-1)\nu}, \theta_{2+(\ell-2)\nu}, \ldots, \theta_{2+\nu}, \theta_2.$$

Para tal, e para $j = \ell, \ldots, 1$, usamos o estimador do núcleo $\hat{\theta}_{2+(j-1)\nu}(g)$ para estimar $\theta_{2+(j-1)\nu}$, com janela $g = \hat{g}_{2+(j-1)\nu}$ dada por (3.9.5) onde $\theta_{2+j\nu}$ é substituída pela estimativa previamente calculada. Este procedimento multietápico define um estimador de θ_2 a ℓ etapas que vamos denotar por $\hat{\theta}_{2,\ell}(\hat{\sigma})$, e também uma janela plug-in a ℓ etapas definida por

$$\hat{h}_{\mathrm{PI},\ell} = c_K \, \hat{\theta}_{2,\ell}(\hat{\sigma})^{-1/5} \, n^{-1/5}.$$

Na prática escolhe-se habitualmente $\nu = 1$, isto é, L é um núcleo de segunda ordem. Se f possui derivadas limitadas até à ordem $4 + 2\ell$ e se existe $\sigma_f \neq 0$ tal que $\hat{\sigma} - \sigma_f = O_p(n^{-1/2})$ é possível mostrar que

$$\frac{\hat{h}_{\mathrm{PI},1}}{h_{\mathrm{EQMI}}} - 1 = O_p\big(n^{-2/7}\big)$$

e que

$$\frac{\hat{h}_{\mathrm{PI},\ell}}{h_{\mathrm{EQMI}}} - 1 = O_p\big(n^{-5/14}\big), \ \ell \geq 2$$

(cf. Tenreiro, 2003). Estes resultados justificam a habitual recomendação de se utilizar $\ell = 2$ quando $\nu = 1$. Apesar da utilização de núcleos de ordem superior à segunda não revelar ganhos práticos relativamente aos núcleos de segunda ordem, ela permite melhorar ligeiramente as ordens

de convergência anteriores no caso de f possuir derivadas limitadas até à ordem $4 + 2\nu\ell$:

$$\frac{\hat{h}_{\mathrm{PI},\ell}}{h_{\mathrm{EQMI}}} - 1 = O_p\big(n^{-2/5}\big),\ \ell \geq 1.$$

A ordem de convergência anterior reflecte uma limitação da janela assintoticamente óptima, h_{EQMIA}, como aproximação da janela óptima h_{EQMI}. Com efeito, sendo K uma densidade de probabilidade simétrica e limitada com $\int |u|^5 K(u)du < \infty$, e possuindo f derivadas contínuas de quadrado integrável até à quarta ordem, é possível provar que

$$h_{\mathrm{EQMI}} = h_{\mathrm{EQMIA}} + d_K\, \theta_2^{-8/5}\, \theta_3\, n^{-3/5} + O(n^{-4/5}), \tag{3.9.6}$$

onde

$$d_K = \frac{1}{20}\, R(K)^{3/5}\, m_2(K)^{-11/5}\, m_4(K)$$

(cf. Hall et al., 1991). Concluímos assim que

$$\frac{h_{\mathrm{EQMIA}}}{h_{\mathrm{EQMI}}} - 1 = O\big(n^{-2/5}\big),$$

não sendo, por isso, possível melhorar a ordem de convergência $n^{-2/5}$ utilizando a metologia *plug-in* a partir da aproximação h_{EQMIA} de h_{EQMI}. No entanto, utilizando uma melhor aproximação de h_{EQMI} como aquela que podemos deduzir de (3.9.6) e definindo a janela *plug-in* por

$$\hat{h}^*_{\mathrm{PI},\ell} = c_K\, \hat{\theta}_{2,\ell}(\hat{\sigma})^{-1/5}\, n^{-1/5} + d_K\, \hat{\theta}_{2,\ell}(\hat{\sigma})^{-8/5}\, \hat{\theta}_{3,\ell}(\hat{\sigma})\, n^{-3/5},$$

é possível provar que se $\nu \geq 2$ e $\ell \geq 2$ então

$$n^{1/2}\left(\frac{\hat{h}^*_{\mathrm{PI},\ell}}{h_{\mathrm{EQMI}}} - 1\right) \xrightarrow{d} N(0, \sigma^2(f)),$$

onde $\sigma^2(f)$ é dado por (3.9.4). Quer em termos da ordem de convergência, quer relativamente à sua variância assintótica, a janela $\hat{h}^*_{\mathrm{PI},\ell}$ possui propriedades de optimalidade assintóticas já apontadas ao método de validação cruzada de Chiu.

O método *plug-in* de Sheather e Jones

Uma abordagem diferente para a selecção automática da janela, que tem as suas origens nos trabalhos de Hall (1980) e Sheather (1983, 1986), é inicialmente sugerida por Park e Marron (1990), sendo posteriormente melhorada por Sheather e Jones (1991). Partindo mais uma vez da expressão para a janela assintoticamente óptima h_{EQMIA}, e tendo por motivação o facto da janela h_{EQMIA} poder ser aproximada por

$$h_{\text{EQMIA}} \approx c_K \, \hat{\theta}_2(g_{\text{EQMA}})^{-1/5} n^{-1/5},$$

onde o estimador $\hat{\theta}_2(g)$ é definido por (3.9.2) com L um núcleo de segunda ordem, e da janela assintoticamente óptima g_{EQMA} poder ser expressa em termos de h_{EQMIA} através da relação

$$g_{\text{EQMA}} = \alpha(h_{\text{EQMIA}}),$$

com

$$\alpha(h) = c_K^{-5/7}(2L^{(4)}(0)/m_2(L))^{1/7}(\theta_2/\theta_3)^{1/7}h^{5/7},$$

propõe-se a escolha da janela $\hat{h}_{\text{SJ}}$ que satisfaça a equação

$$h = c_K \, \hat{\theta}_2(\hat{\alpha}(h))^{-1/5} n^{-1/5},$$

onde $\hat{\alpha}$ é definida como α, mas onde as quantidades desconhecidas θ_2 e θ_3 são substituídas pelos estimadores a uma etapa $\hat{\theta}_{2,1}(\hat{\sigma})$ e $\hat{\theta}_{3,1}(\hat{\sigma})$, respectivamente, onde $\hat{\sigma}$ é um qualquer estimador da escala σ das observações. No caso em que L é o núcleo normal standard, temos

$$\hat{\theta}_{2,1}(\hat{\sigma}) = \hat{\theta}_2\big(1.241 \, \hat{\sigma} \, n^{-1/7}\big)$$

e

$$\hat{\theta}_{3,1}(\hat{\sigma}) = \hat{\theta}_3\big(1.230 \, \hat{\sigma} \, n^{-1/9}\big),$$

com $\hat{\theta}_2$ e $\hat{\theta}_3$ dados por (3.9.2) e

$$\hat{\alpha}(h) = 1.133 \, c_K^{-5/7} \, (\hat{\theta}_{2,1}(\hat{\sigma})/\hat{\theta}_{3,1}(\hat{\sigma}))^{1/7} h^{5/7}.$$

Sob certas condições de regularidade em f, é possível provar que

$$\frac{\hat{h}_{\mathrm{SJ}}}{h_{\mathrm{EQMI}}} - 1 = O_p\big(n^{-5/14}\big)$$

(cf. Sheather e Jones, 1991), obtendo-se uma ordem de convergência igual à do método *plug-in* directo a duas etapas. Uma modificação do método de Sheather e Jones, que aparentemente melhora o seu comportamento a distância finita, foi recentemente proposta por Liao et al. (2010).

3.9.6 Aplicação a um conjunto de dados reais

Retomamos neste parágrafo o conjunto de dados considerado em §2.8.3. Na Figura 3.9.7 apresentamos três estimativas do núcleo obtidas a partir das janelas $\hat{h}_{\mathrm{SIL}}$, $\hat{h}_{\mathrm{CV}}$ e $\hat{h}_{\mathrm{SJ}}$. Na implementação desta última janela tomámos para estimador $\hat{\sigma}$ da escala das observações o estimador (3.9.1). Tal como já acontecia relativamente ao estimador do histograma as janelas obtidas pela regra de Silverman são significativamente maiores que as restantes janelas. Na origem deste facto pode estar uma subestimação das quantidades $R(f'')$ uma vez que a janela de Silverman toma a distribuição normal como distribuição de referência e a distribuição subjacente às observações é marcadamente não normal.

3.10 O estimador automático do núcleo

Sendo a janela h_n escolhida na prática em função das observações, isto é,

$$h_n = h_n(X_1, \ldots, X_n),$$

é natural pretendermos saber se o estimador do núcleo com janela aleatória, designado habitualmente por estimador automático do núcleo, continua a gozar de boas propriedades de convergência como acontece com o estimador com janela determinista. Esta pergunta foi sendo respondida afirmativamente por autores como Wagner (1975), Deheuvels e Hominal (1980), Devroye e Wagner (1979), Devroye e Penrod (1984) e Tenreiro (2001), que

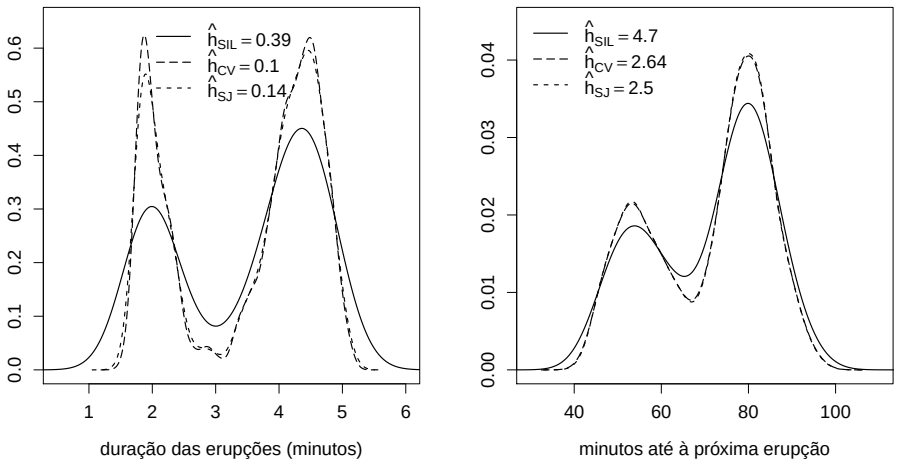

Figura 3.9.7: *Estimativas do núcleo obtidas a partir das janelas* $\hat{h}_{\mathrm{SIL}}$, $\hat{h}_{\mathrm{CV}}$ *e* $\hat{h}_{\mathrm{SJ}}$.

obtiveram para o estimador automático do núcleo, versões dos resultados de convergência que apresentámos atrás. Enunciamos a seguir dois desses resultados, relativos às convergências L_∞ e L_1, que generalizam os Teoremas 3.3.1 e 3.4.1, respectivamente. O primeiro é devido a Wagner (1975) e o segundo surge em Devroye e Györfi (1985, Cap. 6).

Teorema 3.10.1 (Convergência L_∞)**.** *Sejam* K *um núcleo de variação limitada e* $f \in \mathscr{U}$*. Se* $h_n \xrightarrow{p} 0$ *e* $nh_n^2/\log\log n \xrightarrow{p} +\infty$ *então*

$$||f_n - f||_\infty \xrightarrow{p} 0.$$

O resultado permanece válido com $\xrightarrow{qc}$ *no lugar de* $\xrightarrow{p}$*.*

Teorema 3.10.2 (Convergência L_1)**.** *Sejam* $K \in \mathscr{F}$ *limitado de suporte limitado e integrável à Riemann e* $f \in \mathscr{F}$*. Se* $h_n \xrightarrow{p} 0$ *e* $nh_n \xrightarrow{p} +\infty$ *então*

$$||f_n - f||_1 \xrightarrow{p} 0.$$

O resultado permanece válido com $\xrightarrow{qc}$ *no lugar de* $\xrightarrow{p}$*.*

O primeiro dos resultados anteriores pode ser obtido de forma simples adaptando a demonstração do Teorema 3.3.1. O segundo resultado é obtido por Devroye e Györfi (1985) usando um resultado de convergência uniforme relativamente à janela h, isto é, a partir de um resultado do tipo

$$\sup_{h \in H_n} ||f_{nh} - f||_1 \xrightarrow{p} 0,$$

onde f_{nh} é o estimador do núcleo com janela determinista h e H_n é uma apropriada sucessão de intervalos deterministas que converge para zero quando o tamanho da amostra aumenta e que satisfaz

$$I(h_n \notin H_n) \xrightarrow{p} 0,$$

sempre que h_n verifica as condições $h_n \xrightarrow{p} 0$ e $nh_n \xrightarrow{p} +\infty$. A mesma técnica é descrita em Deheuvels e Hominal (1980) com o objectivo de generalizar ao estimador automático o Teorema 3.3.2. Utilizando técnicas de processos empíricos, outros resultados mais recentes sobre a convergência uniforme relativamente à janela que permitem obter a convergência L_∞ para o estimador automático do núcleo são devidos a Einmahl e Mason (2005) e Mason e Swanepoel (2010).

3.11 Estimação em pontos fronteiros

Quando f tem como suporte um intervalo do tipo $[a, +\infty[$ não sendo f contínua no ponto $x = a$, o estimador do núcleo deixa de ser adequado para estimar f numa vizinhança de $x = a$. Este facto é ilustrado na Figura 3.11.1 para o estimador com núcleo de Epanechnikov e janela $h = 0.5$ baseado numa amostra de tamanho 200 da distribuição exponencial de parâmetro 1. A principal razão para o comportamento do estimador para pontos x na região $[a, a+h]$, dita região de fronteira, prende-se com o facto da vizinhança $[x - h, x + h]$ de x, definida pela janela h, ser parcialmente desprovida de observações.

No sentido de analisar o viés do estimador do núcleo na região de fronteira $[a, a + h]$, vamos assumir que o ponto x é da forma $x = a + \alpha h$ para

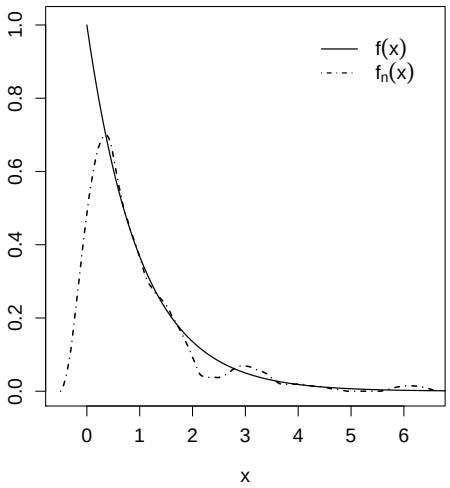

Figura 3.11.1: *Estimativa baseada numa amostra de tamanho 200 da distribuição* $E(1)$ *com* $h = 0.5$ *e núcleo de Epanechnikov.*

algum $\alpha \in]0,1[$ e que K é um núcleo limitado de suporte $[-1,1]$. Se f possui derivada de segunda ordem limitada e contínua à direita de $x = a$, temos

$$\begin{aligned} \mathrm{E}f_n(x) &= \int_a^{+\infty} K_{h_n}(x-y)f(y)dy = \int_{-1}^{\alpha} K(z)f(x-zh)dz \\ &= \int_{-1}^{\alpha} K(z)\left[f(x) - zhf'(x) + z^2h^2\int_0^1 (1-t)f''(x-tzh)dt\right]dz, \end{aligned}$$

o que permite concluir que

$$\mathrm{E}f_n(x) = m_{0,\alpha}(K)f(x) - h\,m_{1,\alpha}(K)f'(x) + \frac{h^2}{2}\,m_{2,\alpha}(K)f''(x) + o(h^2),$$

onde

$$m_{\ell,\alpha}(K) = \int_{-1}^{\alpha} z^\ell K(z)dz.$$

Uma forma de corrigir o viés anterior é proposta por Gasser e Müller (1979), Rice (1984), Gasser et al. (1985) e Müller (1991), e consiste em modificar o estimador do núcleo para pontos $x \in [a, a+h[$. A ideia é substituir $K(\cdot)$ por núcleos especiais $K(\cdot;(x-a)/h)$, ditos núcleos de fronteira, que

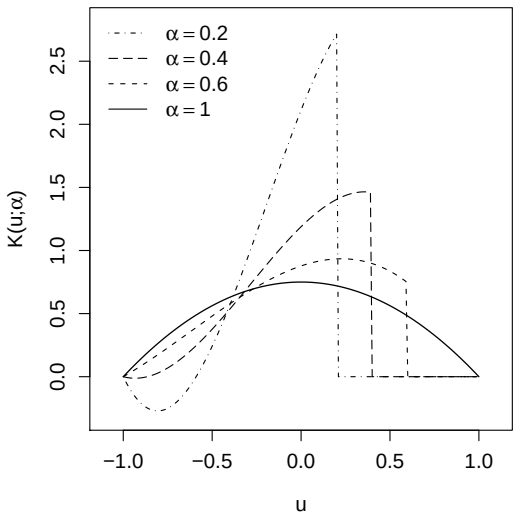

Figura 3.11.3: *Núcleos de fronteira baseados no núcleo de Epanechnikov.*

dependem da distância $\alpha = (x-a)/h$ de x à fronteira, e tais que $K(\cdot;\alpha)$ é um núcleo de segunda ordem em $[-1,\alpha]$, para $\alpha \in [0,1[$. O estimador do núcleo passa então a ser definido por

$$\tilde{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K_{x,h_n}\left(\frac{x-X_i}{h_n}\right), \tag{3.11.2}$$

onde

$$K_{x,h}(u) = \begin{cases} K(u;(x-a)/h), & a \le x < a+h \\ K(u), & x \ge a+h, \end{cases}$$

e $\tilde{f}_n(x) = 0$ para $x < a$. Neste caso, e para $x = a + \alpha h$, obtemos

$$\mathrm{E}\tilde{f}_n(x) = f(x) + \frac{h^2}{2}\int_{-1}^{\alpha} z^2 K(z;\alpha)dz f''(x) + o(h^2),$$

e

$$\mathrm{Var}\tilde{f}_n(x) = \frac{1}{nh}\int_{-1}^{\alpha} K(z;\alpha)^2 dz f(x) + O(n^{-1}),$$

recuperando-se assim a ordem de convergência h^2 para o viés do estimador.

Dado um núcleo K simétrico e não-negativo com suporte $[-1,1]$, uma forma simples de construir núcleos $K(\cdot;\alpha)$ de segunda ordem em $[-1,\alpha]$, é

considerar

$$K(u;\alpha) = \big(A_\alpha + B_\alpha u\big)K(u)I(-1 \le u \le \alpha),$$

onde A_α e B_α são determinados pelas condições $m_{0,\alpha}(K(\cdot;\alpha)) = 1$ e $m_{1,\alpha}(K(\cdot;\alpha)) = 0$, sendo dados por

$$A_\alpha = \frac{m_{2,\alpha}(K)}{m_{0,\alpha}(K)m_{2,\alpha}(K) - m_{1,\alpha}^2(K)}$$

e

$$B_\alpha = \frac{-m_{1,\alpha}(K)}{m_{0,\alpha}(K)m_{2,\alpha}(K) - m_{1,\alpha}^2(K)}.$$

Tomando para K o núcleo de Epanechnikov, os núcleos de fronteira $K(\cdot;\alpha)$ são representados na Figura 3.11.3 para alguns valores de α. Neste caso temos $m_{0,\alpha}(K) = (2 + 3\alpha - \alpha^3)/4$, $m_{1,\alpha}(K) = -3(-1 + \alpha^2)^2/16$ e $m_{2,\alpha}(K) = (2 + 5\alpha^3 - 3\alpha^5)/20$.

No caso de f ter como suporte o intervalo $[a, b]$, a metologia anterior pode ser facilmente adaptada. O estimador com núcleo de fronteira é dado por (3.11.2) com

$$K_{x,h}(u) = \begin{cases} K(u;(x-a)/h), & a \le x < a + h \\ K(u), & a + h \le x \le b - h \\ K(-u;(b-x)/h), & b - h < x \le b, \end{cases}$$

e $\tilde{f}_n(x) = 0$ para $x < a$ ou $x > b$.

Retomando os dados considerados atrás, apresentamos na Figura 3.11.4 a estimativa anteriormente obtida e a nova estimativa obtida utilizando núcleos de fronteira. A partir de uma amostra de tamanho 500 da distribuição com densidade $f(x) = 1 + \sin(2\pi x)$, para $x \in [0, 1]$, representamos na Figura 3.11.5 as estimativas obtidas, com e sem correcção de fronteira, usando o núcleo de Epanechnikov e $h_n = 0.15$.

Procedimentos alternativos para corrigir o problema de viés em pontos fronteiros são considerados, entre outros, por Schuster (1985), Jones (1993), Marron e Ruppert (1994), Jones et al. (1996b), Cowling e Hall (1996), Cheng et al. (1997), Chiu (2000), Karunamunia e Zhang (2008) e Zhang e Karunamuni (2010).

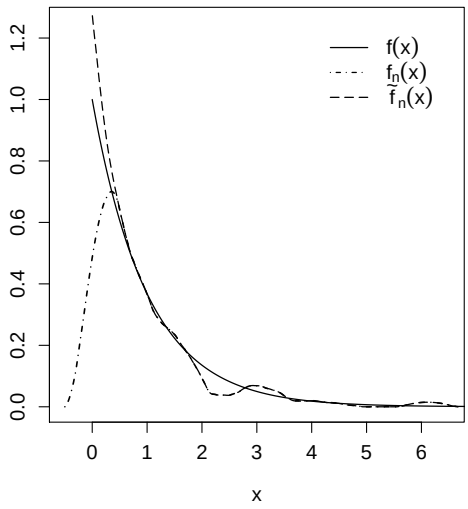

Figura 3.11.4: *Estimativas, com e sem correcção de fronteira, baseadas numa amostra de tamanho 200 da distribuição* $E(1)$ *com* $h_n = 0.5$ *e núcleo de Epanechnikov.*

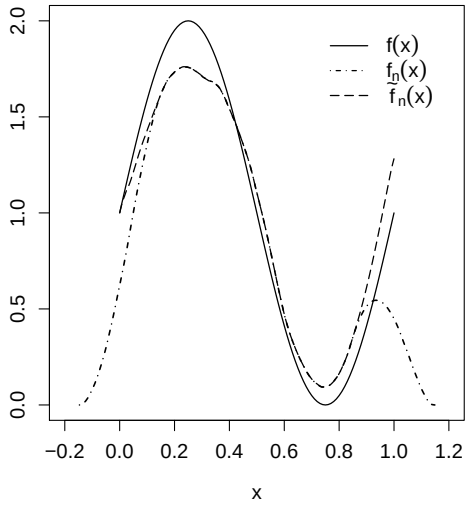

Figura 3.11.5: *Estimativas, com e sem correcção de fronteira, baseadas numa amostra de tamanho 500 da distribuição de densidade* $f(x) = 1 + \sin(2\pi x)$, $x \in [0, 1]$, *com* $h_n = 0.15$ *e núcleo de Epanechnikov.*

3.12 Estimador multivariado do núcleo

Na sua forma mais geral, o estimador do núcleo associado às observações d dimensionais $X_1, \ldots, X_n$ é definido por

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_H(x - X_i)$$

(cf. Deheuvels, 1977b), onde a janela $H = H_n$ é uma matriz simétrica e definida positiva de dimensão $d \times d$, e

$$K_H(x) = |H|^{-1/2} K(H^{-1/2} x),$$

onde $H^{-1/2}$ representa a raiz quadrada de H e o núcleo K é uma função de $\mathbb{R}^d$ em $\mathbb{R}$ com $\int K(x)dx = 1$. O núcleo K é habitualmente uma densidade de probabilidade, sendo o núcleo normal o mais usado na literatura:

$$K(x) = (2\pi)^{-d/2} \exp\big(-x^t x/2\big),$$

onde x^t denota o transposto de x.

Uma forma simplificada do estimador anterior é obtida assumindo que a janela matricial H_n tem uma forma diagonal $H = diag(h_1^2, \ldots, h_d^2)$ o que dá origem ao estimador

$$f_n(x) = \frac{1}{n h_1 \ldots h_d} \sum_{i=1}^{n} K\left(\frac{x_1 - X_{i1}}{h_1}, \ldots, \frac{x_d - X_{id}}{h_d}\right)$$

(cf. Epanechnikov, 1969). Uma maior simplificação pode ainda ser obtida assumindo que $H = h^2 I$, onde I é a matriz identidade. Neste caso o estimador toma a forma clássica considerada por Cacoullos (1966):

$$f_n(x) = \frac{1}{n h^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

Vários autores consideram esta simplificação demasiado restritiva, advogando que pelo menos uma janela diferente segundo a direcção de cada um dos eixo coordenados deve ser considerada (ver Wand e Jones, 1993; Scott e Sain, 2005).

A escolha da janela

Sendo H diagonal e K um núcleo com $\int zK(z)dz = 0$ e $\int zz'K(z)dz = m_2(K)I$, com $m_2(K) = \int z_i^2 K(z)dz$ independente de i, sob certas condições de regularidade sobre f e assumindo que $h_i \to 0$ e $nh_1 \dots h_d \to +\infty$, o erro quadrático médio integrado assintótico do estimador do núcleo toma a forma

$$\text{EQMIA} = \frac{R(K)^d}{nh_1 \dots h_d} + \frac{1}{4}\, m_2(K) \sum_{i,j=1}^{d} h_i^2 h_j^2 \int f_{ii}(x) f_{jj}(x) dx,$$

onde f_{ii} denota a segunda derivada de f relativamente à sua i-ésima variável (cf. Wand, 1992; Wand e Jones, 1995).

Tomando para K o núcleo normal standard e como distribuição de referência a distribuição normal $N(0, \Sigma)$, com $\Sigma = diag(\sigma_1^2, \dots, \sigma_d^2)$, a janela assintoticamente óptima é dada por

$$h_{k,\text{EQMIA}} = \left(\frac{8}{d+5}\right)^{1/(d+4)} \sigma_k\, n^{-1/(d+4)},$$

expressão esta que pode ser usada para propor uma escolha prática da janela. Outros procedimentos para a escolha automática da janela que, contrariamente ao anterior, produzem janelas assintoticamente equivalentes à janela assintoticamente óptima, são estudados em Stone (1984), Sain et al. (1994), Wand e Jones (1994), Duong e Hazelton (2003, 2005) e Chacón e Duong (2010).

Usando a janela anterior com σ_k substituído pelo desvio-padrão empírico, apresentamos na Figura 3.12.1 uma estimativa da densidade obtida a partir de uma amostra de tamanho 500 da mistura de normais bivariada indicada. Retomando os dados descritos em §2.8.3, apresentamos na Figura 3.12.2 uma estimativa da densidade conjunta das duas variáveis observadas.

A maldição da dimensão

Para a distribuição normal multivariada standard $N(0, I)$, e admitindo que o núcleo normal standard é tomado para K, apresentamos na Tabela 3.12.3

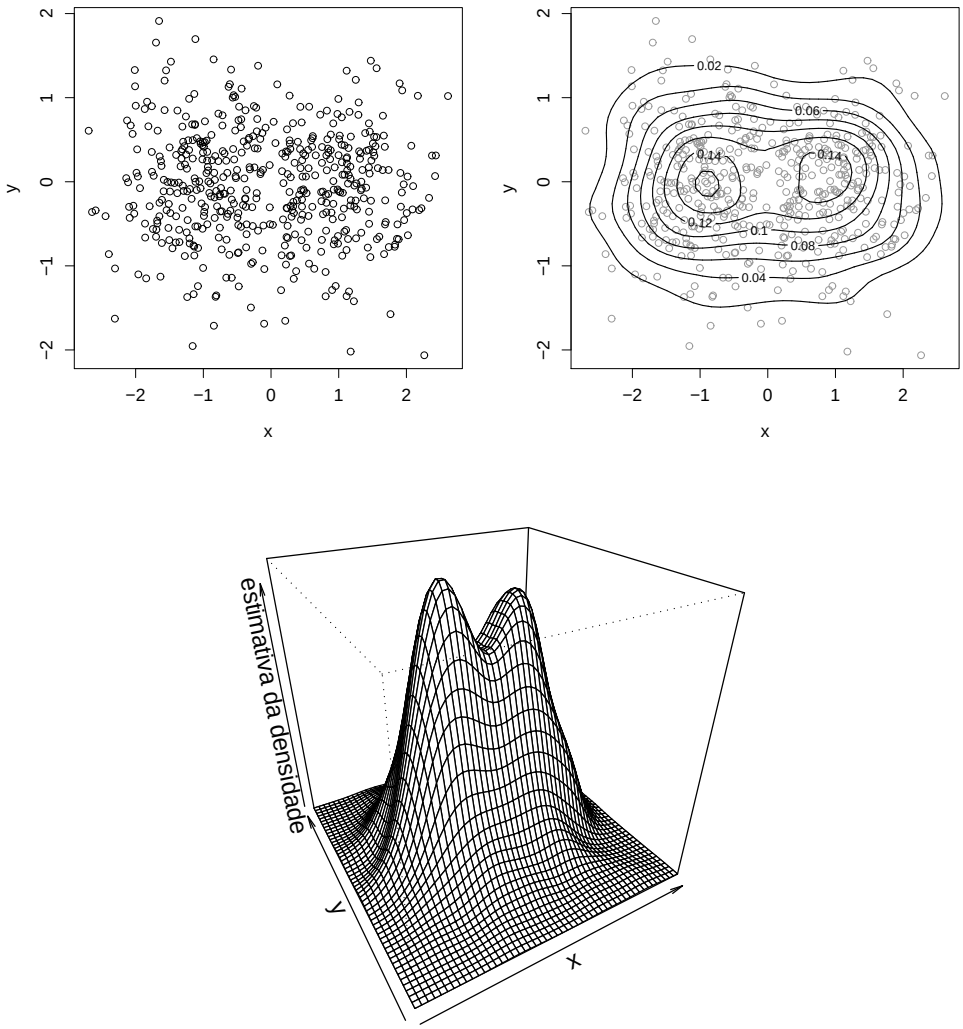

Figura 3.12.1: *Gráfico de dispersão, estimativa da densidade e respectivas curvas de nível obtidas a partir de uma amostra de tamanho* 500 *da mistura de normais* $\frac{1}{2}N\big(-1,0,(\frac{2}{3})^2,(\frac{2}{3})^2,0\big)+\frac{1}{2}N\big(1,0,(\frac{2}{3})^2,(\frac{2}{3})^2,0\big)$.

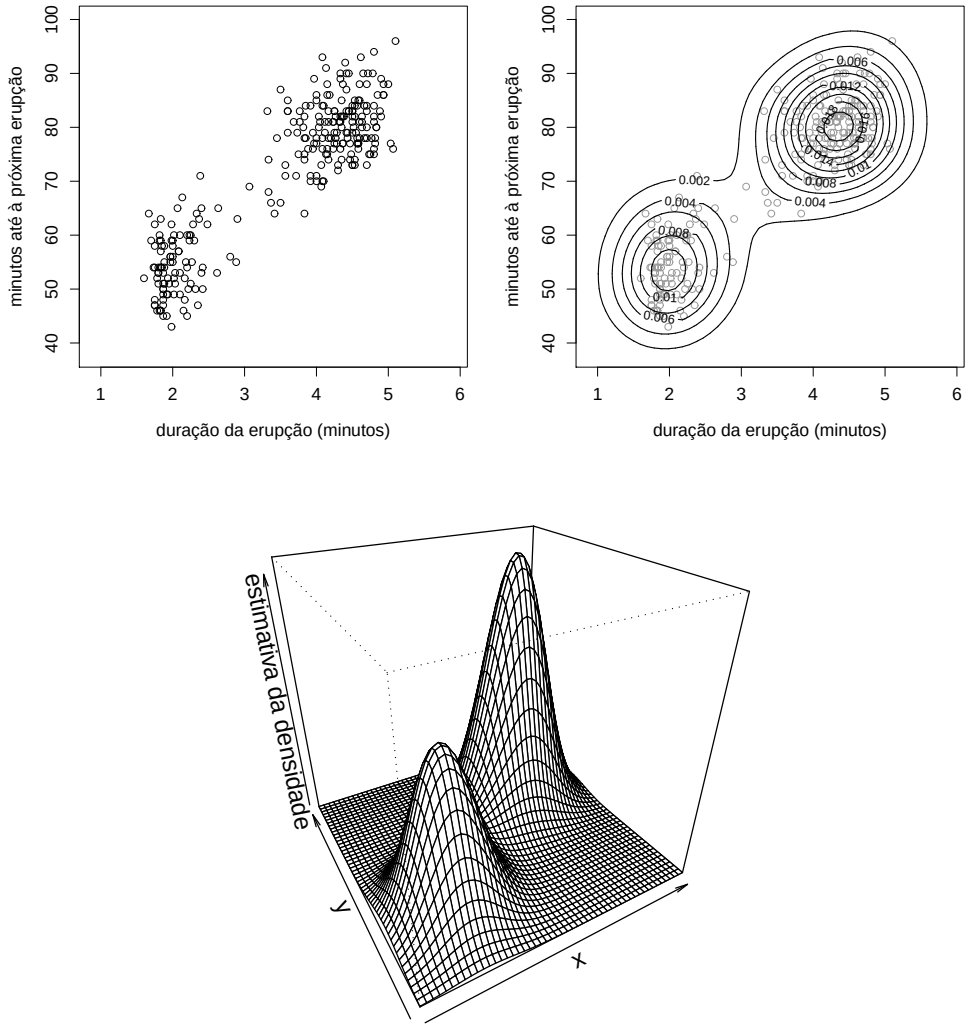

Figura 3.12.2: *Gráfico de dispersão, estimativa da densidade e respectivas curvas de nível obtidas a partir das* 272 *observações relativas às variáveis descritas em §2.8.3 (x - duração da erupção; y - minutos até à próxima erupção).*

d	Tamanhos equivalentes da amostra		
1	10	50	100
2	23	157	361
3	60	569	1502
4	175	2299	6970
5	560	10138	35303

Tabela 3.12.3: *Tamanhos equivalentes da amostra, em função da dimensão d, para a distribuição normal multivariada standard segundo o critério de Epanechnikov.*

os tamanhos das amostras que permitem obter, para várias dimensões d, um mesmo valor da medida normalizada da qualidade do estimador dada por

$$\text{EQMIA}/R(f) = \frac{1}{16}(d^2 + 4d + 10)\Big(\frac{8}{d+5}\Big)^{4/(d+4)} n^{-4/(d+4)}.$$

Tal como para o histograma, a perda de qualidade do estimador com o aumento da dimensão dos dados é evidente. No entanto, por comparação com a Tabela 2.10.2, vemos que essa perda de qualidade é muito mais sentida pelo histograma do que pelo estimador do núcleo.

Sobre a utilização do estimador do núcleo em contextos multidimensionais veja-se Simonoff (1996), Scott e Sain (2005) e Scott (2004).

Bibliografia

Abou-Jaoudé, S. (1976a). Conditions de convergence L_1 en probabilité de l'histogramme pour une densité. *Ann. Inst. H. Poincaré Probab. Statist.* 3, 213–231.

Abou-Jaoudé, S. (1976b). Sur une condition nécessaire et suffisante de L_1-convergence presque complète de l'estimateur de la partition fixe pour une densité. *C. R. Acad. Sci. Paris, Ser. A* 283, 1107–1110.

Abou-Jaoudé, S. (1977). *La convergence L_1 et L_∞ de certains estimateurs d'une densité de probabilité.* Doctorat d'État, Université Paris VI.

Akaike, H. (1954). An approximation to the density function. *Ann. Inst. Statist. Math.* 6, 127–132.

Bartlett, M.S. (1963). Statistical estimation of density functions. *Sankhya Ser. A* 25, 245–254.

Berlinet, A., Devroye, L. (1994). A comparison of kernel density estimates. *Pub. Inst. Stat. Univ. Paris* 38, 3–59.

Bertrand-Retali, M. (1974). Convergence uniforme stochastique d'un estimateur d'une densité de probabilité dans $\mathbb{R}^s$. *C. R. Acad. Sci. Paris, Ser. A* 278, 451–453.

Bertrand-Retali, M. (1978). Convergence uniforme d'un estimateur de la

densité par la méthode du noyau. *Rev. Roumaine Math. Pures Appl.* 13, 361–385.

Bochner, S. (1955). *Harmonic analysis and the theory of probability.* University of California Press.

Bogachev, V.I. (2007). *Measure theory.* Springer.

Bosq, D., Lecoutre, J.-P. (1987). *Théorie de l'estimation fonctionnelle.* Economica.

Bowman, A.W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* 71, 353–360.

Bowman, A.W., Azzalini, A. (1997). *Applied smoothing techniques for data analysis.* Oxford University Press.

Cacoullos, T. (1966). Estimation of a multivariate density. *Ann. Inst. Statist. Math.* 18, 179–189.

Cao, R., Cuevas, A., González Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation. *Comput. Statist. Data Anal.* 17, 153–176.

Cencov, N.N. (1962). Evaluation of an unknown distribution density from observations. *Dokl. Akad. Nauk. SSSR* 147, 45-48.

Chacón, J.E., Duong, T. (2010). Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *TEST* 19, 375–398.

Chacón, J.E., Montanero, J., Nogales, A.G. (2007a). A note on kernel density estimation at a parametric rate *J. Nonparametr. Stat.* 19, 13–21.

Chacón, J.E., Montanero, J., Nogales, A.G. and Pérez, P. (2007b). On the existence and limit behavior of the optimal bandwidth in kernel density estimation. *Statist. Sinica* 17, 289–300.

Cheng, M-Y., Fan, J., Marron, J.S. (1997). On automatic boundary corrections. *Ann. Statist.* 25, 1691–1708.

Chiu, S.-T. (1991). Bandwidth selection for kernel density estimation. *Ann. Statist.* 19, 1883–1905.

Chiu, S.-T. (1992). An automatic bandwidth selector for kernel density estimation. *Biometrika* 4, 771–782.

Chiu, S.-T. (1996). A comparative review of bandwidth selection for kernel density estimation. *Statist. Sinica* 6, 129–145.

Chiu, S.-T. (2000). Boundary adjusted density estimation and bandwidth selection. *Statist. Sinica* 10, 1345–1367.

Cohn, D.L. (1980). *Measure theory.* Birkhäuser.

Cowling, A., Hall, P. (1996). On pseudodata methods for removing boundary effects in kernel density estimation. *J. Roy. Statist. Soc. Ser. B* 58, 551–563.

Deheuvels, P. (1974). Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité. *C. R. Acad. Sci. Paris, Ser. A* 278, 1217–1220.

Deheuvels, P. (1977a). Estimation non paramétrique de la densité par histogrammes généralisés. *Rev. Statist. Appl.* 25, 5–42.

Deheuvels, P. (1977b). Estimation non paramétrique de la densité par histogrammes généralisés II. *Pub. Inst. Stat. Univ. Paris* 22, 1–23.

Deheuvels, P., Hominal, P. (1980). Estimation automatique de la densité. *Rev. Statist. Appl.* 28, 25–55.

Devroye, L. (1983). The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates. *Ann. Statist.* 11, 896–904.

Devroye, L. (1987). *A course in density estimation.* Birkhäuser.

Devroye, L., Györfi, L. (1985). *Nonparametric density estimation: the L_1 view.* Wiley.

Devroye, L., Penrod, C.S. (1984). The consistency of automatic kernel density estimates. *Ann. Statist.* 12, 1231–1249.

Devroye, L., Wagner, T.J. (1979). The L_1 convergence of kernel density estimates. *Ann. Statist.* 7, 1136–1139.

Duong, T., Hazelton, M.L. (2003). Plug-in bandwidth matrices for bivariate kernel density estimation. *J. Nonparametr. Stat.* 15, 17–30.

Duong, T., Hazelton, M.L. (2005). Convergence rates for unconstrained bandwidth matrix selectors in multivariate kernel density estimation *J. Multivariate Anal.* 93, 417–433.

Dvoretzky, A., Kiefer, J., Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* 27, 642–669.

Einmahl, U., Mason, D.M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* 33, 1380–1403.

Epanechnikov, V.A. (1969). Nonparametric estimation of a multivariate probability density. *Theory Probab. Appl.* 14, 153–158.

Fan, J., Gijbels, I. (1997). *Local polynomial modelling and its applications.* Chapman & Hall.

Fan, J., Marron, J.S. (1992). Best possible constant for bandwidth selection *Ann. Statist.* 20, 2057–2070.

Farrel, R.H. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density at a point. *Ann. Math. Statist.* 43, 170–180.

Ferraty, F., Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice.* Springer.

Fix, E., Hodges, J.L. (1989). Discriminatory analysis. Nonparametric estimation: consistency properties *Int. Stat. Rev.* 57, 238–247. O artigo original surge em Report Number 4, Project Number 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.

Freedman, D., Diaconis, P. (1981). On the histogram as a density estimator: L_2 theory. *Z. Wahrsch. Verw. Gebiete* 57, 453–476.

Gasser, T., Müller, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing techniques for curve estimation*, Gasser, T., Rosenblatt, M. (Eds.), Lecture Notes in Mathematics 757, 23–68.

Gasser, Th., Müller, H.-G., Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* 47, 238–252.

Geffroy, J. (1974). Sur l'estimation d'une densité dans un espace métrique. *C. R. Acad. Sci. Paris Sér. A* 278, 1449–1452.

Glad, I.K., Hjort, N.L., Ushakov, N.G. (2003). Correction of density estimators that are not densities. *Scand. J. Statist.* 30, 415–427.

Glick, N. (1974). Consistency conditions for probability estimators and integrals of density estimators. *Util. Math.* 6, 61–74.

Glivenko, V.I. (1934). *A course in probability theory.* Moscow (em Russo).

Hall, P. (1980). Objective methods for the estimation of window size in the nonparametric estimation of a density. Manuscrito não publicado.

Hall, P. (1983). Large sample optimality of least squares cross-validation in density estimation. *Ann. Statist.* 11, 1156–1174.

Hall, P., Marron, J.S. (1987a). Extent to which least-squares cross-validation minimizes integrated square error in nonparametric density estimation. *Probab. Theory Related Fields* 74, 567–581.

Hall, P., Marron, J.S. (1987b). Estimation of integrated squared density derivatives. *Statis. Probab. Lett.* 6, 109–115.

Hall, P., Marron, J.S. (1991). Lower bounds for bandwidth selection in density estimation. *Probab. Theory Related Fields* 90, 149–173.

Hall, P., Marron, J.S., Park, B.U. (1992). Smoothed cross-validation. *Probab. Theory Related Fields* 92, 1–20.

Hall, P., Sheather, S.J., Jones, M.C., Marron, J.S. (1991). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* 78, 263–269.

Härdle, W. (1990). *Applied nonparametric regression.* Cambridge

Härdle, W. (1991). *Smoothing techniques: with implementation in S.* Springer.

Härdle,W., Müller, M., Sperlich, S., Werwatz, A. (2004). *Nonparametric and semiparametric models.* Springer.

Hodges, J.L., Lehmann, E.L. (1956). The efficiency of some nonparametric competitors of the t-test. *Ann. Math. Statist.* 27, 324–335.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* 19, 293–325.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* 58, 13–30.

Jones, M.C. (1993). Simple boundary correction for kernel density estimation. *Stat. Comput.* 3, 135–146.

Jones, M.C., Foster, P.J. (1993). Generalized jacknifing and higher order kernels. *J. Nonparametr. Stat.* 3, 81–94.

Jones, M.C., Foster, P.J. (1996). A simple nonnegative boundary correction method for kernel density estimation. *Statist. Sinica* 6, 1005–1013.

Jones, M.C., Kappenman, R.F. (1992). On a class of kernel density estimate bandwidth selectors. *Scand. J. Statist.* 19, 337–350.

Jones, M.C., Marron, J.S., Park, B.U. (1991). A simple root n bandwidth selector. *Ann. Statist.* 19, 1919–1932.

Jones, M.C., Marron, J.S., Sheather, S.J. (1996a). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.* 91, 401–407.

Jones, M.C., Marron, J.S., Sheather, S.J. (1996b). Progress in data-based bandwidth selection for kernel density estimation. *Comput. Statist.* 11, 337–381.

Jones, M.C., Sheather, S.J. (1991). Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statis. Probab. Lett.* 11, 511–514.

Jones, M.C., Signorini, D.F. (1997). A comparison of higher-order bias kernel density estimators. *J. Amer. Statist. Assoc.* 92, 1063–1073.

Karunamunia, R.J., Zhang, S. (2008). Some improvements on a boundary corrected kernel density estimator. *Statis. Probab. Lett.* 78, 499–507.

Lecoutre, J.-P. (1985). The L_2-optimal cell width for the histogram. *Statis. Probab. Lett.* 3, 303–306.

Lehmann, E.L. (1959). *Testing statistical hypotheses.* Wiley.

Lejeune, M. (1985). Estimation non-paramétrique par noyaux: régression polynomiale mobile. *Rev. Statist. Appl.* 33, 43–67.

Lejeune, M., Sarda, P. (1992). Smooth estimators of distribution and density functions. *Comput. Statist. Data Anal.* 14, 457–471.

Liao, J.G., Wu, Y., Lin, Y. (2010). Improving Sheather and Jones' bandwidth selector for difficult densities in kernel density estimation. *J. Nonparametr. Stat.* 22, 105–114.

Lima, E.L. (1995). *Curso de análise, Vol. 1.* Projecto Euclides.

Loader, C.R. (1996). Local likelihood density estimation. *Ann. Statist.* 24, 1602–1618.

Marron, J.S., Nolan, D. (1989). Canonical kernels for density estimation. *Statis. Probab. Lett.* 7, 191–195.

Marron, J.S., Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *J. Roy. Statist. Soc. Ser. B* 56, 653–671.

Marron, J.S., Wand, M.P. (1992). Exact mean integrated squared error. *Ann. Statist.* 20, 712–736.

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.* 3, 1269–1283.

Mason, D.M., Swanepoel, J.W.H. (2010). A general result on the uniform in bandwidth consistency of kernel-type function estimators. *TEST*.

Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika* 78, 521–530.

Nadaraya, E.A. (1965). On non-parametric estimates of density functions and regression curves. *Theory Probab. Appl.* 10, 186–190.

Nadaraya, E.A. (1974). On the integral mean square error of some non-parametric estimates for the density function. *Theory Probab. Appl.* 19, 133–141.

Nadaraya, E.A. (1989). *Nonparametric estimation of probability densities and regression curves.* Kluwer.

Park, B.U., Marron, J.S. (1990). Comparison of data-driven bandwidth selectors. *J. Amer. Statist. Assoc.* 85, 66–72.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* 33, 1065–1076.

Prakasa Rao, B.L.S. (1983). *Nonparametric functional estimation.* Academic Press.

R Development Core Team (2009). *R: a language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Révész, P. (1972). On empirical density function. *Period. Math. Hungar.* 2, 85–110.

Rice, J. (1984). Boundary modification for kernel regression. *Comm. Statist. Theory Methods* 13, 893–900.

Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. *Ann. Math. Statist.* 27, 832–837.

Rosenblatt, M. (1971). Curve estimates. *Ann. Math. Statist.* 42, 1815–1842.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.* 9, 65–78.

Ruppert, D., Wand, M.P. (1992). Multivariate locally weighted least squares regression. *Ann. Statist.* 22, 1346–1370.

S original by Matt Wand. R port by Brian Ripley. (2009). *KernSmooth: functions for kernel smoothing for Wand & Jones (1995).* URL http://CRAN.R-project.org/package=KernSmooth

Sain, S.R., Baggerly, K.A., Scott, D.W. (1994). Cross-validation of multivariate densities. *J. Amer. Statist. Assoc.* 89, 807–817.

Schuster, E.F. (1969). Estimation of a probability density function and its derivatives. *Ann. Math. Statist.* 40, 1187–1195.

Schuster, E.F. (1970). Note on the uniform convergence of density estimates. *Ann. Math. Statist.* 41, 1347–1348.

Schuster, E.F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Comm. Statist. Theory Methods* 14, 1123–1136.

Scott, D.W. (1979). On optimal and data-based histograms. *Biometrika* 66, 605–610.

Scott, D.W. (1985). Frequency polygons: theory and applications. *J. Amer. Statist. Assoc.* 80, 348–354.

Scott, D.W. (1992). *Multivariate density estimation*. Wiley.

Scott, D.W. (2004). Multivariate density estimation and vizualization. In *Handbook of computational statistics: concepts and methods*, Gentle, J.E., Hardle, W., Mori, Y. (Eds.), Springer, 517–538.

Scott, D.W., Sain, S.R. (2005). Multi-dimensional density estimation. In *Handbook of statistics, Vol. 24: Data mining and computational statistics*, Rao, C.R., Wegman, E.J., Solka, J.L. (Eds.), Elsevier, 229–262.

Scott, D.W., Terrel, G.R. (1987). Biased and unbiased cross-validation in density estimation. *J. Amer. Statist. Assoc.* 82, 1131–1146.

Scott, D.W., Wand, M.P. (1991). Feasibility of multivariate density estimates. *Biometrika* 78, 197–206.

Sheather, S.J. (1983). A data-based algorithm for choosing the window width when estimating the density at a point. *Comput. Statist. Data Anal.* 1, 229–238.

Sheather, S.J. (1986). An improved data-based algorithm for choosing the window width when estimating the density at a point. *Comput. Statist. Data Anal.* 4, 61–65.

Sheather, S.J. (2004). Density estimation. *Statist. Sci.* 19, 588–597.

Sheather, S.J., Jones, M.C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Ser. B* 53, 683–690.

Simonoff, J.S. (1996). *Smoothing methods is statistics.* Springer.

Silverman, B.W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* 6, 177–189.

Silverman, J. (1986). *Density estimation for statistics and data analysis.* Chapman & Hall.

Smirnov, N.V. (1950). On the construction of confidence regions for the density of distribution of random variables. *Doklady Akad. Nauk SSSR* 74, 189–191 (em Russo).

Smirnov, N.V. (1951). On the approximation of probability densities of random variables. *Scholary Notes of Moscow State Polytechnical Institute* 16, 69–96 (em Russo).

Stone, C.J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.* 12, 1285–1297.

Stone, C.J. (1985). An asymptotically optimal histogram selection rule. *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II* (Berkeley, Calif., 1983), 513–520.

Sturges, H.A. (1926). The choice of a class interval. *J. Amer. Statist. Assoc.* 21, 65–66.

Tenreiro, C. (2001). On the asymptotic behaviour of the integrated square error of kernel density estimators with data-dependent bandwidth. *Statis. Probab. Lett.* 53, 283–292.

Tenreiro, C. (2003). On the asymptotic normality of multistage integrated density derivatives kernel estimators. *Statis. Probab. Lett.* 64, 311–322.

Terrel, G.R. (1990). The maximal smoothing principle in density estimation. *J. Amer. Statist. Assoc.* 85, 470–477.

Terrell, G.R., Scott, D.W. (1985). Oversmoothed nonparametric density estimates. *J. Amer. Statist. Assoc.* 80, 209–214.

Thompson, J.R., Tapia, R.A. (1990). *Nonparametric function estimation, modeling, and simulation.* SIAM.

Tiago de Oliveira, J. (1963). Estatística de densidades: resultados assintóticos. *Rev. Fac. Ciên. Lisboa* 9, 111–206.

Tsybakov, A. (2004). *Introduction à l'estimation non-paramétrique.* Springer.

van der Vaart, A.W. (1998). *Asymptotic statistics*. Cambridge.

Van Ryzin, J. (1969). On strong consistency of density estimates. *Ann. Math. Statist.* 40, 1765–1772.

Wagner, T.J. (1975). Nonparametric estimates of probability densities. *IEEE Trans. Inform. Theory* 21, 438–440.

Wand, M.P. (1992). Error analysis for general multivariate kernel estimators. *J. Nonparametr. Stat.* 2, 1–15.

Wand, M.P. (1997). Data-based choice of histogram bin width. *Amer. Statist.* 51, 59–64.

Wand, M.P., Jones, M.C. (1993). Comparison of smoothing parameterizations in bivariate kernel density estimation. *J. Amer. Statist. Assoc.* 88, 520–528.

Wand, M.P., Jones, M.C. (1994). Multivariate plug-in bandwidth selection. *Comput. Statist.* 9, 97–116.

Wand, M.P., Jones, M.C. (1995). *Kernel smoothing.* Chapman & Hall.

Wand, M.P., Schucany, W.R. (1990). Gaussian-based kernels. *Canad. J. Statist.* 18, 197–204.

Wasserman, L. (2006). *All of nonparametric statistics.* Springer.

Watson, G.S., Leadbetter, M.R. (1963). On the estimation of the probability density. I. *Ann. Math. Statist.* 34, 480–491.

Woodroofe, M. (1970). On choosing a delta-sequence. *Ann. Math. Statist.* 41, 1665–1671.

Wu, T.-J. (1995). Adaptive root n estimates of integrated squared density derivatives. *Ann. Statist.* 23, 1474–1495.

Zhang, S., Karunamuni, R.J. (2010). Boundary performance of the beta kernel estimators. *J. Nonparametr. Stat.* 22, 81–104.

Índice Remissivo

$\mathscr{C}$, 5
$\mathscr{C}_{k\alpha}$, 38
convergência
  em média quadrática, 5
  em média quadrática integrada, 5
  em probabilidade, 3
  L_1, 20
  quase certa, 3
  uniforme, 17

desigualdade exponencial, 16, 19
distância
  L_1, 3
  L_2, 3
  L_∞, 3

EQM, 2
EQMI, 3
erro quadrático
  integrado, 33
  médio, 2
  médio integrado, 2
  médio integrado assintótico, 26
estimador
  assintoticamente cêntrico, 11
  automático do núcleo, 94
  da janela móvel, 4, 44
  do histograma, 9, 40
  do núcleo, 43, 98, 99, 101
  do polígono de frequências, 36

$\mathscr{F}$, 20
Fórmula de Taylor, 12
função
  absolutamente contínua, 22
  de distribuição empírica, 4
  de variação limitada, 57
  Lipschitziana de ordem α, 11, 18
  uniformemente contínua, 18

IVAR, 3
IVIES, 3

janela, 4, 7, 10, 44
  assintoticamente óptima, 27, 39, 40, 63, 102

com distribuição de referência normal, 31, 80
de Silverman, 32, 80
de Sturges, 30
pelo método de validação cruzada, 33, 81
pelo princípio da distribuição mais suave, 32, 81

$\mathscr{L}_x(\alpha)$, 11
Lema de Bochner, 49

mudança de escala, 47

$\mathscr{N}$, 67
núcleo
canónico, 72
de Bertrand-Retali–Geffroy, 59
de Epanechnikov, 68
de fronteira, 99
de Parzen–Rosenblatt, 50
momentos, 48

oráculo, 27

$\mathscr{U}$, 18

Var, 2
Viés, 2

Escola Superior de Tecnologia de Viseu
ESTV
Banco de Portugal
Caixa Geral de Depositos
cmuc
Centro de Matemática
Universidade de Coimbra
FCT
Fundação para a Ciência e a Tecnologia
Timberlakeconsultores
Estatística, Matemática e Econometria
Instituto Nacional de Estatística
Statistics Portugal
LIVRARIA ESCOLAR EDITORA
PSE Produtos e Serviços de Estatística, Lda
TERMAS DE S. PEDRO DO SUL
DSD
DÃO SUL