

CARLOS MANUEL REBELO TENREIRO DA CRUZ

Sobre a Estimação Não-Paramétrica  
duma Densidade de Probabilidade



Faculdade de Ciências e Tecnologia  
Universidade de Coimbra

1989

Este trabalho destina-se ao preenchimento de parte dos requisitos necessários à prestação de Provas de Aptidão Pedagógica e Capacidade Científica.

Quero deixar um agradecimento à Dr<sup>a</sup> Maria de Nazaré Mendes Lopes por me ter sugerido o tema, no âmbito do qual se insere este trabalho, bem como, pelo apoio que sempre me deu durante a sua realização e pela leitura atenta e crítica do mesmo.

Agradeço também, a todos os que contribuíram para a realização da parte computacional inerente a este trabalho.

## ÍNDICE

### INTRODUÇÃO

CAPÍTULO I : O ESTIMADOR DO HISTOGRAMA .....	p.1
1.1. Definição do estimador do histograma e primeiras propriedades .....	p.2
1.2. Questões de Mensurabilidade .....	p.5
1.3. Convergência $L_1$ do estimador do histograma .....	p.6
1.4. Convergência uniforme do estimador do histograma .....	p.17
1.5. Escolha assintoticamente ótima de $h_n$ .....	p.25
CAPÍTULO II : O ESTIMADOR DO NÚCLEO .....	p.41
2.1. Primeiras definições e propriedades .....	p.42
2.2. Mais Mensurabilidade .....	p.45
2.3. Convergência $L_1$ do estimador do núcleo .....	p.46
2.4. Convergência uniforme do estimador do núcleo .....	p.53
2.5. Escolha assintoticamente ótima de $h_n$ e de $K$ .....	p.67
CAPÍTULO III : O ESTIMADOR AUTOMÁTICO DO NÚCLEO	
BREVE ANÁLISE .....	p.81
3.1. Escolha "paramétrica" de $h_n$ .....	p.83
3.2. Método do tipo "máximo de verosimilhança" para a escolha de $h_n$ .....	p.85
3.3. Minimização do EQI como critério para a escolha de $h_n$ .....	p.86
3.4. Escolha de $h_n$ "segundo Kappenmann" .....	p.88
3.5. Estudo de simulação. Resultados. ....	p.88
BIBLIOGRAFIA .....	p.94

## INTRODUÇÃO

Nas últimas três décadas surgiu uma vasta literatura no domínio da estimação não paramétrica da densidade de probabilidade. Esta avalanche de investigação foi em grande parte provocada por um artigo de Rosenblatt em 1956 que começa por considerar para estimar  $f(y)$ , com  $y \in \mathbb{R}$ , a função mensurável da amostra

$$\frac{F_n(y+h_n) - F_n(y-h_n)}{2h_n}$$

onde  $F_n$  é a função de repartição empírica associada à  $n$ -amostra  $(X_1, X_2, \dots, X_n)$  de lei absolutamente contínua de densidade  $f$  e  $h_n > 0$ ,  $n \in \mathbb{N}$ . No fim deste artigo Rosenblatt propõe uma classe de estimadores mais geral que tomaria a forma

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

onde  $\lim_n h_n = 0$  e  $\int_{\mathbb{R}} K(y) dy = 1$ . Este estimador passaria a ser conhecido por estimador do núcleo.

É realmente à custa desta ideia que nos anos que se seguiram surgem diversos trabalhos onde são estudadas as propriedades desta classe de estimadores das quais realçamos as convergências em média quadrática, média quadrática integrada, uniforme e  $L_1$ . Como principais artigos nestas áreas podemos citar o importante trabalho de Parzen em 1962 e os trabalhos de Cacoullos (1966), Nadaraya ((1965) e (1974)), Epanechnikov (1968), Bertrand-Retali (1978), Abou-Jaoude (1977) e Devroye (1983).

É no ano de 1972 que P. Révész aborda o estudo do primeiro estimador que consideraremos neste trabalho : o estimador do histograma.

Através do histograma, cujo aparecimento data de finais do século passado, surge um método usado pelos estatísticos para obterem uma informação global sobre a amostra extraída e desta informação serem inferidas algumas características da lei da população de onde se extraiu a amostra. Com o passar dos anos o histograma começou a ser apresentado

não só com ideia anterior mas também como uma aproximação da densidade de probabilidade da lei da população (supondo que esta existe).

Apesar do seu longo historial é realmente só em 1972 que são estudadas as propriedades do histograma como estimador da densidade de probabilidade, tendo sido obtidas propriedades de convergência que justificam a atitude empírica de tomar as suas estimativas como aproximações para a densidade de probabilidade. No caso real ele é definido do seguinte modo: Tomemos  $\dots < x_{-2}(n) < x_{-1}(n) < x_0(n) < x_1(n) < x_2(n) < \dots$  uma sucessão de partições de  $\mathbb{R}$  tal que  $x_{i+1}(n) - x_i(n) = h_n$  ( $i \in \mathbb{Z}$ ) onde  $h_n$  é uma sucessão de números reais convergindo para zero. Definimos o estimador do histograma (associado a esta sucessão de partições) por

$$f_n(y) = \frac{1}{nh_n} \sum_{j=1}^n \mathbb{1}_{[x_i(n), x_{i+1}(n)]}(X_j), \text{ para todo o } y \in [x_i(n), x_{i+1}(n)], \text{ com } i \in \mathbb{Z},$$

onde  $(X_1, X_2, \dots, X_n)$  é uma  $n$ -amostra de lei absolutamente contínua.

Os artigos mais importantes dentro das áreas de convergência que já referimos para o estimador do núcleo são o já citado artigo de Révész e ainda os artigos de Bertrand-Retali e Geffroy em 1974, Abou-Jaoude ((1976) e (1976a)), Scott (1979) e Lecoutre (1985).

É sobre estes dois tipos de estimadores da densidade, mas definidos em espaços mais gerais, que nos iremos debruçar ao longo deste trabalho. Ocupar-nos-emos essencialmente do estudo da convergência uniforme e  $L_1$  dos mesmos.

A convergência uniforme é historicamente dos primeiros tipos de convergência a ser analisado e o seu interesse reside na sua forte interpretação geométrica (quer local, quer global) apesar de não possuir significação probabilística directa.

A convergência  $L_1$  possui a grande vantagem de trabalhar com funções para as quais basta exigir a integrabilidade, sendo esta propriedade "natural" para as densidades de probabilidade. A sua interpretação geométrica é clara apesar de não nos fornecer informação acerca da aproximação local entre a densidade e uma sua estimativa. A nível probabilístico a sua importância é realçada por um resultado de H. Scheffé em 1947 (ver Devroye, 1984a, p.2): Sendo  $f$  e  $g$  densidades de probabilidade sobre  $(\mathcal{X}, \mathcal{C}, \mu)$  então

$$\int_{\mathcal{X}} |f(x) - g(x)| \mu(dx) = \sup_{A \in \mathcal{C}} |P_f(A) - P_g(A)|.$$

Um dos critérios mais usados, pelo seu fácil manuseamento, para quantificar a discrepância global entre um estimador da densidade e a verdadeira densidade, supondo esta de quadrado integrável, é o erro quadrático médio integrado, ou seja, no caso real,  $\int_{\mathbb{R}} E(f_n(y) - f(y))^2 dy$ . Este critério é também aplicado neste trabalho ao estudo da escolha dos parâmetros que interferem num e noutro dos estimadores e também para a obtenção das ordens de

convergência de cada um deles ( a escolha dos parâmetros dos estimadores no que refere à sucessão  $(h_n)$  pode ser encontrada, utilizando outro critério de escolha, em Devroye (1984a)).

Verificaremos nas partes finais dos 1º e 2º capítulos que a sucessão  $(h_n)$  escolhida de modo a minimizar assintoticamente o erro quadrático médio integrado é função da densidade  $f$  que desconhecemos. Devido a este facto surge nos últimos anos uma vasta literatura que aborda o problema da escolha da sucessão  $h_n$ , escolha essa que se pretende seja definida pela amostra obtida e ainda que o estimador definido à custa dessa sucessão apresente boas propriedades quer de consistência, quer de eficiência ( $n$  considerado fixo). Dos trabalhos que abordam este tema realçamos os de Wagner (1975), Deheuvels (1980) e Devroye (1984).

Vamos de seguida referir mais em pormenor o percurso tomado neste trabalho, o qual dividimos em três capítulos. O primeiro é dedicado ao estudo do estimador do histograma e o segundo ao estudo do estimador do núcleo cujas definições já aqui apresentámos no caso real. O último capítulo vamos dedicá-lo ao estimador automático do núcleo, que, em termos genéricos, vai diferir do estimador do núcleo pelo facto de  $h_n$  ser uma estatística.

Os trajectos seguidos na elaboração dos dois primeiros capítulos possuem linhas comuns. Depois das definições dos estimadores, focaremos propriedades locais dos mesmos, tais como a convergência assintótica cêntrica e a convergência em média quadrática. Tais propriedades servirão para nos alertar acerca da importância que vai desempenhar na estimação, uma conveniente escolha da sucessão  $h_n$ .

O estudo da convergência  $L_1$  e uniforme dos dois estimadores é o passo que é dado a seguir. Os resultados de Geffroy(1974), Abou-Jaoude(1977), Bertrand-Retali(1978) e Devroye(1983) são apresentados em pormenor.

Os últimos parágrafos de ambos os capítulos são preenchidos com a escolha assintoticamente óptima da sucessão  $h_n$ . No caso do estimador do núcleo abordaremos também a escolha do núcleo  $K$ . Os capítulos terminam com a aplicação dos métodos a amostras simuladas.

No último capítulo, não querendo fazer um estudo pormenorizado sobre o estimador automático do núcleo, julgamos no entanto, ser indispensável a sua referência num trabalho sobre estimação não-paramétrica da densidade de probabilidade. Abordaremos quatro tipos de escolhas possíveis para a sucessão  $h_n$  e referiremos os principais resultados conhecidos para os estimadores resultantes das referidas escolhas.

O principal objectivo deste capítulo é o de realçar o muito que ainda pode ser feito quer nos estimadores que vamos considerar, quer na criação de novos estimadores (novas possibilidades de escolha de  $h_n$ ) que para além das propriedades de consistência, possuam também boas propriedades de eficiência. A dualidade consistência/eficiência é posta em evidência apresentando-se casos de estimadores com boas propriedades de consistência mas com um fraco rendimento a nível de eficiência, havendo outros com boas prestações a nível de eficiência não havendo contudo resultados que garantam a sua consistência.

No início de cada capítulo é dada, em detalhe, informação acerca dos resultados que focaremos, bem como as respectivas referências bibliográficas.

Neste trabalho efectuaremos, como já referimos, alguns estudos de simulação onde necessitamos de gerar a nível computacional amostras de algumas distribuições de probabilidade. As ideias para os métodos que utilizámos para tal, podem ser encontrados em Gmurman (1977, p.372-379).



## **CAPÍTULO I : O ESTIMADOR DO HISTOGRAMA**

Este primeiro capítulo é dedicado ao estudo do estimador do histograma e vai centrar-se principalmente na obtenção dos resultados fundamentais acerca da convergência  $L_1$  e uniforme deste estimador que são apresentadas no Teorema 1.A do parágrafo 1.3., devido a Abou-Jaoude ((1976) e (1976a)) e no Teorema 1.C, do parágrafo 1.4., de Geffroy (1974).

Abordaremos também a convergência em média quadrática e a convergência em média quadrática integrada que terão como principal finalidade a análise do problema da escolha do parâmetro que interfere neste estimador e cujo resultado mais importante, que apresentaremos no parágrafo 1.5., é de Lecoutre (1985).

O estudo da convergência  $L_1$  e uniforme do histograma justificam-se só por si, devido à grande importância histórica e popularidade deste estimador, mas terá, no entanto aqui um duplo interesse pois será também importante no estudo da convergência  $L_1$  e uniforme do estimador do núcleo que faremos no capítulo seguinte.

## 1.1. Definição do estimador do histograma e primeiras propriedades

Seja  $(\mathfrak{X}, \mathfrak{C}, \mu)$  um espaço mensurado. Suponhamos que existe uma sucessão  $(\mathfrak{P}_n)_{n \in \mathbb{N}}$  de partições de  $\mathfrak{X}$ ,

$$\mathfrak{P}_n = \{ \Delta_{n,r} \mid r \in R_n \}$$

onde  $R_n \subset \mathbb{N}$  e  $\Delta_{n,r} \in \mathfrak{C}$  para todo  $n \in \mathbb{N}$  e  $r \in R_n$ .

Admitamos ainda que para  $n$  suficientemente grande (s.g.)  $0 < \mu(\Delta_{n,r}) < \infty$ , para todo  $r \in R_n$ .

Sejam  $P_f$  uma lei de probabilidade sobre  $(\mathfrak{X}, \mathfrak{C})$  dominada por  $\mu$ , admitindo  $f$  como versão da densidade de probabilidade e  $X^n = (X_1, X_2, \dots, X_n)$ , com  $n \in \mathbb{N}$ , uma  $n$ -amostra de lei  $P_f$ .

Definimos o estimador do histograma de  $f$  associado à partição  $\mathfrak{P}_n$  de  $\mathfrak{X}$  por:

$$\hat{f}_n(x; X^n) = \hat{f}_n(x) = \sum_{r \in R_n} \frac{\nu_n(\Delta_{n,r})}{n\mu(\Delta_{n,r})} \mathbb{1}_{\Delta_{n,r}}(x), \text{ para } x \in \mathfrak{X},$$

onde  $v_n(B) = \sum_{i=1}^n \mathbb{1}_B(X_i)$ , para  $B \in \mathcal{A}$ .

Vamos designar por  $\mathcal{F}$  o conjunto das densidades de probabilidade sobre  $(\mathcal{X}, \mathcal{A}, \mu)$ .

Imediatamente se conclui que para todo o  $x \in \mathcal{X}^n$ , com  $n \in \mathbb{N}$ ,  $\hat{f}_n \in \mathcal{F}$ .

Ao longo deste capítulo vamos por diversas vezes analisar o importante caso particular em que  $\mathcal{X} = \mathbb{R}^s$  ( $s \in \mathbb{N}$ ),  $\mathcal{A} = \mathcal{B}_{\mathbb{R}^s}$  (tribo dos borelianos) e  $\mu = \lambda$  (medida de Lebesgue).

Quando tal acontecer consideramos definida em  $\mathbb{R}^s$  a seguinte sucessão de partições de  $\mathbb{R}^s$  em borelianos :

$$\mathcal{P}_n = (\Delta_{n, q_1, q_2, \dots, q_s} \mid (q_1, q_2, \dots, q_s) \in \mathbb{Z}^s), n \in \mathbb{N}, \text{ onde } \Delta_{n, q_1, q_2, \dots, q_s} = \prod_{i=1}^s [q_i h_n, (q_i+1)h_n[$$

com  $h_n$  número real positivo.

Vamos apresentar desde já, propriedades locais importantes do estimador do histograma neste caso particular.

Seja  $x \in \mathcal{X}$ . Denotemos por  $\mathcal{F}\mathcal{C}_x$  o conjunto das densidades de probabilidade contínuas em  $x$ .

As proposições 1.1.1 e 1.1.2, estabelecem-nos, respectivamente, condições necessárias e suficientes (n/s) para que para todo o ponto  $x \in \mathbb{R}^s$  e toda a densidade de probabilidade  $f \in \mathcal{F}\mathcal{C}_x$ , o estimador do histograma seja assintoticamente centrado e consistente em média quadrática.

A demonstração da proposição 1.1.2 permite-nos ainda concluir que (quando da aplicação prática do histograma) a escolha de  $h_n$  "grande" implica um viés,  $E[\hat{f}_n(x)] - f(x)$ , "grande". Por outro lado, uma escolha de  $h_n$  "pequeno" irá implicar uma variância,  $V[\hat{f}_n(x)]$ , "grande". Assim escolhas "desequilibradas" de  $h_n$  contribuem para uma má aproximação local no sentido do erro quadrático médio.

**Proposição 1.1.1:**

$$(\forall x \in \mathbb{R}^s \forall f \in \mathcal{F}\mathcal{C}_x \lim_n E[\hat{f}_n(x)] = f(x)) \Leftrightarrow \lim_n h_n = 0$$

*dem:*

Seja  $x \in \mathbb{R}^s$  e  $n \in \mathbb{N}$ , existe  $(q_1, q_2, \dots, q_s) \in \mathbb{Z}^s$  tal que  $x \in \Delta_{n, q_1, q_2, \dots, q_s}$ .

$$\text{Assim } E[\hat{f}_n(x)] = \frac{1}{h_n^s} \int_{\Delta_{n, q_1, q_2, \dots, q_s}} f(t) dt \text{ e portanto}$$

$$|E[\hat{f}_n(x)] - f(x)| \leq \frac{1}{h_n^s} \int_{\Delta_{n,q_1,q_2,\dots,q_s}} |f(t) - f(x)| dt.$$

Seja  $\varepsilon > 0$ . Sendo  $f$  contínua em  $x$  existe  $\delta > 0$  tal que se  $\|x - t\| < \delta$  então  $|f(x) - f(t)| < \varepsilon$ .

Por hipótese existe uma ordem  $n_0 \in \mathbb{N}$  a partir da qual  $\text{diam}(\Delta_{n,q_1,q_2,\dots,q_s}) < \delta$ , logo para  $n \geq n_0$ ,  $|E[\hat{f}_n(x)] - f(x)| < \varepsilon$ .

Para provarmos a implicação recíproca suponhamos que  $h_n$  possui limite finito ou infinito (se tal não acontecer consideramos  $h_{n_k}$  subsucessão de  $h_n$  que tenha limite finito ou infinito).

$$\text{Consideremos } f(x_1, \dots, x_s) = \frac{1}{\prod_{i=1}^s} \frac{1}{1+x_i^2}.$$

$$\text{Se } h_n \rightarrow \alpha \neq 0, n \rightarrow \infty, \text{ então } \lim_n E[\hat{f}_n(0, \dots, 0)] = \lim_n \frac{1}{(\prod h_n)^s} \prod_{i=1}^s \int_0^{h_n} \frac{1}{1+x_i^2} dx_i \\ = \left(\frac{\arctg \alpha}{\alpha \pi}\right)^s \neq f(0, \dots, 0).$$

Se  $h_n \rightarrow \infty, n \rightarrow \infty$ , por raciocínio análogo  $\lim_n E[\hat{f}_n(0, \dots, 0)] \neq f(0, \dots, 0)$ .  $\square$

**Proposição 1.1.2:**

$$(\forall x \in \mathbb{R}^s \forall f \in \mathcal{F} \mathcal{C}_x \lim_n E[\hat{f}_n(x) - f(x)]^2 = 0) \Leftrightarrow (\lim_n h_n = 0, \lim_n n h_n^s = \infty)$$

*dem:*

Sejam  $x \in \mathbb{R}^s, n \in \mathbb{N}$  e  $f \in \mathcal{F} \mathcal{C}_x$  então

$$E[\hat{f}_n(x) - f(x)]^2 = (E[\hat{f}_n(x)] - f(x))^2 + \frac{E[\hat{f}_n(x)]}{n h_n^s} (1 - h_n^s E[\hat{f}_n(x)]) \quad (1.1.1).$$

Sendo por hipótese  $\lim_n h_n = 0$ , basta provar, pela proposição 1.1.1, que

$$\frac{E[\hat{f}_n(x)]}{n h_n^s} (1 - h_n^s E[\hat{f}_n(x)]) \rightarrow 0, n \rightarrow \infty$$

o que é verdade pois  $\lim_n n h_n^s = \infty$ .

Reciprocamente, se para todo o  $x \in \mathbb{R}^s$  e  $f \in \mathcal{F} \mathcal{C}_x$   $\lim_n E[\hat{f}_n(x) - f(x)]^2 = 0$ , vem, pela proposição anterior, por (1.1.1) e tendo-se  $h_n^s E[\hat{f}_n(x)] \leq 1$ , que  $\lim_n h_n = 0$ . Como por (1.1.1)

$$\lim_n \frac{E[\hat{f}_n(x)]}{nh_n^s} (1 - h_n^s E[\hat{f}_n(x)]) = 0 \text{ então } \lim_n nh_n^s = \infty. \square$$

## 1.2. Questões de Mensurabilidade

Nos próximos parágrafos vamos abordar o problema da convergência estocástica, do ponto de vista uniforme e  $L_1$ , do estimador do histograma. Seja então  $\mathfrak{M}$ , um modo de convergência estocástica (convergência em probabilidade (p.), quase certa (q.c.) ou quase completa (q.co.)).

Pretendemos determinar condições para que  $\int_{\mathfrak{X}} |\hat{f}_n(x) - f(x)| \mu(dx) \xrightarrow{0, n \rightarrow \infty, \mathfrak{M}}$

e  $\sup_{x \in \mathfrak{X}} |\hat{f}_n(x) - f(x)| \xrightarrow{0, n \rightarrow \infty, \mathfrak{M}}$ , para  $f$  pertencente, em cada um dos casos, a

subconjuntos convenientes de  $\mathcal{F}$ .

Vamos denotar por  $\mathcal{F} \mathcal{C}(\mathfrak{X})$  o conjunto das densidades de probabilidade contínuas em  $\mathfrak{X}$ .

Para as expressões anteriores terem significado têm as aplicações

$$(\mathfrak{X}^n, \mathcal{Q}^n, \otimes_{i=1}^n P_f) \longrightarrow (\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+})$$

$$\mathcal{B}^* : (x_1, x_2, \dots, x_n) \longrightarrow \int_{\mathfrak{X}} |\hat{f}_n(x) - f(x)| \mu(dx), \text{ com } f \in \mathcal{F}$$

$$\mathcal{B}^{\wedge} : (x_1, x_2, \dots, x_n) \longrightarrow \sup_{x \in \mathfrak{X}} |\hat{f}_n(x) - f(x)|, \text{ com } f \in \mathcal{F} \mathcal{C}(\mathfrak{X})$$

de ser mensuráveis para todo o  $n \in \mathbb{N}$  (ver parágrafo 1.4. sobre a restrição de  $f \in \mathcal{F} \mathcal{C}(\mathfrak{X})$  no caso da aplicação  $\mathcal{B}^{\wedge}$ ).

A mensurabilidade de  $\mathcal{B}^*$  resulta de  $\mu$  ser  $\sigma$ -finita, da aplicação  $(x; x^n) \longrightarrow |\hat{f}_n(x; x^n) - f(x)| =$

$$\left| \sum_{r \in \mathbb{R}_n} \sum_{i=1}^n \frac{1}{n \mu(\Delta_{n,r})} \mathbb{1}_{\Delta_{n,r}}(x) \mathbb{1}_{\Delta_{n,r}}(x_i) - f(x) \right| \text{ ser } \mathcal{Q}^{n+1}\text{-mensurável e de uma das$$

conclusões do teorema de Fubini.

Para a mensurabilidade de  $\hat{f}$  basta notar que sendo  $\hat{f}_{n,r} = \frac{v_n(\Delta_{n,r})}{n\mu(\Delta_{n,r})}$ ,

$$\{x^n \in \mathfrak{X}^n \mid \sup_{x \in \mathfrak{X}} |\hat{f}_n(x) - f(x)| < \alpha\} =$$

$$= \bigcap_{r \in \mathbb{R}_n} (\{x^n \in \mathfrak{X}^n \mid \hat{f}_{n,r} < \alpha + \inf_{x \in \Delta_{n,r}} f(x)\} \cap \{x^n \in \mathfrak{X}^n \mid \hat{f}_{n,r} > -\alpha + \sup_{x \in \Delta_{n,r}} f(x)\})$$

para  $\alpha > 0$  e ter em conta que para todo o  $x \in \mathfrak{X}$ ,  $\hat{f}_n(x)$  é  $\mathbb{Q}^n$ -mensurável.

### 1.3. Convergência $L_1$ do estimador do histograma

Este parágrafo é dedicado ao estudo da convergência  $L_1$  do estimador do histograma, ou seja, pretendemos determinar condições necessárias ou suficientes para que se tenha

$$\forall f \in \mathfrak{F}_0 \int_{\mathfrak{X}} |\hat{f}_n(x) - f(x)| \mu(dx) \longrightarrow 0, n \rightarrow \infty, \mathfrak{M}. \quad (1.3.1)$$

sendo  $\mathfrak{F}_0$  um subconjunto conveniente de  $\mathfrak{F}$ .

As condições que procuramos podem ser de diversos tipos : condições sobre o espaço  $(\mathfrak{X}, \mathbb{Q})$ , sobre a sucessão de partições  $(\mathcal{P}_n)_{n \in \mathbb{N}}$  ou por exemplo sobre  $\mathfrak{F}_0$  (naturalmente que nos vai interessar que (1.3.1) se verifique para um conjunto, o mais lato possível de densidades de probabilidade). A proposição seguinte explicita-nos um possível caminho para a determinação dessas condições :

**Proposição 1.3.1 :** Se para  $f \in \mathfrak{F}$   $\int_{\mathfrak{X}} |\hat{f}_n(x) - f(x)| \mu(dx) \longrightarrow 0, n \rightarrow \infty, p.$  então

$$\int_{\mathfrak{X}} |E[\hat{f}_n(x)] - f(x)| \mu(dx) \longrightarrow 0, n \rightarrow \infty.$$

**dem :**

Seja a variável aleatória (v.a.)  $\int_{\mathfrak{X}} |\hat{f}_n(x) - f(x)| \mu(dx)$  limitada então por hipótese

$$E\left(\int_{\mathfrak{X}} |\hat{f}_n(x) - f(x)| \mu(dx)\right) \longrightarrow 0, n \rightarrow \infty. \text{ Mas } \int_{\mathfrak{X}} |E[\hat{f}_n(x)] - f(x)| \mu(dx) \leq$$

$$\leq E\left(\int_{\mathfrak{X}} |\hat{f}_n(x) - f(x)| \mu(dx)\right), \square$$

Da proposição anterior concluímos então que :

$$\Delta_n \rightarrow 0, n \rightarrow \infty, \mathfrak{M}, \text{ sse } \begin{cases} \Delta_{n,1} \rightarrow 0, n \rightarrow \infty \\ \Delta_{n,2} \rightarrow 0, n \rightarrow \infty, \mathfrak{M}. \end{cases} \quad (1.3.2)$$

$$\text{onde } \Delta_n = \int_{\mathfrak{X}} |\hat{f}_n(x) - f(x)| \mu(dx), \Delta_{n,1} = \int_{\mathfrak{X}} |E[\hat{f}_n(x)] - f(x)| \mu(dx) \text{ e } \Delta_{n,2} = \int_{\mathfrak{X}} |\hat{f}_n(x) - E[\hat{f}_n(x)]| \mu(dx).$$

Introduzamos as seguintes notações :

$$\mathfrak{B}_n = \text{tribo gerada por } (\mathfrak{P}_n), n \in \mathbb{N}$$

$$\mathfrak{B}'_n = \text{tribo gerada por } \bigcup_{m \geq n} \mathfrak{P}_m, n \in \mathbb{N}$$

$$\mathfrak{B} = \bigcap_{n \in \mathbb{N}} \mathfrak{B}'_n.$$

Se (1.3.1) se verifica então pela proposição 1.3.1,  $E[\hat{f}_n] \rightarrow f, n \rightarrow \infty, L_1$  (convergência  $L_1$ ), o que permite concluir que  $f$  é  $\mathfrak{B}^\wedge$ -mensurável ( $\mathfrak{B}^\wedge$  designa o complemento de  $\mathfrak{B}$  relativamente a  $\mu$ ). Assim é visto existir uma subsucessão  $E[\hat{f}_{n_j}]$  de  $E[\hat{f}_n]$  que converge  $\mu$  q.c.

para  $f$  e portanto dado  $\alpha \in \mathbb{R}^+$ ,

$$\{x \in \mathfrak{X} \mid f(x) > \alpha\} = \{x \mid \lim_{n_j} E[\hat{f}_{n_j}(x)] > \alpha\} \cup \{x \mid \lim_{n_j} E[\hat{f}_{n_j}(x)] \neq f(x), f(x) > \alpha\}$$

onde  $\mu(\{x \mid \lim_{n_j} E[\hat{f}_{n_j}(x)] \neq f(x)\}) = 0$  e  $\{x \mid \lim_{n_j} E[\hat{f}_{n_j}(x)] > \alpha\} \in \mathfrak{B}$ .

Concluímos então que para o problema que vamos abordar estar bem formulado (ver (1.3.1)) teremos de exigir que sendo  $f \in \mathfrak{F}_0$ , logo  $\mathcal{C}$ -mensurável,  $f$  seja também  $\mathfrak{B}^\wedge$ -mensurável. Como o ideal será termos, para todo o  $f \in \mathfrak{F}, \Delta_n \rightarrow 0, n \rightarrow \infty, \mathfrak{M}$ , vamos supor a par-

tir deste momento que sendo  $f \in \mathcal{F}$ ,  $f$  seja  $\mathcal{B}^\wedge$ -mensurável, ou equivalentemente que  $\mathcal{A}^\wedge = \mathcal{B}^\wedge$ .

Seja  $\mathcal{X} = \mathbb{R}^s$ , a seguinte proposição dá-nos condições para que  $\mathcal{B} = \mathcal{B}_{\mathbb{R}^s}$ .

**Proposição 1.3.2:** Uma condição n/s para que  $\mathcal{B} = \mathcal{B}_{\mathbb{R}^s}$  é que  $\lim_{n \rightarrow \infty} h_n = 0$ .

*dem:*

A condição necessária é de demonstração trivial. Para provarmos a implicação recíproca tomemos para  $m \in \mathbb{N}$  e  $(x_1, x_2, \dots, x_s) \in \mathbb{R}^s$

$$A_m^{(x_1, x_2, \dots, x_s)} = \bigcup_{\substack{(q_1, q_2, \dots, q_s) \in \mathbb{Z}^s \\ q_i h_m \leq x_i, i=1, \dots, s}} \prod_{i=1}^s [(q_i - 1)h_m, q_i h_m] \in \mathcal{B}_m.$$

Por hipótese existe uma subsucessão  $h_{n_k}$  de  $h_n$  tal que  $\lim_{k \rightarrow \infty} h_{n_k} = 0$ , e portanto

$$\text{para todo } k \in \mathbb{N} \quad \bigcup_{m \geq k} A_m^{(x_1, x_2, \dots, x_s)} = \prod_{i=1}^s ]-\infty, x_i[.$$

Assim  $\prod_{i=1}^s ]-\infty, x_i[ \in \mathcal{B}$ . A conclusão é agora imediata tendo em conta que  $(\prod_{i=1}^s ]-\infty, x_i[$

$(x_1, x_2, \dots, x_s) \in \mathbb{R}^s$  gera  $\mathcal{B}_{\mathbb{R}^s}$ .  $\square$

Tendo em conta (1.3.2) vamos começar por determinar condições n/s para que  $\Delta_{n,1}^\wedge \rightarrow 0, n \rightarrow \infty$ .

**Definição 1.3.1:** Dizemos que  $(\mathcal{P}_n)_{n \in \mathbb{N}}$  satisfaz a condição (a) se:

$$\forall A \in \mathcal{A} (0 < \mu(A) < \infty) \quad \forall \varepsilon > 0 \quad \exists n_0 \in \mathbb{N} : \forall n \geq n_0 \quad \exists A_n \in \mathcal{B}_n : \mu(A \Delta A_n) < \varepsilon$$

No caso de  $\mathcal{X} = \mathbb{R}^s$  esta condição pode ser formulada de um modo simples que é dado pela proposição seguinte

**Proposição 1.3.3:**  $(\mathcal{P}_n)_{n \in \mathbb{N}}$  satisfaz a condição (a) sse  $\lim_{n \rightarrow \infty} h_n = 0$ .

*dem:*

Suponhamos que existe uma subsucessão de  $h_n$ , convergente para  $\alpha \neq 0$ . Assim existe um subconjunto infinito  $N_0$  de  $\mathbb{N}$ , tal que  $|h_n - \alpha| < \frac{\alpha}{3}$  para  $n \in N_0$ . Consideremos  $A = ]0, \frac{\alpha}{2}[^s$ . Para  $n \in N_0$  e  $A_n \in \mathcal{B}_n$  temos que  $\lambda(A \Delta A_n) \geq \lambda(]0, \frac{\alpha}{2}[^s \Delta ]0, h_n[^s) \geq (\frac{\alpha}{6})^s$ .

Se existir uma subsucessão de  $h_n$  que tende para  $+\infty$ , então existe um subconjunto infini-



to  $N_1$  de  $\mathbb{N}$ , tal que  $h_n > 2$  para  $n \in N_1$ . Consideremos  $A = [0, 1]^S$ . Para  $n \in N_1$  e  $A_n \in \mathcal{B}_n$  temos que  $\lambda(A \Delta A_n) \geq 1$ .

Num e noutro caso  $h_n$  não satisfaz (a).

Suponhamos agora que  $h_n \rightarrow 0, n \rightarrow \infty$ .

Basta-nos estabelecer a condição (a) para todo o aberto limitado de  $\mathbb{R}^S$ . De facto, sendo  $A \in \mathcal{B}_{\mathbb{R}^S}$  tal que  $0 < \lambda(A) < \infty$  e  $A$  limitado, existem abertos limitados  $O_i$  que contêm  $A$  tais que  $\lambda(O_i) \downarrow \lambda(A), i \rightarrow \infty$ .

Seja então  $A$  um aberto limitado de  $\mathbb{R}^S$ . Para  $\varepsilon > 0$ , podemos sempre determinar um conjunto finito de ladrilhos  $\prod_{i=1}^s [a_i^j, b_i^j], j=1, \dots, k$ , tais que  $\lambda(A - \bigcup_{j=1}^k \prod_{i=1}^s [a_i^j, b_i^j]) < \varepsilon$  e

portanto basta estabelecer a condição (a) para conjuntos da forma  $\prod_{i=1}^s [a_i, b_i]$ , com  $a_i, b_i \in \mathbb{R}$ .

Como  $h_n \rightarrow 0, n \rightarrow \infty$ , a propriedade é neste caso trivialmente verificada.  $\square$

Vamos denotar por  $\mathcal{F}'$  o subconjunto de  $\mathcal{F}$  das densidades uniformes sobre elementos  $C \in \mathcal{C}$  com  $0 < \mu(C) < \infty$ .

Para a demonstração dos resultados seguintes tomemos para  $A \in \mathcal{C}$  e  $n \in \mathbb{N}$ :

$$J_n = \{ r \in \mathbb{R}_n \mid \mu(\Delta_{n,r} \cap A) \neq 0, \mu(\Delta_{n,r} \cap A^c) \neq 0 \}$$

$$\alpha_{n,r} = \mu(\Delta_{n,r} \cap A), \beta_{n,r} = \mu(\Delta_{n,r} \cap A^c)$$

$$J_n^{(1)} = \{ r \in J_n \mid \beta_{n,r} \leq \alpha_{n,r} \}$$

$$J_n^{(2)} = J_n - J_n^{(1)}$$

**Teorema 1.3.1:** Se para todo o  $f \in \mathcal{F}'$ ,  $\Delta_{n,1} \rightarrow 0, n \rightarrow \infty$  então  $(\mathcal{P}_n)_{n \in \mathbb{N}}$  verifica (a).

*dem:*

Suponhamos que  $(\mathcal{P}_n)$  não verifica (a). Existirão portanto  $A \in \mathcal{C}, 0 < \mu(A) < \infty, \varepsilon > 0$  e um subconjunto infinito  $N_0 \subset \mathbb{N}$  tais que:  $\forall n \in N_0 \forall A_n \in \mathcal{B}_n \mu(A \Delta A_n) \geq \varepsilon$  (1.3.3).

Tomando  $f = \frac{1}{\mu(A)} \mathbb{1}_A$ , temos que  $\Delta_{n,1} = \sum_{r \in \mathbb{R}_n} \int_{\Delta_{n,r}} \left| \frac{P_f(\Delta_{n,r})}{\mu(\Delta_{n,r})} - \frac{1}{\mu(A)} \mathbb{1}_A(x) \right| \mu(dx)$

$$= \frac{1}{\mu(A)} \sum_{r \in \mathbb{R}_n: \alpha_{n,r} \neq 0} \left( \int_{\Delta_{n,r} \cap A} (\mathbb{1}_A(x) - \frac{\mu(\Delta_{n,r} \cap A)}{\mu(\Delta_{n,r})}) \mu(dx) + \int_{\Delta_{n,r} \cap A^c} \frac{\mu(\Delta_{n,r} \cap A)}{\mu(\Delta_{n,r})} \mu(dx) \right).$$

Como  $\mu(\Delta_{n,r}) = \alpha_{n,r} + \beta_{n,r}$  então  $\Delta_{n,1} = \frac{1}{\mu(A)} \sum_{r \in \mathbb{J}_n} \frac{2\alpha_{n,r}\beta_{n,r}}{\alpha_{n,r} + \beta_{n,r}}$ .

Assim  $\Delta_{n,1} \geq \frac{1}{\mu(A)} \left( \sum_{r \in \mathbb{J}_n^{(1)}} \beta_{n,r} + \sum_{r \in \mathbb{J}_n^{(2)}} \alpha_{n,r} \right)$ . Como por hipótese existe uma ordem

$n_0 \in \mathbb{N}$ , a partir da qual  $\Delta_{n,1} \leq \frac{\varepsilon}{\mu(A)}$ , temos para  $n \geq n_0$ ,  $\sum_{r \in \mathbb{J}_n^{(1)}} \beta_{n,r} + \sum_{r \in \mathbb{J}_n^{(2)}} \alpha_{n,r} < \varepsilon$  (1.3.4).

Considerando  $A_n = \bigcup_{(r \in \mathbb{R}_n: \beta_{n,r} = 0 \text{ ou } \beta_{n,r} \leq \alpha_{n,r})} \Delta_{n,r} \in \mathfrak{B}_n$ , tem-se por (1.3.4) que  $\mu(A \Delta A_n) =$

$$= \mu(A \cap A_n^c) + \mu(A^c \cap A_n) = \sum_{r \in \mathbb{J}_n^{(1)}} \beta_{n,r} + \sum_{r \in \mathbb{J}_n^{(2)}} \alpha_{n,r} < \varepsilon, \text{ para } n \geq n_0, \text{ por (1.3.4) o que}$$

contradiz (1.3.3).  $\square$

Consideremos  $f$   $\mathcal{G}$ -mensurável e integrável e definamos o seguinte operador :

$$f \longrightarrow E^{\mathfrak{B}_n}(f)(x) = \frac{1}{\mu(\Delta_{n,r})} \int_{\Delta_{n,r}} f(x) \mu(dx), \text{ se } x \in \Delta_{n,r}$$

Imediatamente se prova que sendo  $\alpha, \beta \in \mathbb{R}$  e  $f, g \in L_1(\mathfrak{X})$ ,  $E^{\mathfrak{B}_n}(\alpha f + \beta g) = \alpha E^{\mathfrak{B}_n}(f) +$

$+\beta E^{\mathfrak{B}_n}(g)$ ; se  $f \leq g$  então  $E^{\mathfrak{B}_n}(f) \leq E^{\mathfrak{B}_n}(g)$  e que  $\int_{\mathfrak{X}} E^{\mathfrak{B}_n}(f)(x) \mu(dx) = \int_{\mathfrak{X}} f(x) \mu(dx)$ .

Temos então a seguinte

**Proposição 1.3.4:** Se  $(\mathfrak{P}_n)_{n \in \mathbb{N}}$  satisfaz (a) então para todo o  $f \in L_1(\mathfrak{X})$

$$\lim_n \int_{\mathfrak{X}} |E^{\mathfrak{B}_n}(f)(x) - f(x)| \mu(dx) = 0$$

**dem:**

Basta provar o resultado para  $f = \mathbb{1}_A$  com  $0 < \mu(A) < \infty$ , pois atendendo às propriedades do

operador  $E^{\mathfrak{B}_n}$ , expressas atrás, facilmente se passa para as funções escalonadas positivas, mensuráveis positivas e finalmente às integráveis.

Usando as mesmas notações do teorema anterior para  $f = \mathbb{1}_A$ ,

$$\int_{\mathfrak{X}} |E^{\mathfrak{B}_n}(f)(x) - f(x)| \mu(dx) = \sum_{r \in J_n} \frac{2\alpha_{n,r}\beta_{n,r}}{\alpha_{n,r} + \beta_{n,r}}$$

Por hipótese existe uma sucessão  $A_n \in \mathfrak{B}_n$  tal que  $\mu(A \Delta A_n) \rightarrow 0, n \rightarrow \infty$ . Vamos supor que para  $r \in J_n^{(1)}$ ,  $\Delta_{n,r} \subset A_n$  e para  $r \in J_n^{(2)}$ ,  $\Delta_{n,r} \cap A_n = \emptyset$  (se tal não acontecer construímos  $A'_n$  juntando ou tirando  $\Delta_{n,r}$  a  $A_n$  conforme o caso, tendo-se então  $\mu(A \Delta A'_n) \leq \mu(A \Delta A_n)$ ).

$$\text{Assim } \int_{\mathfrak{X}} |E^{\mathfrak{B}_n}(f)(x) - f(x)| \mu(dx) \leq 2 \sum_{r \in J_n^{(1)}} \beta_{n,r} + 2 \sum_{r \in J_n^{(2)}} \alpha_{n,r}.$$

Por outro lado, como  $A_n \in \mathfrak{B}_n$  tem-se que  $\mu(A \Delta A_n) \geq \sum_{r \in J_n^{(1)}} \beta_{n,r} + \sum_{r \in J_n^{(2)}} \alpha_{n,r}$ . Das

expressões anteriores a conclusão pretendida.  $\square$

Como para  $f \in \mathfrak{F}$ ,  $E^{\mathfrak{B}_n}(f) = E[\hat{f}_n]$ , concluímos da proposição anterior que :

**Teorema 1.3.2:** *Se a sucessão de partições  $(\mathfrak{P}_n)_{n \in \mathbb{N}}$  satisfaz (a) então  $\hat{\Delta}_{n,1} \rightarrow 0, n \rightarrow \infty$ , para todo o  $f \in \mathfrak{F}$ .*

Preocupemo-nos agora em determinar condições n/s para que  $\hat{\Delta}_{n,2} \rightarrow 0, n \rightarrow \infty$ ,  $\mathfrak{M}$ ; para tal seja  $(\mathfrak{P}_{\varphi(n)})_{n \in \mathbb{N}}$  ( $\varphi: \mathbb{N} \rightarrow \mathbb{N}$ ) a sucessão de partições de  $\mathfrak{X}$  cuja existência é garantida no parágrafo 1.1 e introduzamos as seguintes notações :

Para  $M > 0, n \in \mathbb{N}$  e  $C \in \mathfrak{C}$  de medida finita ponhamos

$$J_n(M, C) = \left\{ r \in R_{\varphi(n)} \mid \mu(\Delta_{\varphi(n), r} \cap C) \leq \frac{M}{n} \right\},$$

$$A_n(M, C) = \bigcup_{r \in J_n(M, C)} (\Delta_{\varphi(n), r} \cap C),$$

$$a_n(M, C) = \mu(A_n(M, C)),$$

$$e \text{ a} = \sup_{M,C} (\overline{\lim} a_n(M,C)).$$

**Definição 1.3.2:** Dizemos que a sucessão de partições  $(\mathbb{P}_{\varphi(n)})_{n \in \mathbb{N}}$  verifica a condição (b) se  $a = 0$ .

No caso particular de  $\mu$  ser finita seja  $a' = \sup_M (\overline{\lim} a_n(M, \mathfrak{G}))$ .

**Definição 1.3.3:** Sendo  $\mu$  finita, dizemos que a sucessão de partições  $(\mathbb{P}_{\varphi(n)})_{n \in \mathbb{N}}$  verifica a condição (b') se  $a' = 0$ .

As demonstrações dos dois teoremas seguintes são efectuadas supondo que  $\mu$  não é finita, pois sendo-o, estas facilmente se estabelecem.

**Teorema 1.3.2:** Se  $\Delta_{n,2} \xrightarrow{p..} 0, n \rightarrow \infty$ , para todo o  $f \in \mathfrak{F}'$ , então  $(\mathbb{P}_n)_{n \in \mathbb{N}}$  satisfaz a condição (b). Se  $\mu$  é finita e  $\Delta_{n,2} \xrightarrow{p..} 0, n \rightarrow \infty$ , para  $f = \frac{1}{\mu(\mathfrak{G})} \mathbb{1}_{\mathfrak{G}}$ , então  $(\mathbb{P}_n)_{n \in \mathbb{N}}$  satisfaz a condição (b').

dem:

Suponhamos que  $a \neq 0$ . Sejam então  $M > 0$  e  $C \in \mathcal{C}$ , de medida finita, tais que  $\overline{\lim} a_n(M,C) \geq \varepsilon$ , com  $\varepsilon > 0$ . Assim existe  $N_0 \subset \mathbb{N}$ , infinito, tal que para todo o  $n \in N_0$   $a_n(M,C) \geq \varepsilon$ .

Como  $\mu(C) > 0$ , tomemos  $\lambda = \frac{1}{\mu(C)}$  e  $f = \lambda \mathbb{1}_C$ . Ponhamos  $p_{n,r} = \int_{\Delta_{n,r}} f(x) \mu(dx) = \lambda \mu(\Delta_{n,r} \cap C)$ .

Definamos a v.a.  $Z_n = \sum_{r \in R_n} p_{n,r} \mathbb{1}_{\bigcap_{i=1}^n \{X_i \notin \Delta_{n,r}\}}$ , que verifica  $Z_n \leq 1$  e  $Z_n \leq \Delta_{n,2}$ . Assim

$E(Z_n) \xrightarrow{p..} 0, n \rightarrow \infty$  (1.3.6) pois por hipótese  $Z_n \xrightarrow{p..} 0, n \rightarrow \infty$ , p..

Para  $n \in N_0$ ,  $E(Z_n) = \sum_{r \in R_n} p_{n,r} (1 - p_{n,r})^n \geq \sum_{r \in J_n(M,C)} p_{n,r} (1 - \lambda \frac{M}{n})^n$ . Assim para  $n \in N_0$  e  $n$

s.g.  $E(Z_n) \geq \frac{\exp(-\lambda M)}{2} \sum_{r \in J_n(M,C)} p_{n,r} \geq \frac{\exp(-\lambda M)}{2} \lambda \varepsilon$  o que contraria (1.3.6).  $\square$

Por diversas vezes no decorrer deste trabalho vamos usar dois importantes resultados que são expressos nas seguintes proposições:

**Proposição 1.3.5** : (desigualdade de Bernstein) Seja  $X_1, X_2, \dots, X_n, n \in \mathbb{N}$ , uma sucessão de v.a. reais e independentes, tais que para todo o  $i=1, 2, \dots, n$

$$|E(X_i - E(X_i))^k| \leq H^{k-2} \frac{k!}{2} E(X_i - E(X_i))^2 \quad (1.3.7)$$

para  $k \geq 2$  e sendo  $H$  uma constante conveniente. Designando por  $\eta^2 = E(\sum_{i=1}^n (X_i - E(X_i))^2)$

então  $P(\sum_{i=1}^n (X_i - E(X_i)) > t\eta) \leq \exp(-\frac{t^2}{4})$ , para  $0 \leq t \leq \frac{\eta}{H}$ .

(Para demonstração ver Fréchet, 1937, p.103-135)

Um caso particular, que nos vai especialmente interessar, é aquele em que as v.a.  $X_i$  para  $i=1, 2, \dots, n$  são limitadas. E então fácil concluir que a condição (1.3.7) é verificada com  $H \geq \max_{i=1, 2, \dots, n} (\sup \omega |X_i(\omega)|)$ .

**Proposição 1.3.6** : Seja  $(v_1, v_2, \dots, v_k), k \in \mathbb{N}$ , uma v.a. multinomial de parâmetros  $(n, p_1, p_2, \dots, p_k)$ . Então dado  $\varepsilon > 0$

$$P(\sum_{i=1}^k |p_i - \frac{v_i}{n}| \geq \varepsilon) \leq \frac{\alpha}{\sqrt{n}} \left( \frac{(1+(k-1)/n)^{1+(k-(1/2))/n}}{((k-1)/n)^{(k-(1/2))/n}} \exp(-\frac{\varepsilon^2}{8}) \right)^n$$

onde  $\alpha$  é uma constante positiva independente de  $k, n$  e  $\varepsilon$ .

(Para traços gerais da demonstração ver Bosq, 1987, p.173-174)

Passemos agora a enunciar (e demonstrar) uma condição suficiente de convergência q.co. de  $\Delta_{n,2}$  para zero.

**Teorema 1.3.4** : Se  $(\mathbb{P}_n)_{n \in \mathbb{N}}$  satisfaz (b) ((b') no caso de  $\mu$  ser finita) então  $\Delta_{n,2} \xrightarrow{n \rightarrow \infty} 0$ , q.co., para todo o  $f \in \mathcal{F}$ .

**dem:**

Sejam então  $0 < \varepsilon < \frac{1}{4}$ , qualquer,  $C$  um elemento de  $\mathcal{C}$  tal que  $0 < \mu(C) < \infty$  e  $P_f(C^c) \leq \varepsilon$  (que sabemos existir pois  $\mu$  é  $\sigma$ -finita) e  $M > 0$  tal que

$$\left(1 + \frac{\mu(C)}{M}\right) \left(1 + \frac{M}{\mu(C)}\right)^M \exp(-\frac{\varepsilon^2}{8}) < 1 \quad (1.3.8)$$

Pretendemos provar que  $\sum_{n=1}^{\infty} P_f(\Delta_{n,2} \geq 6\varepsilon) < +\infty$ , onde  $\Delta_{n,2} = \sum_{r \in R_n} |p_{n,r} - \frac{v_n(\Delta_{n,r})}{n}|$

e  $p_{n,r} = P_f(\Delta_{n,r})$  (usamos também, ao longo deste trabalho, a notação  $P_f$  para designar a lei de probabilidade produto  $P_f^n$ ).

Por hipótese  $a = 0$ , logo para  $C$  e  $M$  atrás escolhidos  $\lim_n \mu(A_n(M,C)) = 0$ . Como  $P_f \ll \mu$  então existe uma ordem  $n_0 \in \mathbb{N}$  a partir da qual  $P_f(A_n(M,C)) < \varepsilon$ . Para  $n \geq n_0$  introduzamos as seguintes notações :  $p'_{n,r} = P_f(\Delta_{n,r} \cap C)$ ;  $p''_{n,r} = P_f(\Delta_{n,r} \cap C^c)$ ;  $v'_{n,r} = v_n(\Delta_{n,r})$ ;

$$v'_{n,r} = v_n(\Delta_{n,r} \cap C); v''_{n,r} = v_n(\Delta_{n,r} \cap C^c); p_n = \sum_{r \in J_n(M,C)} p'_{n,r} + \sum_{r \in R_n} p''_{n,r};$$

$$v_n = \sum_{r \in J_n(M,C)} v'_{n,r} + \sum_{r \in R_n} v''_{n,r}.$$

$$\text{Temos que } \Delta_{n,2} \leq \sum_{r \in J_n(M,C)} |p'_{n,r} - \frac{v'_{n,r}}{n}| + |p_n - \frac{v_n}{n}| + p_n + \frac{v_n}{n}.$$

Analisemos cada uma das parcelas do 2º membro da desigualdade anterior:

$$- p_n = P_f(A_n(M,C)) + P_f(C^c) \leq 2\varepsilon \quad (1.3.9).$$

$$- v_n = \sum_{i=1}^n \mathbb{1}_{A_n(M,C)}(X_i) + \sum_{i=1}^n \mathbb{1}_{C^c}(X_i), \text{ logo}$$

$$P_f\left(\frac{v_n}{n} \geq 3\varepsilon\right) \leq P_f\left(v_n - E(v_n) \geq \frac{\varepsilon\sqrt{n}}{\sqrt{\eta}}\sqrt{n\eta}\right),$$

onde  $\eta = P_f(A_n(M,C)) + P_f(C^c) - (P_f(A_n(M,C)) + P_f(C^c))^2$  é tal que  $E(v_n - E(v_n))^2 = n\eta$ .

Verifiquemos agora se existe  $\lambda > 0$  tal que para  $t = \sqrt{\lambda n}$  se tenha  $\begin{cases} t \leq \sqrt{n\eta} \\ t \leq \frac{\varepsilon\sqrt{n}}{\sqrt{\eta}} \end{cases}$ . Tal acontece

se existir  $\lambda > 0$  tal que  $\begin{cases} \lambda \leq \eta \\ \eta\lambda \leq \varepsilon^2 \end{cases}$ . Tomemos assim  $\lambda \leq \min\left(\frac{\varepsilon}{2}, P_f(C^c) - P_f^2(C^c)\right)$ . Pela

proposição 1.3.5 vem que :

$$P_f\left(\frac{v_n}{n} \geq 3\varepsilon\right) \leq P_f\left(v_n - E(v_n) \geq t\sqrt{n\eta}\right) \leq \exp\left(-\frac{\lambda n}{4}\right) \quad (1.3.10).$$

- Reparemos agora que  $v'_{n,r}$ , para  $r \in J_n(M,C)^c$ , dá-nos o número de elementos da amostra que "caem" em  $\Delta_{n,r} \cap C$  e  $v_n$  dos que "caem" em  $C^c \cup \left( \bigcup_{r \in J_n(M,C)} (\Delta_{n,r} \cap C) \right)$ . Temos

portanto uma experiência com  $\text{card}(J_n(M,C)^c)+1$  resultados possíveis (reparar que  $\text{card}(J_n(M,C)^c) \leq \frac{n}{M} \mu(C)$  (1.3.11)), logo  $((v'_{n,r})_{r \in J_n(M,C)^c}, v_n)$  segue uma lei multinomial

de parâmetros  $(n, (p'_{n,r})_{r \in J_n(M,C)^c}, p_n)$ . Pela proposição 1.3.6 vem que

$$P_f \left( \sum_{r \in J_n(M,C)^c} \left| p'_{n,r} - \frac{v'_{n,r}}{n} \right| + \left| p_n - \frac{v_n}{n} \right| \geq \varepsilon \right) \leq$$

$$\leq \frac{\alpha}{\sqrt{n}} \left( 1 + \frac{n}{\text{card} J_n(M,C)^c} \right)^{\frac{1}{2n} + \frac{\text{card} J_n(M,C)^c}{n}} \left( 1 + \frac{\text{card} J_n(M,C)^c}{n} \right) \exp\left(-\frac{\varepsilon^2}{8}\right)^n$$

e ainda

$$\leq \frac{\alpha}{\sqrt{n}} \quad , \text{ tendo em conta (1.3.8), (1.3.11) e o facto da função } x \rightarrow \left(1 + \frac{1}{x}\right)^x$$

ser crescente para  $x > 0$ .

De (1.3.9), (1.3.10) e da expressão anterior temos o pretendido.  $\square$

O teorema seguinte resume o que até aqui foi estabelecido

**Teorema 1.A** : São equivalentes as seguintes afirmações:

- i)  $(\mathcal{P}_n)_{n \in \mathbb{N}}$  verifica as condições (a) e (b) ((a') e (b') no caso de  $\mu$  ser finita)
- ii)  $\forall f \in \mathcal{F} \hat{\Delta}_n \rightarrow 0, n \rightarrow \infty, q.c.(p., q.c.)$
- iii)  $\forall f \in \mathcal{F}' \hat{\Delta}_n \rightarrow 0, n \rightarrow \infty, p.$

Ao enunciarmos este resultado surge o problema da compatibilidade das condições (a) e (b). No entanto sabemos (ver Abou-Jaoude, 1976, p.225-226) que de toda a sucessão de partições de  $\mathcal{X}$  satisfazendo a condição (a) é possível extrair uma subsucessão satisfazendo (a) e (b).

**Observação** : Se  $\mathcal{X} = \bigcup_{i \in I} (x_i)$  com  $I \subset \mathbb{N}$  e  $\mu$  é a medida de contagem sobre  $\mathcal{X}$  então, se

$P \ll \mu$  é a lei parente da amostra  $(X_1, X_2, \dots, X_n)$ , temos que o histograma associado à sucessão de partições  $\mathcal{P}_n = \{ (x_i) \mid i \in I \}$  ( $n \in \mathbb{N}$ ) é definido por  $f_n(x_i) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{(x_i)}(X_j)$ , para  $i \in I$ .

Conclui-se que  $\Delta_{n,1} = 0$  para todo o  $n \in \mathbb{N}$  e assim  $\Delta_{n,2} \rightarrow 0, n \rightarrow \infty, \mathbb{M}$ , sse  $\Delta_n \rightarrow 0, n \rightarrow \infty, \mathbb{M}$ .

Como  $(\mathcal{P}_n)_{n \in \mathbb{N}}$  satisfaz a condição (b) temos que  $\sum_{i \in I} |f_n(x_i) - P((x_i))| \rightarrow 0, n \rightarrow \infty, \mathbb{M}$ .

Como q.c.o.

Principalmente para o estudo do caso  $\mathcal{X} = \mathbb{R}^s$  (apesar do resultado ser mais geral), vamos a seguir obter uma condição, em certos casos equivalente à condição  $(\mathcal{P}_n)_{n \in \mathbb{N}}$  satisfaz (a) e (b) e que no caso de  $\mathbb{R}^s$  torna mais perceptível a formulação do teorema anterior.

Designemos por  $\alpha_n$  e  $\beta_n$  respectivamente as quantidades  $\inf_{r \in \mathbb{R}_n} \mu(\Delta_{n,r})$  e  $\sup_{r \in \mathbb{R}_n} \mu(\Delta_{n,r})$ .

**Definição 1.3.4:** Dizemos que a sucessão  $(\mathcal{P}_n)_{n \in \mathbb{N}}$  satisfaz a condição (c) se  $\alpha_n^{-1} = o(n)$ .

**Proposição 1.3.7:** Se  $(\mathcal{P}_n)_{n \in \mathbb{N}}$  satisfaz as condições (a) e (c) então satisfaz a condição (b). No caso de  $\overline{\lim} \left( \frac{\beta_n}{\alpha_n} \right) < \infty$  tem-se que se  $(\mathcal{P}_n)_{n \in \mathbb{N}}$  satisfaz a condição (b) então satisfaz a condição (c).

*dem:*

Provemos a primeira das afirmações:

Seja então  $M > 0$  e  $C \in \mathcal{C}$  tal que  $0 < \mu(C) < \infty$ . Por hipótese, verificando  $(\mathcal{P}_n)$  a condição (a), temos que para  $n$  s.g. existe  $A_n \in \mathcal{B}_n$  tal que  $\mu(A_n \Delta C) \rightarrow 0, n \rightarrow \infty$  (1.3.12).

Designemos por  $I_n = \{ r \in \mathbb{R}_n \mid \Delta_{n,r} \subset A_n \}$ . Facilmente se conclui que  $\mu(A_n) \rightarrow \mu(C), n \rightarrow \infty$ , e como consequência disso  $\text{card } I_n \leq \frac{2\mu(C)}{\alpha_n}$  para  $n$  s.g..

$$\text{Tomemos agora } a_n(M, C) = \sum_{r \in I_n \cap J_n(M, C)} \mu(\Delta_{n,r} \cap C) + \sum_{r \in I_n^c \cap J_n(M, C)} \mu(\Delta_{n,r} \cap C)$$

$$\leq \frac{2\mu(C)M}{n\alpha_n} + \sum_{r \in I_n^c} \mu(\Delta_{n,r} \cap C).$$



Como  $A_n \in \mathcal{B}_n$ ,  $\bigcup_{r \in I_n^c} \Delta_{n,r} = A_n^c$ , e assim  $\sum_{r \in I_n^c} \mu(\Delta_{n,r} \cap C) \leq \mu(A_n \Delta C)$ .

De (1.3.12), do facto de  $\alpha_n^{-1} = o(n)$  e das expressões anteriores concluímos o que pretendemos.

Para a demonstração da segunda condição suponhamos que  $(\mathbb{P}_n)$  não satisfaz (c), ou seja,

$$\exists L > 0 \exists N_0 \subset \mathbb{N}(\text{infinito}) : \forall n \in N_0, n \alpha_n < L \quad (1.3.13).$$

Sejam  $C \in \mathcal{C}$  tal que  $0 < \mu(C) < \infty$ ,  $M$  um majorante da sucessão  $\frac{\beta_n}{\alpha_n}$  e  $L_1 = LM$ . Por hipótese

$$\lim_{n \rightarrow \infty} a_n(L_1, C) = 0, \text{ e assim para } n \in N_0 \text{ e n s.g. temos que } \sum_{r \in J_n(L_1, C)} \mu(\Delta_{n,r} \cap C) < \mu(C) \quad (1.3.14).$$

Por outro lado, para  $r \in R_n$  e  $n \in N_0$ , atendendo a (1.3.13)  $\mu(\Delta_{n,r} \cap C) \leq \frac{L_1}{n}$ , e portanto

$$J_n(L_1, C) = R_n. \text{ Assim } \sum_{r \in J_n(L_1, C)} \mu(\Delta_{n,r} \cap C) = \mu(C).$$

Da expressão anterior e de (1.3.14) a contradição pretendida.  $\square$

Das proposições 1.3.3 e 1.3.7 e do teorema 1.A podemos obter :

**Teorema 1.B :** Em  $(\mathbb{R}^s, \mathcal{B}_{\mathbb{R}^s}, \lambda)$  são equivalentes as seguintes afirmações :

i)  $h_n \rightarrow 0, n \rightarrow \infty$ ;  $nh_n^s \rightarrow \infty, n \rightarrow \infty$ .

ii)  $\forall f \in \mathcal{F} \hat{\Delta}_n \rightarrow 0, n \rightarrow \infty$ , q.co.(p., q.c.)

iii)  $\forall f \in \mathcal{F}' \hat{\Delta}_n \rightarrow 0, n \rightarrow \infty$ , p.

#### 1.4. Convergência uniforme do estimador do histograma

A finalidade deste parágrafo é estabelecer condições necessárias ou suficientes para a referida convergência uniforme do estimador do histograma, ou seja, condições para que

$$\forall f \in D \quad \Gamma_n - \sup_{x \in \mathcal{X}} | \hat{f}_n(x) - f(x) | \rightarrow 0, n \rightarrow \infty, \mathfrak{M}. \quad (1.4.1)$$

com  $D$  um subconjunto conveniente de  $\mathcal{F}$ .

Vamos verificar que se  $D$  for muito "vasto" então não se tem a convergência pretendida. Com o exemplo seguinte mostramos que se tomarmos para conjunto das densidades

(considerando  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \lambda)$ ) a estimar  $\mathcal{F}\mathcal{C}(\mathbb{R})$  então é falso que

$$\forall f \in \mathcal{F}\mathcal{C}(\mathbb{R}) \forall \varepsilon > 0 P_f(\Gamma_n \geq \varepsilon) \rightarrow 0, n \rightarrow \infty.$$

Seja para  $0 < \varepsilon < 1$  a função  $f_{\varepsilon}$  definida por :

$$f_{\varepsilon}(x) = \begin{cases} 1 + \frac{k^2}{\varepsilon} (x-k), & \text{se } x \in [k - \frac{\varepsilon}{k^2}, k], \text{ com } k \text{ impar natural} \\ 1 + \frac{k^2}{\varepsilon} (k-x), & \text{se } x \in [k, k + \frac{\varepsilon}{k^2}], \text{ com } k \text{ impar natural} \\ 0, & \text{senão} \end{cases}$$

Evidentemente  $f_{\varepsilon} \geq 0$  e  $\int_{\mathbb{R}} f_{\varepsilon}(x) dx = \varepsilon \sum_{k \in \mathbb{N}} \frac{1}{(2k-1)^2}$ . Tomemos  $0 < \varepsilon_0 < 1$  de modo que

$f_{\varepsilon_0}$  seja uma densidade de probabilidade sobre  $\mathbb{R}$ .

Seja  $n \in \mathbb{N}$  e  $X^n$  uma  $n$ -amostra de lei  $P_{f_{\varepsilon_0}}$ ;  $f_n(\cdot; X^n)$  é nula fora de um intervalo real

limitado. Então  $\sup_{x \in \mathbb{R}} |f_n(x) - f_{\varepsilon_0}(x)| = 1$ , para todo o  $X^n \in \mathbb{R}^n$ , logo  $P_{f_{\varepsilon_0}}(\Gamma_n \geq 1) = 1$ .

Assim voltando ao caso geral de  $\mathcal{X}$ , concluímos que para termos (1.4.1) devemos impôr que  $\mathcal{F}\mathcal{C}(\mathcal{X}) \not\subset \mathcal{D}$ .

Façamos as seguintes hipóteses suplementares sobre o espaço  $(\mathcal{X}, \mathcal{C}, \mu)$  :

Suponhamos que  $(\mathcal{X}, \rho)$  é um espaço métrico sendo  $\mathcal{C}$  a sua tribo boreliana e que  $\mu$  é difusa e finita a distância finita (1).

Sejam  $\alpha_n$  e  $\beta_n$  definidas como no parágrafo anterior e suponhamos que a sucessão de partições  $(\mathcal{P}_n)_{n \in \mathbb{N}}$  de  $\mathcal{X}$ , satisfaz as seguintes condições de regularidade :

$$(d) \lim_n \sup_{r \in \mathbb{R}_n} \text{diam}(\Delta_{n,r}) = 0$$

$$(e) \lim \left( \frac{\beta_n}{\alpha_n} \right) < \infty$$

É trivial provar que sob a condição (d), para todo o  $x \in \mathcal{X}$  e  $f \in \mathcal{F}\mathcal{C}_{\mathcal{X}}$ , o estimador do histograma é assintoticamente cêntrico. Nas condições anteriores o estimador do histograma

(1) No caso de estarmos nas condições da observação que se segue ao teorema 1. A reparar que a convergência uniforme do estimador do histograma resulta como consequência da convergência  $L_1$ .

é consistente em média quadrática para todo o  $x \in \mathfrak{X}$  e  $f \in \mathcal{F} \mathfrak{C}_x$  sse  $\beta_n^{-1} = o(n)$ .

Vamos designar por  $\mathcal{D}$  o conjunto das densidades de probabilidade uniformemente contínuas em  $\mathfrak{X}$ .

Supomos habitualmente que  $f \in \mathcal{D}$ , ou seja, a lei que segue a população da qual extraímos a amostra  $X^n$ , possui uma versão u-contínua da derivada de Radon-Nikodym. Para não ser ambígua a expressão  $\sup_{x \in \mathfrak{X}} | \hat{f}_n(x) - f(x) |$  é necessário que essa versão, que supomos existir, seja única. Para tal basta-nos, no caso presente, provar que sendo  $A$  aberto não vazio de  $\mathfrak{X}$  se tem  $\mu(A) > 0$  (ver Prakasa Rao, 1983, p.32). Tal é trivial a partir da proposição seguinte obtida a partir da condição (d) :

*Proposição 1.4.1*: Para toda a bola em  $(\mathfrak{X}, \rho)$  existe uma ordem a partir da qual pelo menos um dos borelianos da partição  $(\mathcal{P}_n)$  está contido nessa bola, ou seja :

$$\forall x \in \mathfrak{X} \forall \varepsilon > 0 \exists n_0 \in \mathbb{N} : n \geq n_0 \exists r \in R_n : \Delta_{n,r} \subset B(x, \varepsilon)$$

Utilizando a proposição anterior e as condições (d) e (e), concluímos que  $\lim_n \beta_n = 0$ .

**Observação** : Reparar que a condição (d) só por si não permite concluir que  $\lim_n \beta_n = 0$ . Como exemplo basta tomar a sucessão  $\Pi_n = \{ [kh_n, (k+1)h_n[ \mid k \in \mathbb{Z} \}$  de partições de  $\mathbb{R}$ , com  $\lim_n h_n = 0$  e a medida  $\mu$  definida para todo o  $A \in \mathcal{B}_{\mathbb{R}}$  por  $\mu(A) = \int_A |x| dx$ .

Vamos agora mostrar através de um exemplo que a condição (d) é essencial para a convergência uniforme do histograma. A sua não verificação implica que o estimador do histograma pode não convergir para  $f$ , em probabilidade, mesmo pontualmente.

De facto consideremos  $\mathfrak{X} = ]0,1[$ ,  $\mu = \lambda$  (medida de Lebesgue sobre  $(]0,1[, \mathcal{B}_{]0,1[})$ ) e  $\rho$  a métrica usual. Seja  $k(n)$  uma sucessão de números naturais e definamos em  $]0,1[$  a seguinte sucessão de partições

$$\Pi_n = \{ [ \frac{r-1}{k(n)}, \frac{r}{k(n)} [ \mid r = 1, \dots, k(n) \} , n \in \mathbb{N}.$$

É claro que (d)  $\Leftrightarrow \lim_n \frac{1}{k(n)} = 0$ .

Seja  $f(x) = 2x$ ,  $x \in ]0,1[$ , e suponhamos que  $\lim_n \frac{1}{k(n)} \neq 0$ , ou seja,

$$\exists \alpha > 0 \exists N_0 \subset \mathbb{N}(\text{infinito}) : \forall n \in N_0 \quad k(n) < \frac{1}{\alpha}.$$

Como  $k(n) \in \mathbb{N}$ , para  $n \in \mathbb{N}$ , concluímos que existe  $\eta \in \mathbb{N}$  e uma parte infinita  $N_1$  de  $\mathbb{N}$  tal que para todo o  $n \in N_1$   $k(n) = \eta$ .

Tomemos  $n \in \mathbb{N}_1$  e  $\varepsilon > 0$ , então  $P_f(|\hat{f}_n(0) - f(0)| > \varepsilon) = P_f\left(\frac{v_n(0, \frac{1}{\eta})}{n} > \frac{\varepsilon}{\eta}\right)$ .

Pelo teorema do limite central  $\frac{v_n(0, \frac{1}{\eta}) - \frac{n}{\eta^2}}{\sqrt{\frac{n}{\eta^2} (1 - \frac{1}{\eta^2})}} \rightarrow N(0,1), n \rightarrow \infty, \mathfrak{X}$ , e assim

$$\overline{\lim} P_f(|\hat{f}_n(0) - f(0)| > \varepsilon) \geq 1 - \underline{\lim} P_f\left(\frac{v_n(0, \frac{1}{\eta}) - \frac{n}{\eta^2}}{\sqrt{\frac{n}{\eta^2} (1 - \frac{1}{\eta^2})}} \leq 0\right) \leq 0 \quad (\text{basta tomar } \varepsilon < \frac{1}{\eta})$$

$$\geq \frac{1}{2}.$$

Debrucemo-nos agora sobre a determinação de condições n/s para que  $\hat{\Gamma}_n \rightarrow 0, n \rightarrow \infty, \mathfrak{M}$ , para todo o  $f \in \mathfrak{D}$ .

**Teorema 1.4.1:** Se para todo o  $f \in \mathfrak{D}$ ,  $\hat{\Gamma}_n \rightarrow 0, n \rightarrow \infty, p$ , então  $\alpha_n^{-1} = o\left(\frac{n}{\log n}\right)$ .

dem:

Consideremos  $B(x_i, r_i), i=1,2$ , duas bolas em  $(\mathfrak{X}, \rho)$  tais que  $\rho(x_1, x_2) > r_1 + r_2$ . Pela proposição 1.4.1 temos que  $\mu(B(x_i, r_i)) > 0, i=1,2$ .

Definamos para  $i=1,2$  as funções  $g_i(x) = (r_i - \rho(x_1, x)) \mathbb{1}_{B(x_i, r_i)}(x)$  que são u-continuas em  $\mathfrak{X}$ ; então sendo  $\lambda_1, \lambda_2 \in \mathbb{R}^+$ , toda a função da forma  $f = \frac{\lambda_1 g_1 + \lambda_2 g_2}{\mu(\lambda_1 g_1 + \lambda_2 g_2)} \in \mathfrak{D}$  (1.4.2) (com

$$\mu(g) = \int_{\mathfrak{X}} g(x) \mu(dx).$$

Seja  $0 < \eta < \frac{r_1}{\mu(g_1)}$  e provemos que existe uma densidade  $f_0$  da forma (1.4.2) tal que

$$\forall x \in B(x_1, \frac{2}{3}r_1) \quad \frac{\eta}{3} \leq f_0(x) \leq \eta$$

De facto, para  $x \in B(x_1, \frac{2}{3}r_1)$  e  $f$  da forma (1.4.2)

$$\frac{\lambda_1 r_1}{3(\lambda_1 \mu(g_1) + \lambda_2 \mu(g_2))} \leq f(x) \leq \frac{\lambda_1 r_1}{\lambda_1 \mu(g_1) + \lambda_2 \mu(g_2)}$$

Tomando  $\lambda_1 = \frac{\lambda_2 \mu(g_2) \eta}{r_1 - \mu(g_1) \eta}$  temos o pretendido.

Consideremos agora o conjunto  $H_n = \{r \in \mathbb{R}_n \mid \Delta_{n,r} \subset B(x_1, \frac{2}{3}r_1)\}$  que para n s.g. é não vazio pela proposição 1.4.1.

Por hipótese  $\forall f \in \mathfrak{D} \forall \varepsilon > 0 \lim_n P_f(\hat{\Gamma}_n < \varepsilon) = 1$ , em particular tomando  $\varepsilon$  de modo que

$\frac{\eta}{3} - \varepsilon > 0$ , temos que :

$$P_{f_0}(\Gamma_n < \varepsilon) \leq P_{f_0}(\forall r \in H_n \frac{\eta}{3} - \varepsilon < \frac{\nu_n(\Delta_{n,r})}{n\mu(\Delta_{n,r})})$$

$$\leq P_{f_0}(\bigcap_{r \in H_n} (\nu_n(\Delta_{n,r}) > 0))$$

e assim  $\lim_n P_{f_0}(\bigcap_{r \in H_n} (\nu_n(\Delta_{n,r}) > 0)) = 1$ .

Sendo os  $\Delta_{n,r}$  disjuntos dois a dois para  $r \in H_n$  então

$$P_{f_0}(\bigcap_{r \in H_n} (\nu_n(\Delta_{n,r}) > 0)) = \prod_{r \in H_n} P_{f_0}(\nu_n(\Delta_{n,r}) > 0 | \bigcap_{r' \in H_n, r' < r} (\nu_n(\Delta_{n,r'}) > 0))$$

$$\leq \prod_{r \in H_n} P_{f_0}(\nu_n(\Delta_{n,r}) > 0)$$

e portanto  $\lim_n \prod_{r \in H_n} P_{f_0}(\nu_n(\Delta_{n,r}) > 0) = 1$ . Como para  $r \in H_n$

$$P_{f_0}(\nu_n(\Delta_{n,r}) > 0) \leq 1 - (1 - \eta\beta_n)^n \text{ concluímos que } \lim_n \prod_{r \in H_n} (1 - (1 - \eta\beta_n)^n) = 1.$$

Atendendo à condição (d) temos que para  $n$  s.g.  $\bigcup_{r \in H_n} \Delta_{n,r} \supset B(x_1, \frac{1}{2}r_1)$  e portanto

sendo o cardinal de  $H_n$  finito e  $L$  um majorante da sucessão  $(\frac{\beta_n}{\alpha_n})$  (que sabemos existir pela condição (e)) vem que  $\text{card } H_n \geq \frac{\mu(B(x_1, \frac{1}{2}r_1))}{L\alpha_n}$ . Assim temos que para todo o  $\eta > 0$

$$\lim_n (1 - (1 - \eta\beta_n)^n)^{1/\alpha_n} = 1.$$

Da expressão anterior obtemos que  $\lim_n \frac{(1 - \eta\beta_n)^n}{\alpha_n} = 0$  logo  $\lim_n (\log(\frac{1}{\alpha_n}) - n\eta\beta_n) = -\infty$  e portanto a partir de certa ordem  $\frac{1}{n\beta_n} \log(\frac{1}{\alpha_n}) < \eta$ , para todo o  $\eta > 0$ . Por maioria de razão  $\frac{1}{n\beta_n} \log(\frac{1}{\beta_n}) < \eta$ , para  $n$  s.g. e para todo o  $\eta > 0$ . Sendo  $\lim_n \beta_n = 0$ , da expressão anterior o  $\lim_n \frac{1}{n\beta_n} = 0$  e então  $\lim_n \frac{\log n}{n\beta_n} = 0$ . Como  $\frac{\log n}{n\alpha_n} \leq \frac{L \log n}{n\beta_n}$  temos a conclusão pretendida.  $\square$

Para a determinação de condições suficientes de convergência uniforme vamos precisar de resultados preliminares que são expressos nos lemas seguintes :

**Lema 1.4.1 :** *Seja  $A \in \mathcal{C}$  de medida finita e positiva e  $a$  um elemento de  $A$ . Então dado  $0 < \beta < \mu(A)$ , existe  $B \in \mathcal{C}$  tal que  $B \subset A$ ,  $a \in B$  e  $\mu(B) = \beta$ .*

**dem :**

Consideremos para  $\varepsilon > 0$ ,  $(a)_\varepsilon = \{y \in A \mid \rho(y, a) < \varepsilon\}$  e definamos a aplicação  $\phi : ]0, +\infty[ \rightarrow ]0, \mu(A)[$  por  $\phi(\varepsilon) = \mu((a)_\varepsilon)$ .  $\phi$  é contínua,  $\lim_{\varepsilon \rightarrow 0} \phi(\varepsilon) = 0$  e  $\lim_{\varepsilon \rightarrow \infty} \phi(\varepsilon) = \mu(A)$ . Assim dado  $0 < \beta < \mu(A)$ , existe  $\varepsilon_0 > 0$ , tal que  $\phi(\varepsilon_0) = \beta$ . O boreliano  $B = (a)_{\varepsilon_0}$  satisfaz as condições exigidas.  $\square$

**Lema 1.4.2 :** *Seja  $M$  um majorante inteiro da sucessão  $\beta_n / \alpha_n$ . Para  $n$  s.g. e  $r \in \mathbb{R}_n$ , existem borelianos  $\Delta^i_{n,r}$ ,  $i = 1, 2, \dots, M$ , tais que*

$$\bigcup_{i=1}^M \Delta^i_{n,r} = \Delta_{n,r} \text{ e } \mu(\Delta^i_{n,r}) = \alpha_n$$

**dem :**

Seja então  $n \in \mathbb{N}$  e  $\Delta_{n,r} \in \mathcal{P}_n$ , com  $r \in \mathbb{R}_n$ . Se  $\mu(\Delta_{n,r}) = \alpha_n$ , consideramos  $\Delta^i_{n,r} = \Delta_{n,r}$ ,  $i = 1, 2, \dots, M$ . Se  $\mu(\Delta_{n,r}) > \alpha_n$ , pelo lema anterior existe um boreliano  $\Delta^1_{n,r} \subset \Delta_{n,r}$  tal que  $\mu(\Delta^1_{n,r}) = \alpha_n$ .

Determinado  $\Delta^i_{n,r}$  ( $i \in \mathbb{N}$ ), se  $\mu(\Delta_{n,r} - (\bigcup_{j=1}^i \Delta^j_{n,r})) > \alpha_n$ , existe  $\Delta^{i+1}_{n,r} \subset \Delta_{n,r} - (\bigcup_{j=1}^i \Delta^j_{n,r})$

tal que  $\mu(\Delta^{i+1}_{n,r}) = \alpha_n$ .

E imediato provar que para algum  $i < M$  se tem  $\mu(\Delta_{n,r} - (\bigcup_{j=1}^i \Delta^j_{n,r})) \leq \alpha_n$ , para  $n$  s.g.

Seja então  $i_1 < M$ , satisfazendo  $\mu(\Delta^*_{n,r}) < \alpha_n$  (com  $\Delta^*_{n,r} = \Delta_{n,r} - (\bigcup_{j=1}^{i_1} \Delta^j_{n,r})$ ).

Considerando para  $\varepsilon > 0$ ,  $(\Delta^*_{n,r})_\varepsilon = \{y \in \Delta_{n,r} \mid \rho(y, \Delta^*_{n,r}) < \varepsilon\}$  tem-se que  $\Delta^*_{n,r} \subset (\Delta^*_{n,r})_\varepsilon \subset \Delta_{n,r}$ . Sendo  $\phi : ]0, +\infty[ \rightarrow ]\mu(\Delta^*_{n,r}), \mu(\Delta_{n,r})[$  definida por  $\phi(\varepsilon) = \mu((\Delta^*_{n,r})_\varepsilon)$  concluímos por raciocínio análogo ao do lema anterior que existe  $\varepsilon_0 > 0$ , tal que  $\phi(\varepsilon_0) = \alpha_n$ . Tomemos então  $\Delta^i_{n,r} = (\Delta^*_{n,r})_{\varepsilon_0}$ ,  $i = i_1 + 1, \dots, M$ .  $\square$

Da continuidade uniforme de  $f$  e da condição (d) resulta que :

**Lema 1.4.3:** Sendo  $f \in \mathcal{D}$  então  $\lim_n \sup_{x \in \mathcal{X}} |E[\hat{f}_n(x)] - f(x)| = 0$ .

**Teorema 1.4.2:** Se  $\alpha_n^{-1} = o(\frac{n}{\log n})$  então  $\Gamma_n \xrightarrow{\hat{}} 0, n \rightarrow \infty$ , q.co. para todo o  $f \in \mathcal{D}$ .

**dem:**

Seja  $f \in \mathcal{D}$ , qualquer e  $0 < \varepsilon < 1$  escolhido de modo que  $Q = \{x \in \mathcal{X} \mid f(x) \geq \varepsilon\}$  seja não vazio.

Introduzamos as seguintes notações :

$$\begin{aligned} I_n &= \{r \in R_n \mid \Delta_{n,r} \cap Q^c \neq \emptyset\}, J_n = R_n - I_n, \bar{f}_{n,r} = E[\hat{f}_{n,r}] \\ u_n &= \sup_{r \in J_n} \left| \frac{\nu_n(\Delta_{n,r})}{n\mu(\Delta_{n,r})} - \bar{f}_{n,r} \right| \text{ se } J_n \neq \emptyset, u_n = 0 \text{ se } J_n = \emptyset; \\ v_n &= \sup_{r \in I_n} \left| \frac{\nu_n(\Delta_{n,r})}{n\mu(\Delta_{n,r})} - \bar{f}_{n,r} \right| \text{ se } I_n \neq \emptyset, v_n = 0 \text{ se } I_n = \emptyset; \\ w_n &= \sup_{r \in I_n} \left( \frac{\nu_n(\Delta_{n,r})}{n\mu(\Delta_{n,r})} \right) \text{ se } I_n \neq \emptyset, w_n = 0 \text{ se } I_n = \emptyset. \end{aligned}$$

Tendo em conta o lema 1.4.3, basta-nos provar que

$$\sum_{n=1}^{\infty} P_f(\sup_{x \in \mathcal{X}} |\hat{f}_n(x) - E[\hat{f}_n(x)]| \geq \varepsilon) < +\infty.$$

Como  $P_f(\sup_{x \in \mathcal{X}} |\hat{f}_n(x) - E[\hat{f}_n(x)]| \geq \varepsilon) \leq P_f(v_n \geq \varepsilon) + P_f(u_n \geq \varepsilon)$  resta-nos mostrar que

$$\sum_{n=1}^{\infty} P_f(v_n \geq \varepsilon) < +\infty \quad (1.4.3) \text{ e } \sum_{n=1}^{\infty} P_f(u_n \geq \varepsilon) < +\infty \quad (1.4.4).$$

Da escolha de  $\varepsilon$ , da  $u$ -continuidade de  $f$  e da condição (d) tem-se que para  $x \in \Delta_{n,r}$  (n s.g.)  $f(x) < 2\varepsilon$ , com  $r \in I_n$ . Assim como  $v_n \leq w_n + \sup_{r \in I_n} \bar{f}_{n,r}$  temos que  $P_f(v_n \geq 3\varepsilon) \leq P_f(w_n \geq \varepsilon)$ .

Portanto provando que  $\sum_{n=1}^{\infty} P_f(w_n \geq \gamma M \varepsilon) < +\infty$  (1.4.5), onde  $\gamma$  é escolhido de modo que

$$1 + \frac{2}{\gamma} + \log\left(\frac{2}{\gamma}\right) < 0 \text{ e } M \text{ é um majorante natural da sucessão } \left(\frac{\beta_n}{\alpha_n}\right)_n, \text{ fica provado (1.4.3).}$$

Provemos (1.4.4) :

$$P_f(u_n \geq \varepsilon) \leq \sum_{r \in J_n} P_f(|\nu_{n,r} - E(\nu_{n,r})| \geq \sigma(\nu_{n,r}) \frac{\varepsilon n \mu(\Delta_{n,r})}{\sigma(\nu_{n,r})}) \text{ com } \nu_{n,r} = \nu_n(\Delta_{n,r}) \text{ e}$$

$\sigma(v_{n,r}) = \sqrt{n \bar{f}_{n,r} \mu(\Delta_{n,r}) (1 - \bar{f}_{n,r} \mu(\Delta_{n,r}))}$ . Escolhendo  $0 < \lambda < \varepsilon$  e sendo  $t = \lambda \sqrt{\frac{n \alpha_n}{2}}$

temos que para  $n$  s.g.  $\begin{cases} t \leq \sigma(v_{n,r}) \\ t \sigma(v_{n,r}) \leq \varepsilon n \mu(\Delta_{n,r}) \end{cases}$  com  $r \in J_n$ .

Assim pela proposição 1.3.5 para  $n$  s.g.

$$P_f(u_n \geq \varepsilon) \leq \frac{2}{\alpha_n \varepsilon} \exp\left(-\frac{\lambda^2 n \alpha_n}{8}\right) \text{ tendo em conta que } \text{card}(J_n) \leq \frac{1}{\alpha_n \varepsilon}$$

Por hipótese  $\alpha_n = \frac{\log n}{n \xi_n}$  com  $\lim_n \xi_n = 0$  e assim concluímos que para  $n$  s.g.

$$P_f(u_n \geq \varepsilon) \leq \frac{2}{\varepsilon} \frac{1}{\lambda^2 / 8 \xi_n - 1} \leq \frac{2}{\varepsilon} \frac{1}{n^p} \text{ com } p > 1.$$

Provemos (1.4.5):

$$P_f(w_n \geq \gamma M \varepsilon) \leq \sum_{i=1}^M \sum_{r \in I_n} P_f\left(\frac{v_n(\Delta_{n,r}^i)}{n \alpha_n} \geq \gamma \varepsilon\right) \quad (1.4.6) \text{ onde pelo lema 1.4.2, } \Delta_{n,r}^i,$$

$i=1, \dots, M$  são tais que  $\bigcup_{i=1}^M \Delta_{n,r}^i = \Delta_{n,r}$  e  $\mu(\Delta_{n,r}^i) = \alpha_n$ . Necessitamos agora de majorar de modo conveniente cada uma das parcelas anteriores. Como  $v_n(\Delta_{n,r}^i) \sim B(n, \alpha_n f_{n,r}^i)$  com

$$f_{n,r}^i = \frac{1}{\mu(\Delta_{n,r}^i)} \int_{\Delta_{n,r}^i} f(x) \mu(dx) \text{ tem-se que}$$

$$P_f(v_n(\Delta_{n,r}^i) \geq \gamma n \varepsilon \alpha_n) \leq \sum_{j=[\gamma n \varepsilon \alpha_n]}^n \frac{1}{j!} (n \alpha_n f_{n,r}^i)^j \quad (1)$$

$$\leq \frac{1}{[\gamma n \varepsilon \alpha_n]!} (n \alpha_n f_{n,r}^i)^{[\gamma n \varepsilon \alpha_n]} \exp(n \alpha_n f_{n,r}^i).$$

Definamos agora a função  $g: \mathbb{R}^+ \rightarrow \mathbb{R}^+$  por  $g(y) = y^{[\gamma n \varepsilon \alpha_n]} \exp(n \alpha_n y)$  que é convexa para

$n$  s.g., bastando para tal reparar que  $g$  é diferenciável para  $n$  s.g. e  $g'$  crescente.

Assim das expressões anteriores

$$P_f\left(\frac{v_n(\Delta_{n,r}^i)}{n \alpha_n} \geq \gamma \varepsilon\right) \leq \frac{1}{[\gamma n \varepsilon \alpha_n]!} (n \alpha_n)^{[\gamma n \varepsilon \alpha_n]} g(f_{n,r}^i)$$

(1)  $[\alpha]$  com  $\alpha \in \mathbb{R}^+$ , designa a parte inteira de  $\alpha$ .



$$\leq \frac{1}{(\gamma n \varepsilon \alpha_n)!} (n \alpha_n)^{[\gamma n \varepsilon \alpha_n]} \frac{1}{\alpha_n} \int_{\Delta_{n,r}^i} f(x)^{[\gamma n \varepsilon \alpha_n]} \exp(n \alpha_n f(x)) \mu(dx)$$

por aplicação da desigualdade de Jensen. Como para n s.g. se  $x \in \Delta_{n,r}^i$ , com  $r \in I_n$ ,  $f(x) \leq 2\varepsilon$  então

$$\leq \frac{1}{2\varepsilon \alpha_n (\gamma n \varepsilon \alpha_n)!} (2\varepsilon n \alpha_n)^{[\gamma n \varepsilon \alpha_n]} \exp(2n \alpha_n \varepsilon) \int_{\Delta_{n,r}^i} f(x) \mu(dx).$$

Tendo em conta (1.4.6) obtemos que

$$P_f(w_n \geq \gamma M \varepsilon) \leq \frac{M}{2\varepsilon \alpha_n (\gamma n \varepsilon \alpha_n)!} (2\varepsilon n \alpha_n)^{[\gamma n \varepsilon \alpha_n]} \exp(2n \alpha_n \varepsilon).$$

Aplicando a fórmula de Stirling e sabendo por hipótese que  $\alpha_n = \frac{\log n}{n \xi_n}$  com  $\lim_n \xi_n = 0$ ,

tiramos após longos cálculos que

$$\leq \frac{M}{\sqrt{2\pi}} \exp\left[\frac{\gamma \varepsilon \log n}{\xi_n} (1 + \frac{2}{\gamma} + \log(\frac{2}{\gamma})) (1 + o(1))\right].$$

Da escolha feita de  $\gamma$  e da desigualdade anterior vem para n s.g. que  $P_f(w_n \geq \gamma M \varepsilon) \leq$

$$\leq \frac{M}{\sqrt{2\pi} n^p} \text{ com } p > 1. \square$$

Tendo em conta os teoremas anteriores podemos então enunciar o resultado que nos estabelece condições n/s para a convergência uniforme do estimador do histograma, para os modos de convergência estocástica já referidos. Nessas condições temos a equivalência entre os três tipos de convergência.

***Teorema 1.C:*** São equivalentes as seguintes afirmações:

- i)  $\alpha_n^{-1} = o\left(\frac{n}{\log n}\right)$
- ii)  $\hat{\Gamma}_n \rightarrow 0, n \rightarrow \infty$ , q.co.(p.,q.c.) para todo o  $f \in \mathcal{D}$ .

No caso de  $\mathcal{X} = \mathbb{R}^S$ , basta notar que as condições (d) e (e) são satisfeitas desde que  $\lim_n h_n = 0$ .

### 1.5. Escolha assintoticamente óptima de $h_n$

A escolha de  $h_n$  é muito importante na aplicação prática como já referimos no parágrafo 1.1.. A escolha de  $h_n$  "grande" conduz-nos a um estimador muito suave e escolha de  $h_n$  "pequeno" a um estimador muito sinuoso (ver gráficos 1.5.8 e 1.5.9).

Vamos abordar o problema da determinação do valor optimal  $h_n^*$  de  $h_n$  que minimiza assintoticamente o erro quadrático médio integrável,  $EQMI(h_n) = \int_{\mathbb{R}^s} E(\hat{f}_n(x) - f(x))^2 dx$ .

Provaremos que, para uma vasta classe de densidades de probabilidade, a ordem de convergência optimal para o  $EQMI(h_n)$  é  $n^{-\frac{2}{s+2}}$  e é atingida quando  $h_n$  é tomado da ordem de  $n^{-\frac{1}{s+2}}$ . Antes, precisamos da noção de integral generalizado de Riemann que abordaremos a seguir.

**Definição 1.5.1:** Sendo  $\Pi = \{\Pi_r \in \mathcal{B}_{\mathbb{R}^s} \mid r \in \mathbb{R}\}$  uma partição numerável de  $\mathbb{R}^s$ , dizemos que

$g: \mathbb{R}^s \rightarrow \mathbb{R}$  é  $\Pi$ -escalorada se  $g = \sum_{r \in \mathbb{R}} \alpha_r \mathbb{1}_{\Pi_r}$ , onde  $\alpha_r \in \mathbb{R}$  para todo o  $r \in \mathbb{R}$ .

No que se vai seguir supomos que  $\mathbb{R}^s$ , está munido da sucessão de partições  $\mathcal{P}_n$  definida da forma usual.

**Definição 1.5.2:** Seja  $g: \mathbb{R}^s \rightarrow \mathbb{R}$ , mensurável. Dizemos que  $g$  é enquadrada se existem duas sucessões  $(g_n^+)$  e  $(g_n^-)$  de funções  $\mathcal{P}_n$ -escaloradas e integráveis tais que  $g_n^- \leq g \leq g_n^+$ .

para todo o  $n \in \mathbb{N}$ . Se além disso  $\lim_n \int_{\mathbb{R}^s} g_n^-(x) dx = \lim_n \int_{\mathbb{R}^s} g_n^+(x) dx = k$ , quando  $h_n \rightarrow 0, n \rightarrow$

$\rightarrow \infty$ , dizemos que  $k$  é o Integral Generalizado de Riemann (I.G.R.) de  $g$ .

Se  $k$  é o I.G.R. de  $g$  então  $g$  satisfaz as seguintes propriedades:

- $g$  é limitada
- $\lim g(x) = 0, \|x\| \rightarrow \infty$

-  $g$  é integrável em  $\mathbb{R}^s$  e  $\int_{\mathbb{R}^s} g(x) dx = k$ .

-  $g$  é integrável à Riemann em  $\mathbb{R}^s$  e  $(R) \int_{\mathbb{R}^s} g(x) dx = k$ .

A proposição seguinte é de bastante interesse para o estabelecimento do principal resultado deste parágrafo.

**Proposição 1.5.1:** Se  $g: \mathbb{R}^s \rightarrow \mathbb{R}$  é enquadrada então para todo o  $a > 0$ ,  $g_a(x) = \sup \{ |g(t)| : \|t-x\| \leq a \}$  é integrável em  $\mathbb{R}^s$  e  $\int_{\mathbb{R}^s} g_a(x) dx$  permanece limitado quando  $a \rightarrow 0$ .

*dem:*

Tomemos  $n \in \mathbb{N}$ , qualquer, mas fixo.

Seja  $\gamma_n(q_1, q_2, \dots, q_s) = \max \{ |g_n^+(q_1 h_n, q_2 h_n, \dots, q_s h_n)|, |g_n^-(q_1 h_n, q_2 h_n, \dots, q_s h_n)| \}$ . Con-

cluimos que  $\sup_{x \in \Delta_{n, q_1, q_2, \dots, q_s}} |g(x)| \leq \gamma_n(q_1, q_2, \dots, q_s)$  para todo o  $(q_1, q_2, \dots, q_s) \in \mathbb{Z}^s$ .

Definindo para  $x \in \mathbb{R}^s$ ,  $g_n^*(x) = \max \{ |g_n^+(x)|, |g_n^-(x)| \}$  temos que

$$g_n^*(x) = \sum_{(q_1, q_2, \dots, q_s) \in \mathbb{Z}^s} \gamma_n(q_1, q_2, \dots, q_s) \mathbb{1}_{\Delta_{n, q_1, q_2, \dots, q_s}}(x).$$

Atendendo à integrabilidade de  $g_n^*$   $\sum_{(q_1, q_2, \dots, q_s) \in \mathbb{Z}^s} \gamma_n(q_1, q_2, \dots, q_s) < \infty$ .

Seja  $a > 0$ , qualquer e  $\Delta_{n, q_1, q_2, \dots, q_s}^a = \prod_{i=1}^s [q_i h_n - a, (q_i + 1) h_n + a[$ . Temos que  $g_a(x) \leq$

$\leq \sum_{(q_1, q_2, \dots, q_s) \in \mathbb{Z}^s} \gamma_n(q_1, q_2, \dots, q_s) \mathbb{1}_{\Delta_{n, q_1, q_2, \dots, q_s}^a}(x)$ . Das expressões anteriores  $g_a$  é integrável

em  $\mathbb{R}^s$  e ainda  $\int_{\mathbb{R}^s} g_a(x) dx \leq (h_n + 2a)^s \sum_{(q_1, q_2, \dots, q_s) \in \mathbb{Z}^s} \gamma_n(q_1, q_2, \dots, q_s)$ .  $\square$

A proposição seguinte dá-nos condições para que  $g$  possua I.G.R. e permite concluir ainda que a noção de I.G.R. não depende da sucessão  $(h_n)$  que define a partição  $(\mathbb{P}_n)$ .

**Proposição 1.5.2:** Seja  $g$  integrável e limitada em  $\mathbb{R}^s$  possuindo um conjunto de pontos de descontinuidade de medida nula. Então as afirmações seguintes são equivalentes:

- i)  $\exists a > 0$  tal que  $g_a$  é integrável em  $\mathbb{R}^s$
- ii)  $g$  possui I.G.R.
- iii)  $\forall a > 0$   $g_a$  é integrável em  $\mathbb{R}^s$

des :

Pela proposição 1.5.1 basta provar que i)  $\Rightarrow$  ii).

Para  $(q_1, q_2, \dots, q_s) \in \mathbb{Z}^s$  definamos, para todo  $m \in \mathbb{N}$  e  $x \in \Delta_{m, q_1, q_2, \dots, q_s}$

$$g_m^+(x) = \sup_{u \in \Delta_{m, q_1, q_2, \dots, q_s}} g(u) \quad \text{e} \quad g_m^-(x) = \inf_{u \in \Delta_{m, q_1, q_2, \dots, q_s}} g(u).$$

Seja  $m_0 \in \mathbb{N}$  tal que para  $m \geq m_0$ ,  $h_m \sqrt{s} < a$ . Assim para  $m \geq m_0$  se  $x \in \Delta_{m, q_1, q_2, \dots, q_s}$

$$\Delta_{m, q_1, q_2, \dots, q_s} \subset B(x, a).$$

Dado então  $x \in \mathbb{R}^s$ , seja  $(q_1, q_2, \dots, q_s) \in \mathbb{Z}^s$  tal que  $x \in \Delta_{m, q_1, q_2, \dots, q_s}$ . Temos que  $|g_m^+(x)| \leq$

$$\leq \sup_{u \in \Delta_{m, q_1, q_2, \dots, q_s}} |g(u)| \leq g_a(x) \quad \text{e} \quad \text{também} \quad |g_m^-(x)| \leq g_a(x) \quad (1.5.1).$$

Portanto para  $m \geq m_0$ ,  $g_m^+$  e  $g_m^-$  são integráveis e  $g_m^- \leq g \leq g_m^+$ .

Seja então  $\varepsilon > 0$ , qualquer. Existe  $q \in \mathbb{N}$  tal que (sendo  $I = \int_{\mathbb{R}^s} g(x) dx$ )

$$\int_{\mathbb{R}^s - [-q, q]^s} g_a(x) dx < \frac{\varepsilon}{3} \quad \text{e} \quad \left| \int_{[-q, q]^s} g(x) dx - I \right| < \frac{\varepsilon}{3} \quad (1.5.2).$$

De (1.5.1) tiramos que para  $m \geq m_0$

$$\int_{\mathbb{R}^s - [-q, q]^s} |g_m^+(x)| dx + \int_{\mathbb{R}^s - [-q, q]^s} |g_m^-(x)| dx < \frac{\varepsilon}{3} \quad (1.5.3).$$

Se  $g$  integrável à Riemann em  $[-q, q]^s$  temos que

$$\int_{[-q, q]^s} g(x) dx = \lim_{m \rightarrow \infty} \int_{[-q, q]^s} g_m^+(x) dx = \lim_{m \rightarrow \infty} \int_{[-q, q]^s} g_m^-(x) dx$$

e portanto tomando  $m_1 \in \mathbb{N}$  ( $m_1 \geq m_0$ ) tal que

$$\int_{[-q, q]^s} g(x) dx - \int_{[-q, q]^s} g_m^-(x) dx < \frac{\varepsilon}{3} \quad \text{e} \quad \int_{[-q, q]^s} g_m^+(x) dx - \int_{[-q, q]^s} g(x) dx < \frac{\varepsilon}{3}$$

obtemos, para  $m \geq m_1$ , atendendo a (1.5.2) e (1.5.3) que

$$I - \varepsilon \leq \int_{\mathbb{R}^s} g_m^-(x) dx \leq \int_{\mathbb{R}^s} g_m^+(x) dx \leq I + \varepsilon \quad \square$$

**Observação :** No caso de  $s=1$ , suponhamos que  $g$  é limitada, integrável e que o conjunto dos seus pontos de descontinuidade é de medida nula. Se  $g$  é de variação limitada nas vizinhanças do infinito (existem  $a, b \in \mathbb{R}$  tais que  $\int_{-\infty}^a (g) < +\infty$  e  $\int_b^{+\infty} (g) < +\infty$ ) então  $g$  possui I.G.R..

Vamos supor que  $I$  é um subconjunto de  $\mathbb{R}^s$  que admite uma sucessão de partições da forma  $\mathcal{P}_n = \{\Delta_{n,q_1,q_2,\dots,q_s} \mid (q_1,q_2,\dots,q_s) \in \mathcal{Z}_n^s\}$  (com  $n \in \mathbb{N}$ ,  $\mathcal{Z}_n^s \subset \mathbb{Z}^s$  e  $\mathcal{P}_n \subset \mathcal{P}_{n+1}$ ) tal que para  $f \in \mathcal{F}$  a sua aderência contém o suporte de  $f$ .

**Teorema 1.5.1:** *Seja  $f \in \mathcal{F} \cap L_2(\mathbb{R}^s)$ , duas vezes diferenciável no interior de  $I$ . Suponhamos que os quadrados das derivadas parciais de 2ª ordem são enquadrados e que os das derivadas de 1ª ordem possuem I.G.R.. Se  $h_n \rightarrow 0, n \rightarrow \infty$  então*

$$EQMI(h_n) = \frac{1}{nh_n^s} + \frac{h_n^2}{12} \int_{\mathbb{R}^s} \sum_{i=1}^s \left( \frac{\partial f}{\partial x_i}(x) \right)^2 dx + O\left(\frac{1}{n}\right) + o(h_n^2)$$

*dem :*

Para  $n \in \mathbb{N}$ ,  $(q_1, q_2, \dots, q_s) \in \mathcal{Z}_n^s$  e  $x \in \Delta_{n,q_1,q_2,\dots,q_s}$  temos que  $E[f_n^\wedge(x) - f(x)]^2 = (E[f_n^\wedge(x)] - f(x))^2 + \frac{E[f_n^\wedge(x)]^2 - E^2[f_n^\wedge(x)]}{nh_n^s}$ . Assim  $\int_{\Delta_{n,q_1,q_2,\dots,q_s}} E[f_n^\wedge(x) - f(x)]^2 dx = \int_{\Delta_{n,q_1,q_2,\dots,q_s}} (E[f_n^\wedge(x)] - f(x))^2 dx + \frac{1}{nh_n^s} \int_{\Delta_{n,q_1,q_2,\dots,q_s}} E[f_n^\wedge(x)]^2 dx - \frac{1}{n} \int_{\Delta_{n,q_1,q_2,\dots,q_s}} E^2[f_n^\wedge(x)] dx$  para todo o  $(q_1, q_2, \dots, q_s) \in \mathcal{Z}_n^s$ .

Como  $0 \leq \int_{\Delta_{n,q_1,q_2,\dots,q_s}} (E[f_n^\wedge(x)] - f(x))^2 dx = \int_{\Delta_{n,q_1,q_2,\dots,q_s}} f^2(x) dx - \int_{\Delta_{n,q_1,q_2,\dots,q_s}} E^2[f_n^\wedge(x)] dx$  vem

$$\text{que } EQMI(h_n) = \frac{1}{nh_n^s} + \sum_{(q_1,q_2,\dots,q_s) \in \mathcal{Z}_n^s} \int_{\Delta_{n,q_1,q_2,\dots,q_s}} (E[f_n^\wedge(x)] - f(x))^2 dx + O\left(\frac{1}{n}\right) \quad (1.5.4)$$

Estudemos o termo  $\int_{\Delta_{n,q_1,q_2,\dots,q_s}} (E[f_n^\wedge(x)] - f(x))^2 dx$ . Desenvolvendo  $f$  em fórmula de Taylor

centrada no ponto  $x^* = ((q_1 + \frac{1}{2})h_n, (q_2 + \frac{1}{2})h_n, \dots, (q_s + \frac{1}{2})h_n)$  obtemos

$$f(x) = f(x^*) + \sum_{i=1}^s (x_i - x_i^*) \frac{\partial f}{\partial x_i}(x^*) + \frac{1}{2} \sum_{i,j=1}^s (x_i - x_i^*)(x_j - x_j^*) \frac{\partial^2 f}{\partial x_i \partial x_j}(x^* + \alpha(x - x^*)), \alpha \in ]0,1[.$$

Por integração sobre  $\Delta_{n,q_1,q_2,\dots,q_s}$ ,  $E[\hat{f}_n(x)] = f(x^*) + h_n^2 A_{n,q_1,q_2,\dots,q_s}$

$$\text{onde } |A_{n,q_1,q_2,\dots,q_s}| \leq \frac{1}{8} \sum_{i,j=1}^s \sup_{x \in \text{int}(\Delta_{n,q_1,q_2,\dots,q_s})} \left| \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right|.$$

$$\begin{aligned} \text{Tiramos então que} \quad & \int_{\Delta_{n,q_1,q_2,\dots,q_s}} (E[\hat{f}_n(x)] - f(x))^2 dx = \\ & \int_{\Delta_{n,q_1,q_2,\dots,q_s}} \sum_{i=1}^s (x_i - x_i^*)^2 \left( \frac{\partial f}{\partial x_i}(x^*) \right)^2 dx + h_n^3 B_{n,q_1,q_2,\dots,q_s} + h_n^4 C_{n,q_1,q_2,\dots,q_s}, \quad (1.5.5) \end{aligned}$$

onde  $|B_{n,q_1,q_2,\dots,q_s}| \leq$

$$\leq \frac{h_n^3}{4} \sum_{k=1}^s \sup_{x \in \text{int}(\Delta_{n,q_1,q_2,\dots,q_s})} \left| \frac{\partial f}{\partial x_k}(x) \right| \cdot \sum_{i,j=1}^s \sup_{x \in \text{int}(\Delta_{n,q_1,q_2,\dots,q_s})} \left| \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right|$$

$$\text{e } |C_{n,q_1,q_2,\dots,q_s}| \leq \frac{h_n^4}{16} \left( \sum_{i,j=1}^s \sup_{x \in \text{int}(\Delta_{n,q_1,q_2,\dots,q_s})} \left| \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right| \right)^2.$$

$$\text{Assim } \sum_{(q_1,q_2,\dots,q_s) \in \mathcal{Z}_n^s} |B_{n,q_1,q_2,\dots,q_s}| \leq \frac{1}{4} \sum_{i,j,k=1}^s \left( \int_1^s g_{n,k}(u) du \right)^{\frac{1}{2}} \left( \int_1^s h_{n,i,j}(u) du \right)^{\frac{1}{2}} \text{ e}$$

$$\sum_{(q_1,q_2,\dots,q_s) \in \mathcal{Z}_n^s} |C_{n,q_1,q_2,\dots,q_s}| \leq \frac{1}{16} \sum_{i,j,k,l=1}^s \left( \int_1^s h_{n,i,j}(u) du \right)^{\frac{1}{2}} \left( \int_1^s h_{n,k,l}(u) du \right)^{\frac{1}{2}}$$

por utilização da desigualdade de Hölder, onde  $g_{n,k}(u) = \sup_{x \in \text{int}(\Delta_{n,q_1,q_2,\dots,q_s})} \left( \frac{\partial f}{\partial x_k}(x) \right)^2$  e

$h_{n,i,j}(u) = \sup_{x \in \text{int}(\Delta_{n,q_1,q_2,\dots,q_s})} \left( \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right)^2$  para  $n \in \mathbb{N}$ ,  $(q_1, q_2, \dots, q_s) \in \mathcal{Z}_n^s$ ,  $i, j, k, l = 1, 2, \dots, s$  e

$u \in \Delta_{n,q_1,q_2,\dots,q_s}$ .

Como por hipótese e pela proposição 1.5.1 os integrais  $\int_{\mathbb{R}^s} g_{n,k}(u) du$  e  $\int_{\mathbb{R}^s} h_{n,i,j}(u) du$

$(i, j, k, l = 1, 2, \dots, s)$  permanecem limitados quando  $h_n \rightarrow 0, n \rightarrow \infty$ , temos de (1.5.4) e (1.5.5) que

$$EQMI(h_n) = \frac{1}{nh_n^s} + \frac{h_n^2}{12} \sum_{(q_1, q_2, \dots, q_s) \in \mathbb{Z}^s} \int_{\Delta_{n, q_1, q_2, \dots, q_s}} \sum_{i=1}^s \left( \frac{\partial f}{\partial x_i}(x^*) \right)^2 dx + O\left(\frac{1}{n}\right) + O(h_n^3).$$

Da existência do I.G.R. de  $\left(\frac{\partial f}{\partial x_i}\right)^2$   $i=1, 2, \dots, s$ , concluímos que

$$\lim_n \int_{\mathbb{R}^s} \left( \frac{\partial f}{\partial x_i}(x^*) \right)^2 dx = \int_{\mathbb{R}^s} \left( \frac{\partial f}{\partial x_i}(x) \right)^2 dx$$

e portanto  $EQMI(h_n) = \frac{1}{nh_n^s} + \frac{h_n^2}{12} \int_{\mathbb{R}^s} \sum_{i=1}^s \left( \frac{\partial f}{\partial x_i}(x) \right)^2 dx + o(h_n^2) + O\left(\frac{1}{n}\right).$  □

Para cada  $n \in \mathbb{N}$ , designemos por *erro quadrático médio integrado assintótico em*  $h_n \geq 0$ ,  $EQMIA^*(h_n)$ , a quantidade  $\frac{1}{nh_n^s} + \frac{h_n^2}{12} \int_{\mathbb{R}^s} \sum_{i=1}^s \left( \frac{\partial f}{\partial x_i}(x) \right)^2 dx$ , para  $f \in \mathcal{F}$  nas condições do teorema anterior.

**Corolário 1.5.1:** Para cada  $n \in \mathbb{N}$ , o valor  $h_n^* > 0$ , que minimiza o  $EQMIA^*(h_n)$  é

$$h_n^* = \left( \frac{6s}{I_s} \right)^{\frac{1}{s+2}} n^{-\frac{1}{s+2}}$$

onde supomos que  $I_s = \int_{\mathbb{R}^s} \sum_{i=1}^s \left( \frac{\partial f}{\partial x_i}(x) \right)^2 dx \neq 0$ . Tem-se também que

$$EQMI(h_n^*) = \left(1 + \frac{s}{2}\right) \left(\frac{I_s}{6s}\right)^{\frac{s}{s+2}} n^{-\frac{2}{s+2}} + o\left(n^{-\frac{2}{s+2}}\right).$$

Restringindo o conjunto das densidades a estimar, podem ser encontrados resultados que melhoram a informação dada pelos teorema e corolário 1.5.1. Por exemplo, supondo  $f$  de suporte limitado, possuindo derivadas parciais de 1ª e 2ª ordem limitadas no interior de  $I$ , se  $h_n \rightarrow 0$ ,  $n \rightarrow \infty$  então

$$EQMI(h_n) = \frac{1}{nh_n^s} + \frac{h_n^2}{12} \int_{\mathbb{R}^s} \sum_{i=1}^s \left( \frac{\partial f}{\partial x_i}(x) \right)^2 dx + O\left(\frac{1}{n}\right) + O(h_n^3).$$

Assim  $EQMI(h_n^*) = \left(1 + \frac{s}{2}\right) \left(\frac{I_s}{6s}\right)^{\frac{s}{s+2}} n^{-\frac{2}{s+2}} + O\left(n^{-\frac{3}{s+2}}\right).$

Como na prática o valor utilizado de  $h_n$  não pode ser dado pelo corolário anterior, temos todo o interesse em saber como se comporta o  $EQMIA^{\wedge}$  quando  $h_n$  varia. Suponhamos assim que em vez de tomarmos para  $h_n$  o valor óptimo dado pelo corolário anterior escolhiamos  $h_n = ch_n^*$ , com  $c > 0$ . Estudemos a influência de tal escolha sobre o  $EQMIA^{\wedge}(h_n)$ .

É imediato provar que para  $h_n = ch_n^*$  se tem  $EQMIA^{\wedge}(h_n) = \left(\frac{2 + sc^{s+2}}{c^s(s+2)}\right) EQMIA^{\wedge}(h_n^*)$ . A função  $s \rightarrow \frac{2 + sc^{s+2}}{c^s(s+2)}$  é estritamente crescente para  $c \neq 1$ . Se  $c < 1$ ,  $\frac{2 + sc^{s+2}}{c^s(s+2)} \rightarrow +\infty, s \rightarrow \infty$  e se  $c > 1$ ,  $\frac{2 + sc^{s+2}}{c^s(s+2)} \rightarrow c^2, s \rightarrow \infty$ . Assim o aumento da dimensão(s) implica um aumento do  $EQMIA^{\wedge}(h_n)$  relativamente ao  $EQMIA^{\wedge}(h_n^*)$ , tendo-se, no caso de  $c > 1$ , um limite superior para o aumento verificado.

Estes aumentos são expressos no quadro seguinte para alguns valores de  $s$  e  $c$ :

$s \backslash c$	0.5	0.75	1.25	2
1	41%	8%	5%	66%
2	113%	17%	10%	113%
3	235%	29%	14%	145%

Quadro 1.5.1

As conclusões anteriores foram tiradas a partir do  $EQMIA^{\wedge}$ , que sabemos aproximar o  $EQMI$  para  $f \in \mathcal{F}$  nas condições do teorema 1.5.1. No entanto não nos devemos esquecer do carácter assintótico de todo o estudo anterior.

A nível experimental e no caso de  $s=1$ , verificámos que para  $n$  pequeno (testámos  $n=10$  e  $n=50$ ) existe um relativo afastamento entre o  $EQMIA^{\wedge}(h_n^*)$  e o erro quadrático médio esti-



mado  $EQMI^*(h_n^*)$  (1), apesar de que para as densidades testadas se verificou que o  $EQMI^*(h_n^*)$  foi sempre inferior ao  $EQMIA^*(h_n^*)$ .

Nos quadros seguintes apresentamos os resultados da simulação efectuada :

n	$h_n^* \times 10$	$EQMIA^*(h_n^*) \times 10^3$	$EQMI^*(h_n^*) \times 10^3$
10	16,2	92,6	67,6
50	9,46	31,7	26,2
100	7,51	19,9	17,8
500	4,4	6,82	6,17

Quadro 1.5.2 - Lei normal centrada e reduzida

n	$h_n^* \times 10$	$EQMIA^*(h_n^*) \times 10^3$	$EQMI^*(h_n^*) \times 10^3$
10	19,6	76,5	60,3
50	11,5	26,2	25,4
100	9,08	16,5	14,3
500	5,32	5,64	5,37

Quadro 1.5.3 - Lei de Cauchy

n	$h_n^* \times 10$	$EQMIA^*(h_n^*) \times 10^3$	$EQMI^*(h_n^*) \times 10^3$
10	10,6	141,0	77,1
50	6,21	48,3	42,6
100	4,93	30,4	24,1
500	2,88	10,4	9,08

Quadro 1.5.4 - Lei exponencial de parâmetro 1

$$(1) EQMI^*(h_n) = \frac{1}{m(n)} \sum_{i=1}^{m(n)} \int_{\mathbb{R}} (\hat{f}_n(x; X_i^n) - f(x))^2 dx, \text{ (tomamos } m(10)=50, m(50)=40, m(100)=$$

$=m(500)=30$ ), onde  $X_i^n$   $i=1,2,\dots,m(n)$  são  $m(n)$  n-amostras de lei Pf e o integral  $\int_{\mathbb{R}} (\hat{f}_n(x; X_i^n) -$

$- f(x))^2 dx$  é calculado computacionalmente.

n	$h_n^* \times 10^2$	$EQMIA^*(h_n^*) \times 10^2$	$EQMI^*(h_n^*) \times 10^2$
10	21,2	70,6	37,8
50	12,4	24,1	20,9
100	9,9	15,2	12,4
500	5,8	5,2	4,4

Quadro 1.5.5 - Lei exponencial de parâmetro 5

n	$h_n^* \times 10$	$EQMIA^*(h_n^*) \times 10^3$	$EQMI^*(h_n^*) \times 10^3$
10	20,6	72,8	45,6
50	12,0	24,9	16,2
100	9,54	15,7	11,9
500	5,46	5,37	4,83

Quadro 1.5.6 - Mistura de leis normais ( $\frac{1}{2}N(-\frac{3}{2},1) + \frac{1}{2}N(\frac{3}{2},1)$ )

n	$h_n^* \times 10$	$EQMIA^*(h_n^*) \times 10^3$	$EQMI^*(h_n^*) \times 10^3$
10	13,4	112,0	69,2
50	7,83	38,3	29,4
100	6,21	24,1	20,0
500	3,63	8,26	6,96

Quadro 1.5.7 - Lei gama de parâmetro 2

n	$h_n^* \times 10$	$EQMIA^*(h_n^*) \times 10^3$	$EQMI^*(h_n^*) \times 10^3$
10	17,2	87,1	73,0
50	10,1	29,8	25,5
100	7,99	18,8	14,2
500	4,68	6,42	6,41

Quadro 1.5.8 - Lei t(5) de Student

Quanto aos aumentos experimentais do  $EQMI^{\sim}(h_n)$  relativamente ao  $EQMI^{\sim}(h_n^*)$  é de referir que para  $n$  pequeno eles estão por vezes longe dos valores anunciados no Quadro 1.5.1. A título de exemplo temos a lei normal em que os aumentos utilizando  $h_n=2h_n^*$  são respectivamente de 117% e 103% para  $n=10$  e  $n=50$ .

No caso da lei exponencial de parâmetro 5, para  $n=50$ ,  $h_n=2h_n^*$  e  $h_n=h_n^*/2$ , os aumentos são respectivamente de 44,5% e 62,8%. A mesma simulação para a lei gama de parâmetro 2, mas para  $n=10$ , dá respectivamente uma diminuição de 46,1% e um aumento de 82,5% do  $EQMI^{\sim}(h_n)$  relativamente ao  $EQMI^{\sim}(h_n^*)$ .

Para amostras de tamanho 500 foram obtidos resultados mais próximos dos esperados.

Alargámos o estudo de simulação ao caso das densidades das leis exponencial bilateral ( $f(x) = \frac{1}{2} \exp(-|x|)$ ) e triangular ( $f(x) = (1 - |x|) \mathbb{1}_{[-1,1]}(x)$ ). A primeira destas densidades

apesar de não satisfazer as condições do teorema 1.5.1, é fácil verificar, utilizando o mesmo tipo de demonstração, que o resultado do mesmo é válido para esta densidade de probabilidade.

Quanto à densidade triangular vamos recorrer a um resultado que nos garante ser o desenvolvimento obtido para o EQMI válido para uma maior classe de densidades de probabilidade no caso  $s=1$ , classe essa que inclui a densidade triangular (ver Freedman, 1981, p.454).

$n$	$h_n^* \times 10$	$EQMI^{\sim}(h_n^*) \times 10^3$	$EQMI^{\sim}(h_n) \times 10^3$
10	13,4	112,0	78,2
50	7,83	38,3	36,1
100	6,21	24,1	22,2
500	3,63	8,25	7,95

Quadro 1.5.9 - Lei exponencial bilateral

$n$	$h_n^* \times 10$	$EQMI^{\sim}(h_n^*) \times 10^2$	$EQMI^{\sim}(h_n) \times 10^2$
10	6,7	22,4	15,6
50	3,9	7,66	5,28
100	3,1	4,83	3,83
500	1,5	1,65	1,46

Quadro 1.5.10 - Lei triangular

O método do histograma é aplicado a seguir a n-amostras simuladas.

Os gráficos seguintes representam, para alguns valores de n, estimativas de algumas das densidades anteriores. Quando nada é dito em contrário, nas estimativas são utilizados os valores de  $h_n^*$  dados nos quadros anteriores. Na legenda dos gráficos além do tamanho da

amostra utilizada e do erro quadrático integrado,  $EQI = \int_{\mathbb{R}} (f_n^{\hat{}}(x) - f(x))^2 dx$ , da estimativa

relativamente à verdadeira densidade surge também a indicação do plano a que este se refere.

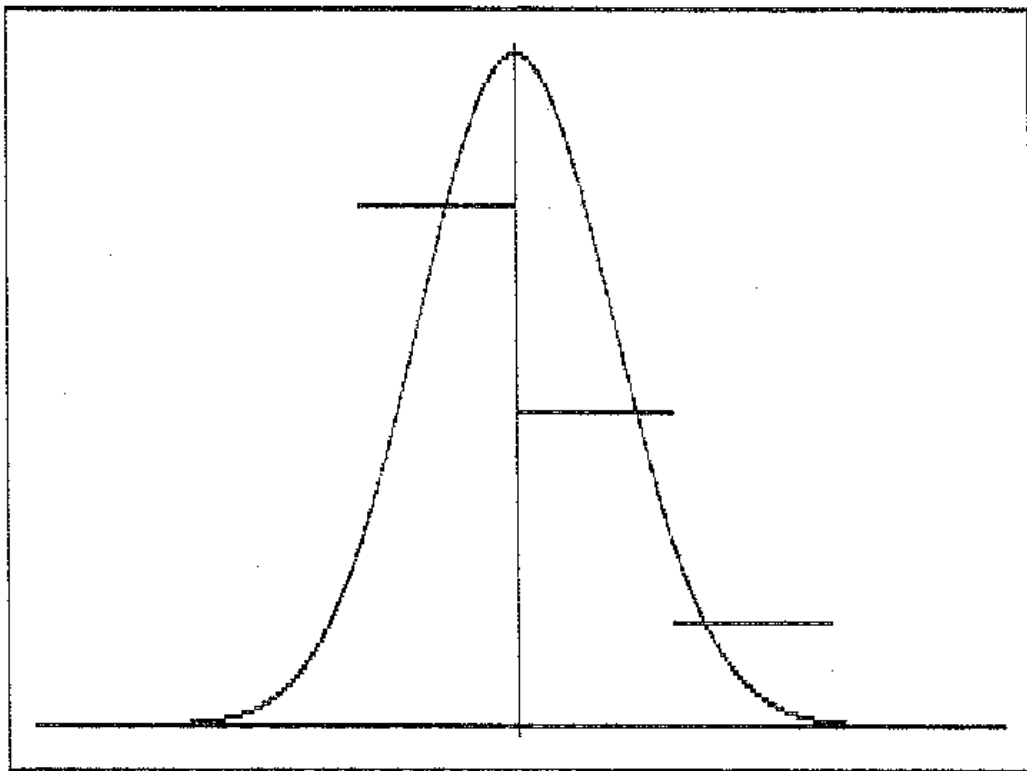


Gráfico 1.5.1 - Lei normal centrada e reduzida.

$n = 10$ ,  $EQI = 0.0497$ ,  $(-5;5) \times (0,0.399)$ .

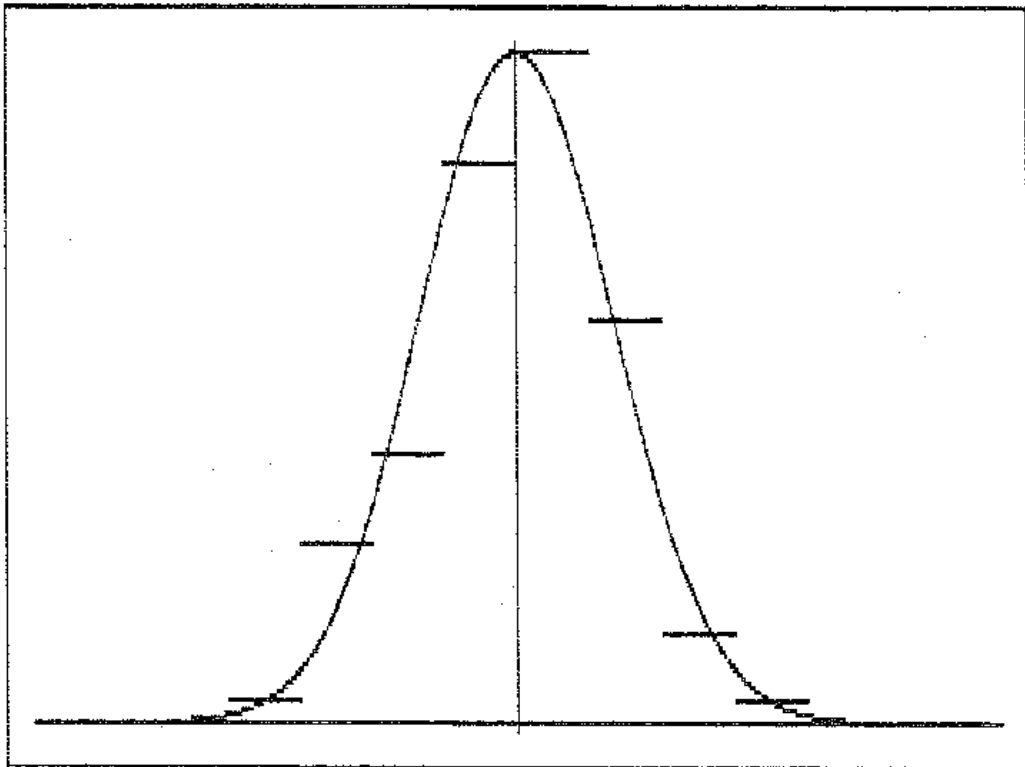


Gráfico 1.5.2 - Lei normal centrada e reduzida.

$n = 100$ ,  $EQI = 0.0119$ ,  $(-5;5) \times (0;0.4)$ .

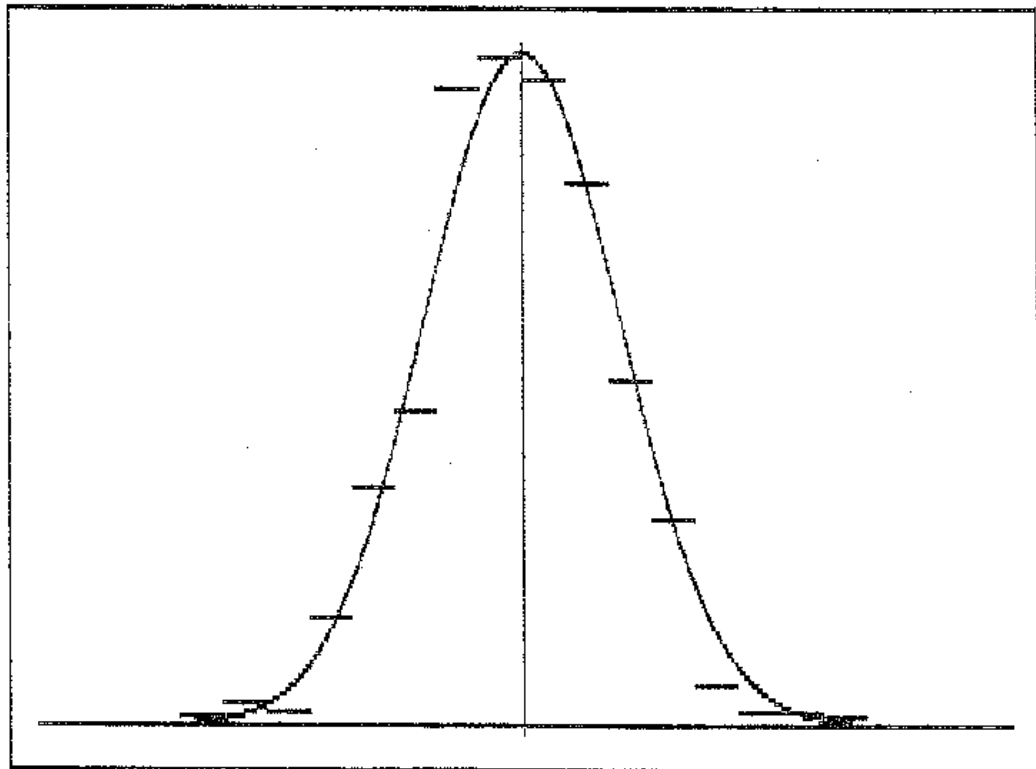


Gráfico 1.5.3 - Lei normal centrada e reduzida.

$n = 500$ ,  $EQI = 0.0051$ ,  $(-5;5) \times (0;0.399)$ .

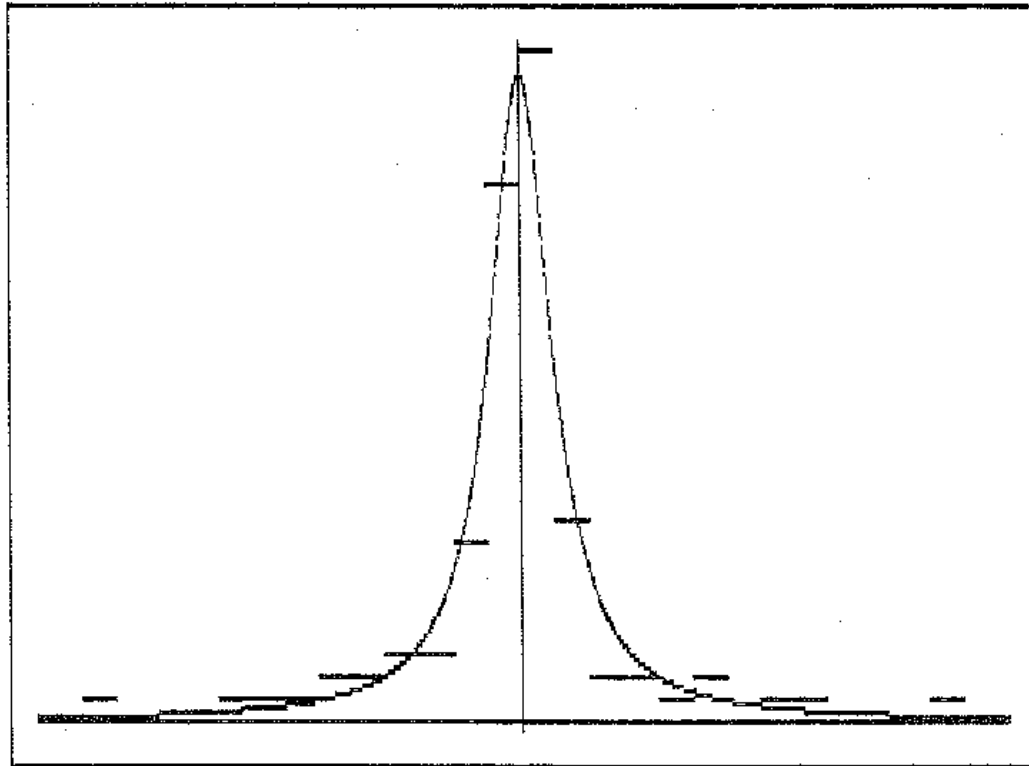


Gráfico 1.5.4 - Lei de Cauchy.

$n = 100$ ,  $EQI = 0.0131$ ,  $(-13;13) \times (0;0.33)$ .

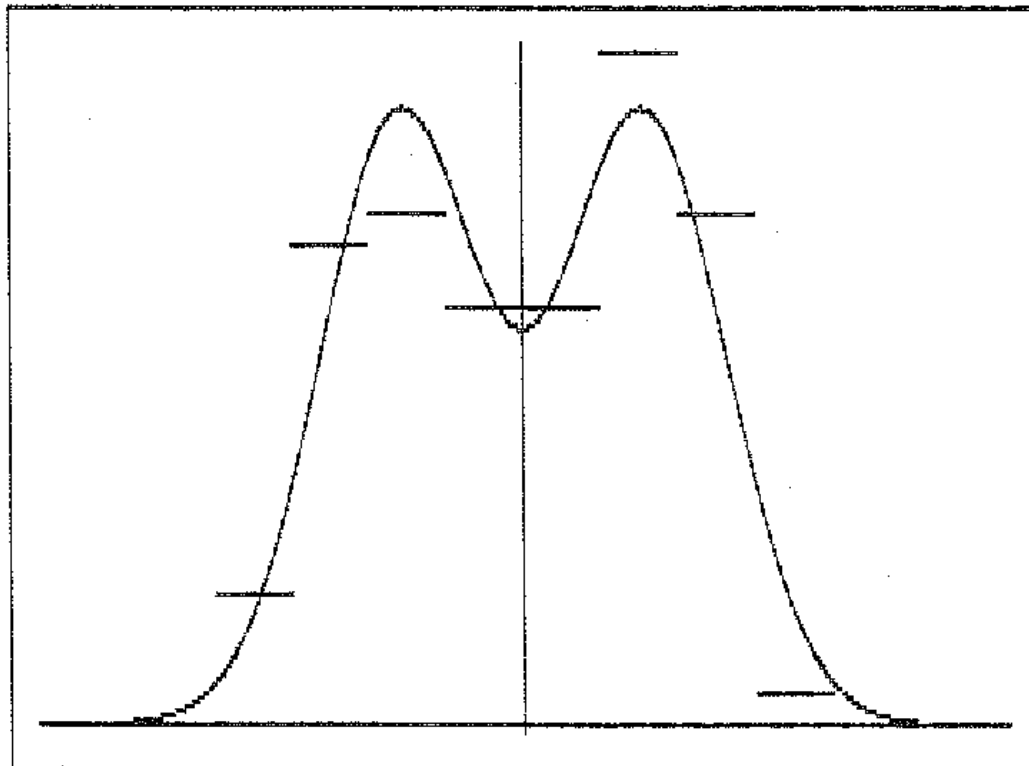


Gráfico 1.5.5 - Mistura de leis normais.

$n = 100$ ,  $EQI = 0.0073$ ,  $(-6;6) \times (0;0.22)$ .

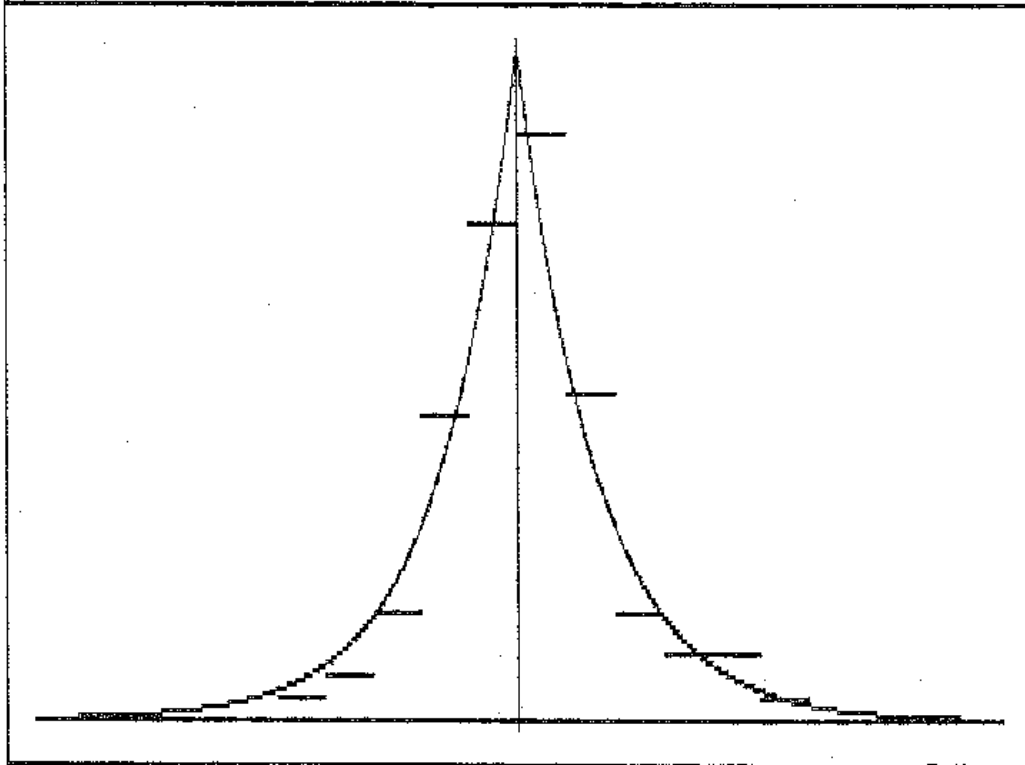


Gráfico 1.5.6 - Lei exponencial bilateral.  
 $n = 100$ ,  $EQI = 0.0136$ ,  $(-6.5; 6.5) \times (0; 0.5)$ .

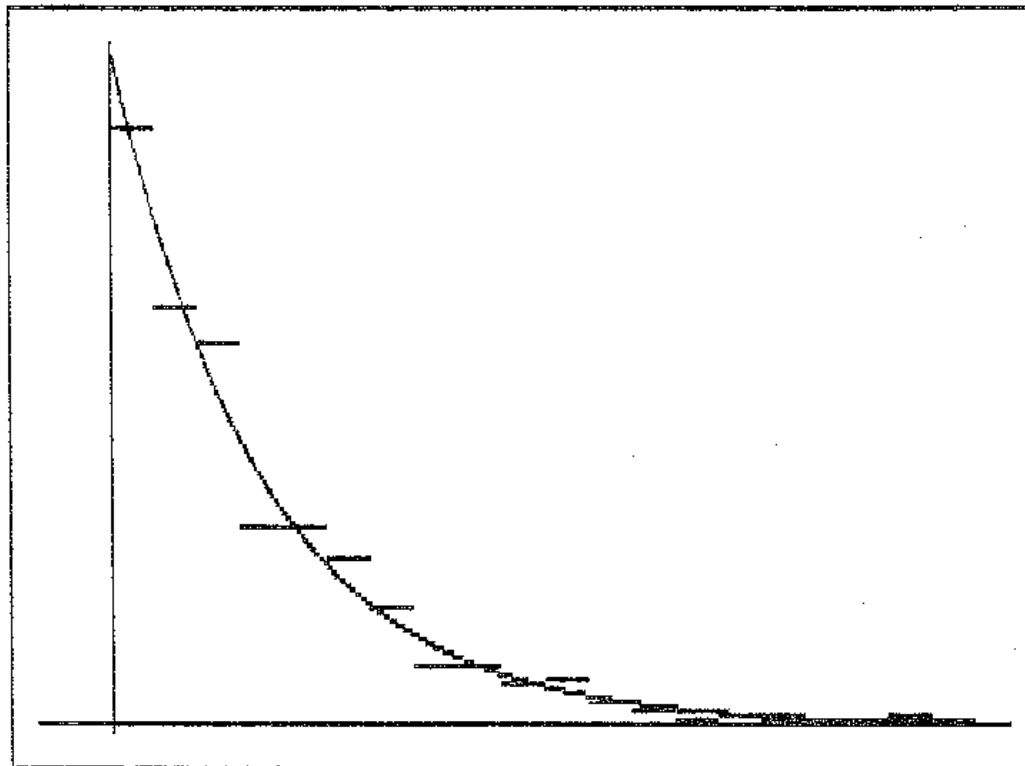


Gráfico 1.5.7 - Lei exponencial(1).  
 $n = 500$ ,  $EQI = 0.0081$ ,  $(-0.5; 6) \times (0; 1)$ .

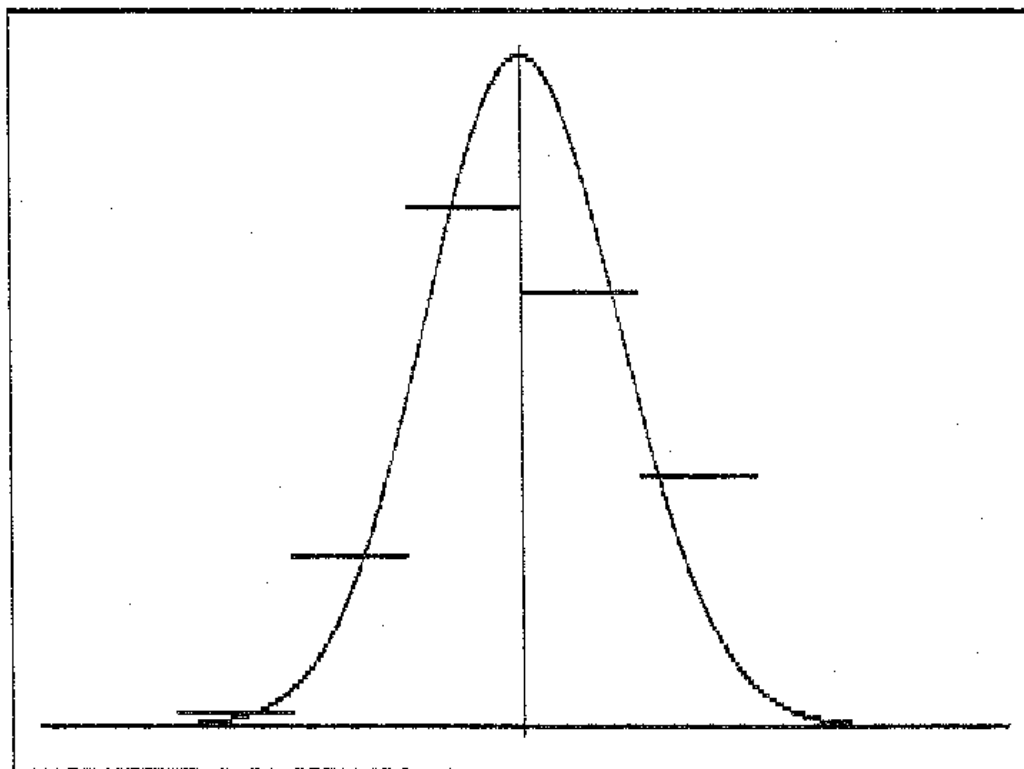


Gráfico 1.5.8 - Lei normal centrada e reduzida.

$n = 100$ ,  $h_n = 1.2$ ,  $EQI = 0.0255$ ,  $(-5;5) \times (0;0.399)$ .

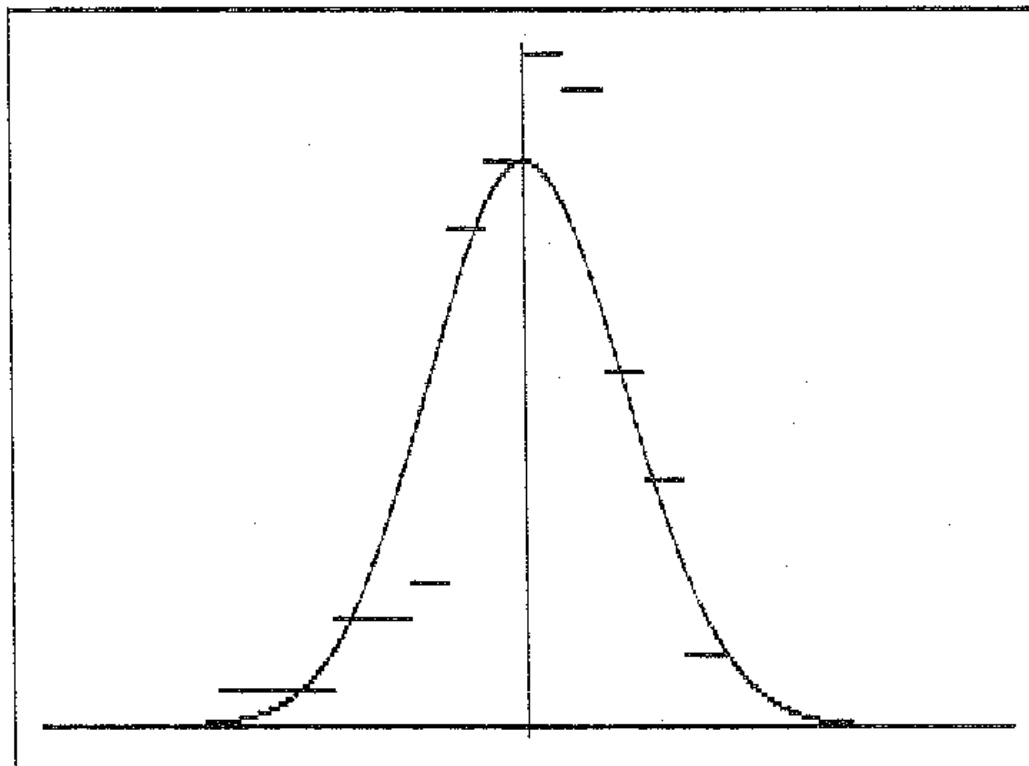


Gráfico 1.5.9 - Lei normal centrada e reduzida.

$n = 100$ ,  $h_n = 0.4$ ,  $EQI = 0.0227$ ,  $(-5;5) \times (0;0.475)$ .



## **CAPÍTULO II : O ESTIMADOR DO NÚCLEO**

A segunda parte deste trabalho centra-se na obtenção dos Teorema 2.A e 2.B, dos parágrafos 2.3. e 2.4., acerca da convergência  $L_1$  e uniforme do estimador do núcleo. O primeiro deles, mas numa forma mais fraca, foi obtido por Abou-Jaoude (1977), sendo mais tarde melhorado por Devroye (1983) tomando a forma que apresentaremos. O Teorema 2.B é de Bertrand-Retali (1978).

Analogamente ao que fizemos para o estimador do histograma, discutiremos a escolha dos parâmetros que interferem no estimador recorrendo para isso à análise da convergência em média quadrática e em média quadrática integrada. Os resultados sobre estes tipos de convergência são obtidos de forma elementar por Parzen (1962) para  $s=1$  e por Cacoullos (1966) no caso multidimensional. O estudo da escolha dos parâmetros é feita nos artigos atrás citados no caso da sucessão  $(h_n)$  e por Epanechnikov (1968) para a escolha da função  $K$ .

## 2.1. Primeiras definições e propriedades

Vamos supor ao longo deste capítulo que  $\mathfrak{X} = \mathbb{R}^s$  está munido da tribo  $\mathfrak{B}_{\mathbb{R}^s}$  dos seus borelianos, onde supomos definida a medida de Lebesgue  $\lambda$ .

**Definição 2.1.1:** Chamamos núcleo em  $\mathbb{R}^s$  a toda a aplicação  $K: (\mathbb{R}^s, \mathfrak{B}_{\mathbb{R}^s}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathfrak{B}_{\mathbb{R}}})$

integrável tal que  $\int_{\mathbb{R}^s} K(x) dx = 1$ .

Sejam  $f \in \mathfrak{F}$ ,  $(h_n)_{n \in \mathbb{N}}$  uma sucessão de números reais positivos e  $K$  um núcleo em  $\mathbb{R}^s$ .

Seja  $X^n = (X_1, X_2, \dots, X_n)$ , com  $n \in \mathbb{N}$ , uma  $n$ -amostra de lei  $P_f$ .

O estimador do núcleo de  $f$ , associado a  $K$  e a  $h_n$  é definido por :

$$f_n(x; X^n) = f_n(x) = \frac{1}{nh_n^s} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right), \text{ para } x \in \mathbb{R}^s.$$

Reparar que para cada  $x \in \mathbb{R}^s$ ,  $f_n(\cdot, x)$  não é necessariamente uma densidade de probabilidade sobre  $(\mathbb{R}^s, \mathcal{B}_{\mathbb{R}^s}, \lambda)$ .

Vamos apresentar algumas propriedades deste estimador análogas às obtidas para o estimador do histograma. Para tal necessitamos do lema seguinte devido a S. Bochner (ver Parzen, 1962, p.1067). Antes disso sendo  $h > 0$ ,  $K$  e  $g$  pertencentes a  $L_1(\mathbb{R}^s)$  definamos para quase todo o  $x \in \mathbb{R}^s$  a aplicação

$$g_h(x) = g_h(K, x) = \frac{1}{h^s} \int_{\mathbb{R}^s} K\left(\frac{x-y}{h}\right) g(y) dy.$$

Imediatamente se prova que  $g_h \in L_1(\mathbb{R}^s)$  e que sendo  $K$  limitada  $g_h$  está definida para todo o  $x \in \mathbb{R}^s$ .

**Lema 2.1.1:** *Sejam  $K, g \in L_1(\mathbb{R}^s)$  e  $K$  tal que:* i)  $\lim_{\|z\| \rightarrow \infty} \|z\|^s |K(z)| = 0$ ,  
ii)  $K$  limitada.

*Se  $g$  é contínua em  $x$  então  $\lim_{h \rightarrow 0} g_h(x) = g(x) \int_{\mathbb{R}^s} K(z) dz$ .*

*dem:*

Seja  $\delta > 0$  de modo que  $\sup_{\|z\| \leq \delta} |g(x-z) - g(x)|$  seja finito temos que

$$\begin{aligned} |g_h(x) - g(x) \int_{\mathbb{R}^s} K(z) dz| &\leq \sup_{\|z\| \leq \delta} |g(x-z) - g(x)| \int_{\mathbb{R}^s} |K(z)| dz + \\ &+ \sup_{\|z\| > \frac{\delta}{h}} \frac{\delta}{h} (\|z\|^s |K(z)|) \delta^{-s} \int_{\|z\| > \delta} |g(x-z)| dz + |g(x)| \int_{\|z\| > \frac{\delta}{h}} |K(z)| dz. \end{aligned}$$

Para concluir a demonstração basta atender à continuidade de  $g$  em  $x$  e à condição i).  $\square$

**Definição 2.1.2:** Dizemos que um núcleo  $K$  é de Parzen-Rosenblatt se  $K$  for limitada e se  $\lim_{\|z\| \rightarrow \infty} \|z\|^s |K(z)| = 0$ .

São núcleos de Parzen-Rosenblatt:

a) o núcleo normal,  $K(x) = \frac{1}{(2\pi)^{1/s}} \exp\left(-\frac{\|x\|^2}{2}\right)$ ,  $x \in \mathbb{R}^s$

b) todo o núcleo limitado com suporte limitado

c) o núcleo de Cauchy,  $K(x) = \frac{1}{\pi(1+x^2)}$ ,  $x \in \mathbb{R}$ .

Passemos então a enunciar e demonstrar as propriedades já anunciadas do estimador do núcleo, notando que o que dissemos sobre escolhas desequilibradas de  $h_n$ , aquando do estimador do histograma, mantém-se válido no estimador do núcleo.

**Proposição 2.1.1:** *Se  $K$  é um núcleo de Parzen-Rosenblatt então*

$$(\forall x \in \mathbb{R}^s \forall f \in \mathcal{F}\mathcal{C}_x \lim_n E[f_n(x)] = f(x)) \Leftrightarrow \lim_n h_n = 0$$

**dem:**

Seja  $x \in \mathbb{R}^s$  e  $f \in \mathcal{F}\mathcal{C}_x$ . Como  $E[f_n(x)] = \frac{1}{h_n^s} \int_{\mathbb{R}^s} K\left(\frac{x-y}{h_n}\right) f(y) dy$  temos pelo lema anterior,

pois  $h_n \rightarrow 0, n \rightarrow \infty$ , que  $E[f_n(x)] \rightarrow f(x), n \rightarrow \infty$ .

Suponhamos agora que  $\lim_n h_n \neq 0$ . Seja  $f(x_1, \dots, x_s) = \frac{1}{\pi^s} \prod_{i=1}^s \frac{1}{1+x_i^2}$  pertencente a  $\mathcal{F}\mathcal{C}_x$ ,

para todo o  $x \in \mathbb{R}^s$ .

Se  $h_n \rightarrow \infty, n \rightarrow \infty$ , temos pelo teorema da convergência dominada que  $\lim_n E[f_n(0)] = 0 \neq f(0)$ . Se  $h_n \rightarrow h \neq 0, n \rightarrow \infty$ , atendendo à continuidade de  $f$  e por nova aplicação do teorema da convergência dominada vem que  $\lim_n E[f_n(x)] = \int_{\mathbb{R}^s} \frac{1}{h^s} K\left(\frac{x-y}{h}\right) f(y) dy$ , para todo o  $x \in \mathbb{R}^s$ . Por hipótese e designando  $K_h(x) = \frac{1}{h^s} K\left(\frac{x}{h}\right)$  para  $x \in \mathbb{R}^s$ , temos que  $f = f * K_h$  ( $*$  - produto de convolução). Denotando por  $\phi_f$  e  $\phi_K$  as transformadas de Fourier de  $f$  e  $K$ , concluímos da igualdade anterior que para  $t \in \mathbb{R}^s$ ,  $\phi_f(t) = \phi_f(t) \phi_K(ht)$ . Pela escolha feita de  $f$   $\phi_f(t) \neq 0$  para todo o  $t \in \mathbb{R}^s$ , o que implica que  $\phi_K(t) = 1$  para todo o  $t \in \mathbb{R}^s$ , o que é falso, visto  $K$  ser um núcleo em  $\mathbb{R}^s$ .  $\square$

**Proposição 2.1.2:** *Se  $K$  é um núcleo de Parzen-Rosenblatt então*

$$(\forall x \in \mathbb{R}^s \forall f \in \mathcal{F}\mathcal{C}_x \lim_n E[f_n(x) - f(x)]^2 = 0) \Leftrightarrow (\lim_n h_n = 0, \lim_n n h_n^s = \infty)$$

**dem:**

Seja  $x \in \mathbb{R}^s$  e  $f \in \mathcal{F}\mathcal{C}_x$ . Facilmente se conclui que  $E[f_n(x) - f(x)]^2 = (E[f_n(x)] - f(x))^2 + V(f_n(x))$

(2.1.1). Como por hipótese  $\lim_n h_n = 0$ , basta pela proposição anterior provar que  $V(f_n(x)) \rightarrow 0, n \rightarrow \infty$ . Sendo

$\int_{\mathbb{R}^s} K^2(z) dz < \infty$  então pelo lema 2.1.1

$$\frac{1}{h_n^s} \int_{\mathbb{R}^s} K^2\left(\frac{x-y}{h_n}\right) f(y) dy \rightarrow f(x) \int_{\mathbb{R}^s} K^2(z) dz, n \rightarrow \infty \quad (2.1.2).$$

Por hipótese  $\lim_n nh_n^s = \infty$ , logo como  $V(f_n(x)) = \frac{1}{nh_n^s} \frac{1}{h_n^s} \int_{\mathbb{R}^s} K^2\left(\frac{x-y}{h_n}\right) f(y) dy - \frac{1}{n} E^2[f_n(x)]$

concluimos o que pretendemos, utilizando a proposição anterior e (2.1.2).

Reciprocamente, usando (2.1.1) vem, por hipótese, que para todo o  $x \in \mathbb{R}^s$  e  $f \in \mathcal{F} \mathcal{C}_x$   $E[f_n(x)] \rightarrow f(x), n \rightarrow \infty$ , o que permite afirmar pela proposição 2.1.1 que  $\lim_n h_n = 0$  e ainda que  $\lim_n V(f_n(x)) = 0$ . Assim  $\lim_n nh_n^s = \infty$  tendo em conta o facto de  $\lim_n nh_n^s V(f_n(x)) = f(x) \int_{\mathbb{R}^s} K^2(z) dz . \square$

## 2.2. Mais Mensurabilidade

Nos próximos parágrafos determinaremos condições para que  $\int_{\mathbb{R}^s} |f_n(x) - f(x)| dx \rightarrow 0, n \rightarrow \infty, \mathcal{M}$ , e  $\sup_{x \in \mathbb{R}^s} |f_n(x) - f(x)| \rightarrow 0, n \rightarrow \infty, \mathcal{M}$ , para  $f$  pertencente, em cada um dos casos, a subconjuntos convenientes de  $\mathcal{F}$ .

Para as expressões anteriores terem significado têm das aplicações

$$(\mathbb{R}^{sn}, \mathcal{B}_{\mathbb{R}^{sn}}, \otimes P_f) \xrightarrow{n} (\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+})$$

$$\mathcal{B} : (x_1, x_2, \dots, x_n) \xrightarrow{n} \int_{\mathbb{R}^s} |f_n(x) - f(x)| dx, \text{ com } f \in \mathcal{F}$$

$$\mathcal{A} : (x_1, x_2, \dots, x_n) \xrightarrow{n} \sup_{x \in \mathbb{R}^s} |f_n(x) - f(x)|, \text{ com } f \in \mathcal{F} \mathcal{C}(\mathbb{R}^s) \text{ (K limitada)}$$

ser mensuráveis para todo o  $n \in \mathbb{N}$ .

A mensurabilidade de  $\mathcal{B}$  resulta do facto de  $\lambda$  ser  $\sigma$ -finita, da aplicação  $(x, x^n) \rightarrow |f_n(x, x^n) - f(x)| = \frac{1}{nh_n^s} \sum_{i=1}^n K\left(\frac{x - x_i}{h_n}\right) - f(x)$  ser  $\mathcal{B}_{\mathbb{R}^{s(n+1)}}$  mensurável e de uma das con-

clusões do teorema de Fubini.

Quanto à aplicação  $\mathcal{A}$  nem sempre podemos garantir a sua mensurabilidade, no en-

tanto sob certas condições essa mensurabilidade pode ser garantida.

Como caso bastante importante, atendendo ao facto de  $f \in \mathcal{F} \mathcal{C}(\mathbb{R}^s)$ , temos a situação em que  $K$  é contínua em  $\mathbb{R}^s$ . Nestas condições facilmente se conclui que para todo o  $\alpha > 0$ ,

$$\{x^n \in \mathbb{R}^{sn} \mid \sup_{x \in \mathbb{R}^s} |f_n(x) - f(x)| < \alpha\} = \{x^n \in \mathbb{R}^{sn} \mid \sup_{x \in \mathbb{Q}^s} |f_n(x) - f(x)| < \alpha\}$$

bastando agora, para concluir a mensurabilidade de  $\mathcal{A}$ , ter em conta que para todo o  $x \in \mathbb{R}^s$ ,  $f_n(x)$  é  $\mathcal{B}_{\mathbb{R}^s}$ -mensurável.

A igualdade anterior pode ainda ser conseguida se admitirmos que para todo o  $x_0 \in \mathbb{R}^s$  se tem que  $\lim_{x \rightarrow x_0, x \in A_{x_0}} K(x) = K(x_0)$  onde

$$A_{x_0} = \{(x_1, \dots, x_s) \in \mathbb{R}^s \mid x_i > (\leq, \geq) x_0^i \text{ para } i=1, \dots, s\}.$$

No caso de  $s=1$  a condição anterior reduz-se a dizer que  $K$  é contínua à esquerda ou à direita.

Vamos sempre supor que a escolha de  $K$  é feita de modo que  $\sup_{x \in \mathbb{R}^s} |f_n(x) - f(x)|$  seja uma v.a..

### 2.3. Convergência $L_1$ do estimador do núcleo

Para o estudo da convergência  $L_1$  do estimador do núcleo, no sentido em que o fizemos para o estimador do histograma no parágrafo 1.3. (ver também 1.3.1), precisamos de alguns resultados auxiliares que vamos estabelecer nos lemas seguintes :

**Lema 2.3.1 :** *Sejam  $K$  um núcleo em  $\mathbb{R}^s$  e  $f \in L_1(\mathbb{R}^s)$ . Se  $\lim_n h_n = 0$  então*

$$\lim_n \int_{\mathbb{R}^s} |f_{h_n}(K, x) - f(x)| dx = 0.$$

*demo :*

Sejam  $\varepsilon > 0$  e  $f \in L_1(\mathbb{R}^s)$ , quaisquer. Existe então  $f^*$  contínua em  $\mathbb{R}^s$ , de suporte compacto

tal que  $\int_{\mathbb{R}^s} |f(x) - f^*(x)| dx < \varepsilon$ . Temos assim que sendo  $C = \int_{\mathbb{R}^s} |K(x)| dx$  então

$$\int_{\mathbb{R}^s} |f_{h_n}(K, x) - f(x)| dx \leq \varepsilon(C+1) + \int_{\mathbb{R}^s} |f^*_{h_n}(K, x) - f^*(x)| dx.$$

Basta portanto provar o resultado para  $f$  contínua de suporte compacto, que supomos con-

tido numa bola  $B(0,r)$ ,  $r > 0$ .

Dado então  $\varepsilon > 0$ , existe  $\delta(\varepsilon) > 0$  tal que, se  $\|x-y\| < \delta$  então  $|f(x) - f(y)| < \varepsilon$ .

$$\begin{aligned} \int_{\mathbb{R}^s} |f_{h_n}(K,x) - f(x)| dx &\leq \int_{\|x\| \leq r} \left| \frac{1}{h_n^s} \int_{\|y\| \leq r} K\left(\frac{x-y}{h_n}\right) (f(y) - f(x)) dy \right| dx + \\ &+ \int_{\|x\| \leq r} \left| \frac{1}{h_n^s} \int_{\|y\| > r} K\left(\frac{x-y}{h_n}\right) f(x) dy \right| dx + \int_{\|x\| > r} \left| \frac{1}{h_n^s} \int_{\|y\| \leq r} K\left(\frac{x-y}{h_n}\right) f(y) dy \right| dx \\ &< 2\varepsilon + \int_{\|x\| \leq r} \left| \frac{1}{h_n^s} \int_{\|y\| \leq r} K\left(\frac{x-y}{h_n}\right) (f(y) - f(x)) dy \right| dx \text{ para } n \text{ s.g. por} \end{aligned}$$

aplicação do teorema da convergência dominada às 2ª e 3ª parcelas do segundo membro da desigualdade anterior, depois de efectuadas convenientes mudanças de variável.

Assim, sendo  $M = \sup_{x \in \mathbb{R}^s} |f(x)|$ , temos para  $n$  s.g.  $\int_{\mathbb{R}^s} |f_{h_n}(K,x) - f(x)| dx <$

$$< 2\varepsilon + \varepsilon \int_{\|x\| \leq r} \frac{1}{h_n^s} \int_{\mathbb{R}^s} |K\left(\frac{x-y}{h_n}\right)| dy dx + 2M \int_{\|x\| \leq r} \frac{1}{h_n^s} \int_{\|x-y\| \geq \delta} |K\left(\frac{x-y}{h_n}\right)| dy dx$$

$< \varepsilon(3 + C\lambda(B(0,r)))$  por nova aplicação do teorema da convergência dominada.  $\square$

**Lema 2.3.2 :** *Sejam  $B$  um boreliano limitado de  $\mathbb{R}^s$  tal que  $\lambda(B) > 0$  e  $f \in L_1(\mathbb{R}^s)$ . Se*

$$\lim_n h_n = 0 \text{ então } \lim_n \frac{1}{\lambda(h_n B)} \int_{x+h_n B} |f(y) - f(x)| dy = 0 \text{ para quase todo o } x.$$

*dem :*

Do teorema da diferenciação de Lebesgue (ver Stein, 1970, p.4-5)

$$\lim \frac{1}{\lambda(B(x,\varepsilon))} \int_{B(x,\varepsilon)} |f(y) - f(x)| dy = 0, \varepsilon \rightarrow 0, \text{ q.c.}$$

Como  $\frac{\lambda(B)}{\lambda(h_n B)} = \frac{\lambda(B(0,r))}{\lambda(h_n B(0,r))}$  onde  $B(0,r)$  contém  $B$ , temos que

$$\frac{1}{\lambda(h_n B)} \int_{x+h_n B} |f(y) - f(x)| dy \leq \frac{\lambda(B(0,r))}{\lambda(B)} \frac{1}{\lambda(B(x,h_n r))} \int_{B(x,h_n r)} |f(y) - f(x)| dy$$

o que prova o que pretendemos.  $\square$

Seja então  $f_n$  o estimador do núcleo de  $f \in \mathfrak{F}$ , associado a  $K$  e à sucessão  $(h_n)_{n \in \mathbb{N}}$  de números reais positivos. Conclui-se imediatamente que :

$$- E[f_n(x)] = f_{h_n}(K,x), \text{ para quase todo o } x \in \mathbb{R}^s \text{ (2.3.1),}$$

$$- \int_{\mathbb{R}^s} E[f_n(x)] dx = 1,$$

$$- \int_{\mathbb{R}^s} f_n(x) dx = 1, \text{ para todo o } x^n \in \mathbb{R}^{sn},$$

- Se  $K$  é um núcleo positivo ( $K \in \mathfrak{F}$ ),  $f_n$  é uma densidade de probabilidade, para todo o  $x^n \in \mathbb{R}^{sn}$ ,

$$- \Delta_n \rightarrow 0, n \rightarrow \infty, \mathfrak{M}. \text{ sse } \begin{cases} \Delta_{n,1} \rightarrow 0, n \rightarrow \infty \\ \Delta_{n,2} \rightarrow 0, n \rightarrow \infty, \mathfrak{M}. \end{cases}$$

$$\text{onde } \Delta_n = \int_{\mathbb{R}^s} |f_n(x) - f(x)| dx, \Delta_{n,1} = \int_{\mathbb{R}^s} |E[f_n(x)] - f(x)| dx \text{ e } \Delta_{n,2} = \int_{\mathbb{R}^s} |f_n(x) - E[f_n(x)]| dx.$$

Começamos então por determinar condições n/s de convergência de  $\Delta_{n,1}$  para zero quando  $n \rightarrow \infty$ . Do lema 2.3.1 e de (2.3.1) temos :

**Teorema 2.3.1** : *Seja  $K$  um núcleo em  $\mathbb{R}^s$ . Se  $h_n \rightarrow 0, n \rightarrow \infty$  então  $\Delta_{n,1} \rightarrow 0, n \rightarrow \infty$ , para todo o  $f \in \mathfrak{F}$ .*

Para determinar condições necessárias para que  $\Delta_{n,1} \rightarrow 0, n \rightarrow \infty$ , precisamos do

**Lema 2.3.2** : *Sejam  $f, K \in \mathfrak{F}$ . Se  $h > 0$  tem-se que  $\int_{\mathbb{R}^s} |f(x) - f_h(K,x)| dx > 0$ .*

**demo** :

Para  $t \in \mathbb{R}^s$ , a função característica de  $f_h(K, \cdot)$  é dada por  $\phi_{f_h(K, \cdot)}(t) = \phi_K(ht) \phi_f(t)$ . Se

$\int_{\mathbb{R}^s} |f(x) - f_h(K,x)| dx = 0$ ,  $\phi_f = \phi_{f_h(K, \cdot)}$  e então  $\phi_f(t) = \phi_K(ht) \phi_f(t)$ , para todo o  $t \in \mathbb{R}^s$ . Como



$\phi_f$  é contínua em  $\mathbb{R}^s$  e  $\phi_f(0) = 1$ , existe uma vizinhança da origem na qual  $\phi_K = 1$ . Assim  $K$  não seria uma densidade de probabilidade.  $\square$

**Teorema 2.3.2:** Seja  $K \in \mathcal{F}$ . Se para algum  $f \in \mathcal{F}$ ,  $\Delta_{n,1} \rightarrow 0, n \rightarrow \infty$  então  $h_n \rightarrow 0, n \rightarrow \infty$ .

*dem:*

Suponhamos em primeiro lugar que  $\lim_n h_n = \infty$ . Por hipótese existe  $K^*$  contínua de

suporte compacto tal que  $\lim_{n \rightarrow \infty} \int_{\mathbb{R}^s} |f(x) - f_{h_n}(K^*, x)| dx = 0$ . Pelo lema de Fatou

$$\lim_n \int_{\mathbb{R}^s} |f(x) - f_{h_n}(K^*, x)| dx \geq \int_{\mathbb{R}^s} \underline{\lim} |f(x) - f_{h_n}(K^*, x)| dx,$$

o que implica que  $\underline{\lim} |f(x) - f_{h_n}(K^*, x)| = 0$  q.c.. Como para todo o  $x$ ,  $\lim_n f_{h_n}(K^*, x) = 0$ ,

concluimos da expressão anterior que  $f(x) = 0$  q.c., o que é falso.

Suponhamos agora que  $\lim_n h_n = c > 0$ . Como

$$\int_{\mathbb{R}^s} |f(x) - f_{h_n}(K, x)| dx \geq \int_{\mathbb{R}^s} |f(x) - f_c(K, x)| dx - \int_{\mathbb{R}^s} |f_c(K, x) - f_{h_n}(K, x)| dx,$$

basta provar, pelo lema 2.3.3, que  $\lim_n \int_{\mathbb{R}^s} |f_c(K, x) - f_{h_n}(K, x)| dx = 0$ .

Seja  $\varepsilon > 0$  e  $K^*$  contínua de suporte compacto tal que  $\int_{\mathbb{R}^s} |K(x) - K^*(x)| dx < \varepsilon$ .

$$\text{Como } \int_{\mathbb{R}^s} |f_c(K, x) - f_{h_n}(K, x)| dx \leq \int_{\mathbb{R}^s} \left| \frac{1}{c^s} K\left(\frac{x}{c}\right) - \frac{1}{h_n^s} K\left(\frac{x}{h_n}\right) \right| dx$$

$$< 2\varepsilon + \int_{\mathbb{R}^s} \left| \frac{1}{c^s} K^*\left(\frac{x}{c}\right) - \frac{1}{h_n^s} K^*\left(\frac{x}{h_n}\right) \right| dx, \text{ então sendo } K^*$$

contínua, temos, por aplicação do teorema da convergência dominada

$$< 3\varepsilon, \text{ para } n \text{ s.g.} \square$$

A partir deste momento vamos supor que  $h_n \rightarrow 0, n \rightarrow \infty$ , ou seja, que sendo  $K$  um núcleo positivo,  $\Delta_{n,1} \rightarrow 0, n \rightarrow \infty$ , para todo o  $f \in \mathcal{F}$ .

Os teoremas seguintes, estabelecem-nos condições n/s para a convergência de  $\Delta_{n,2}$  para zero quando  $n \rightarrow \infty$ , segundo o modo de convergência  $\mathfrak{M}$ .

**Teorema 2.3.3:** *Seja  $K \in \mathfrak{F}$ . Se  $\Delta_{n,2} \rightarrow 0, n \rightarrow \infty, p.$  para algum  $f \in \mathfrak{F}$ , então  $nh_n^s \rightarrow \infty, n \rightarrow \infty.$*

dem:

Sejam  $M \in \mathbb{N}$ , qualquer,  $K^*(x) = K(x) \mathbb{1}_{(K(x) \leq M)}(x)$  e  $f_n^*$  obtida substituindo em  $f_n, K$  por  $K^*$ .

$$\text{Assim } \Delta_{n,2} \geq \int_{\mathbb{R}^s} |f_n^*(x) - f_{h_n}(K^*, x)| dx - 2 \int_{\mathbb{R}^s} |K(x) - K^*(x)| dx \quad (2.3.2).$$

Seja  $L > 0$  tal que  $\int_{B(0,L)} K(z) dz > \frac{1}{2}$  e definamos o acontecimento  $A_n = \bigcap_{i=1}^n (X_i \notin B(x, h_n L))$ ;

$$\begin{aligned} \text{então } \int_{\mathbb{R}^s} E(|f_n^*(x) - f_{h_n}(K^*, x)|) dx &\geq \int_{\mathbb{R}^s} E(|f_n^*(x) - f_{h_n}(K^*, x)| \mathbb{1}_{A_n}) dx \geq U_n - V_n \text{ onde } U_n = \\ &= \int_{\mathbb{R}^s} f_{h_n}(K^*, x) P(A_n) dx \text{ e } V_n = \int_{\mathbb{R}^s} E(f_n^*(x) \mathbb{1}_{A_n}) dx \quad (2.3.3). \end{aligned}$$

Suponhamos que  $\lim_n nh_n^s = r \in [0, +\infty[$ .

Pelo lema de Fatou e pelo lema 2.3.2

$$\begin{aligned} \underline{\lim} U_n &\geq \int_{B(0,L)} K^*(y) dy \int_{\mathbb{R}^s} f(x) \underline{\lim} P(A_n) dx \\ &= \int_{B(0,L)} K^*(y) dy \int_{\mathbb{R}^s} f(x) \underline{\lim} (1 - Pf(B(x, h_n L)))^n dx \text{ por definição de } A_n \\ &\geq \int_{B(0,L)} K^*(y) dy \int_{\mathbb{R}^s} f(x) \exp(-\overline{\lim} \frac{nPf(B(x, h_n L))}{1 - Pf(B(x, h_n L))}) dx \text{ pois } 1-x \geq \exp(\frac{-x}{1-x}) \end{aligned}$$

para  $0 \leq x < 1$ .

Pelo lema 2.3.2 temos então que

$$\underline{\lim} U_n \geq \int_{B(0,L)} K^*(y) dy \int_{\mathbb{R}^s} f(x) \exp(-rL^s \lambda(B(0,1))f(x)) dx \quad (2.3.4).$$

Por outro lado  $V_n \leq \int_{\mathbb{R}^s} f(x) \int_{B^c(0,L)} K^*(y) \exp((1-n)P_f(B(yh_n+x,h_nL))) dy dx$  e como pelo lema 2.3.2

$$f(x) K^*(y) \exp((1-n)P_f(B(yh_n+x,h_nL))) \rightarrow f(x) K^*(y) \exp(-rL^s \lambda(B(0,1))f(x)), n \rightarrow \infty, q.c.,$$

vem pelo teorema da convergência dominada que

$$\overline{\lim} V_n \leq \int_{B^c(0,L)} K^*(y) dy \int_{\mathbb{R}^s} f(x) \exp(-rL^s \lambda(B(0,1))f(x)) dx \quad (2.3.5).$$

Assim por (2.3.2) e (2.3.3)

$$\begin{aligned} \underline{\lim} E(\Delta_{n,2}) + 2 \int_{\mathbb{R}^s} |K(x) - K^*(x)| dx &\geq \underline{\lim} U_n - \overline{\lim} V_n \text{ e ainda, por (2.3.4) e (2.3.5),} \\ &\geq (2 \int_{B(0,L)} K^*(y) dy - 1) \int_{\mathbb{R}^s} f(x) \exp(-rL^s \lambda(B(0,1))f(x)) dx \quad (2.3.6). \end{aligned}$$

Como  $K^*(x) \rightarrow K(x), M \rightarrow \infty$ , então por nova aplicação do teorema da convergência dominada  $\int_{B(0,L)} K^*(y) dy \rightarrow \int_{B(0,L)} K(y) dy, M \rightarrow \infty$  e  $\int_{\mathbb{R}^s} |K(x) - K^*(x)| dx \rightarrow 0, M \rightarrow \infty$ . Também, sendo  $\Delta_{n,2}$  v.a. limitada, se por hipótese  $\Delta_{n,2} \rightarrow 0, n \rightarrow \infty, p.$  então  $\lim_n E(\Delta_{n,2}) = 0$ . Assim, tendo em conta (2.3.6), concluímos que

$$0 \geq (2 \int_{B(0,L)} K(y) dy - 1) \int_{\mathbb{R}^s} f(x) \exp(-rL^s \lambda(B(0,1))f(x)) dx,$$

o que é falso pela escolha feita de L e pelo facto de  $r \in ]0, +\infty[. \square$

**Teorema 2.3.4:** *Seja  $K \in L_1(\mathbb{R}^s)$ . Se  $nh_n^s \rightarrow \infty, n \rightarrow \infty$ , então  $\Delta_{n,2} \rightarrow 0, n \rightarrow \infty, q.co.$  para*

*todo o  $f \in \mathcal{F}$ .*

**dem:**

Dado  $\varepsilon > 0$ , existe uma família finita  $A_1, A_2, \dots, A_k$  de ladrilhos em  $\mathbb{R}^s$ , e de escalares reais

$$\alpha_1, \alpha_2, \dots, \alpha_k \text{ tais que } \int_{\mathbb{R}^s} |K(x) - \sum_{i=1}^k \alpha_i \mathbb{1}_{A_i}(x)| dx < \varepsilon.$$

$$\text{Assim } \Delta_{n,2} < 2\varepsilon + \sum_{j=1}^k |\alpha_j| \int_{\mathbb{R}^s} \left| \frac{1}{nh_n^s} \sum_{i=1}^n \mathbb{1}_{A_j} \left( \frac{x-X_i}{h_n} \right) - \frac{1}{h_n^s} \int_{\mathbb{R}^s} \mathbb{1}_{A_j} \left( \frac{x-y}{h_n} \right) f(y) dy \right| dx,$$

bastando portanto provar o resultado para  $K = \mathbb{1}_P$  onde  $P = \prod_{i=1}^s [a_i, a_i + b_i]$  com

$$\prod_{i=1}^s b_i \neq 0.$$

Seja  $N \in \mathbb{N}$  tal que  $4\lambda(P) \sum_{i=1}^s b_i^{-1} < \varepsilon N$  e  $\min_{i=1,2,\dots,s} b_i > \frac{2}{N}$ . Definamos em  $\mathbb{R}^s$  a sucessão de

partições  $\Pi_n = \left( \prod_{i=1}^s \left[ \frac{q_i h_n}{N}, \frac{(q_i+1)h_n}{N} \right] \mid (q_1, q_2, \dots, q_s) \in \mathbb{Z}^s \right)$  para  $n \in \mathbb{N}$ .

Designando por  $P^* = \prod_{i=1}^s \left[ a_i + \frac{1}{N}, a_i + b_i - \frac{1}{N} \right]$  temos que para  $x \in \mathbb{R}^s$

$$C_x = (x - h_n P) \cup B \subset x - h_n (P - P^*) = C^*_x, \\ (B \in \Pi_n, B \subset x - h_n P)$$

$$\text{Assim para } K = \mathbb{1}_P, \Delta_{n,2} = \frac{1}{h_n^s} \int_{\mathbb{R}^s} \left| \frac{v_n(x - h_n P)}{n} - P_f(x - h_n P) \right| dx$$

$$\leq \frac{1}{h_n^s} \sum_{B \in \Pi_n} \left| \frac{v_n(B)}{n} - P_f(B) \right| \int_{\mathbb{R}^s} \mathbb{1}_{B \subset x - h_n P}(x) dx + \frac{1}{h_n^s} \int_{\mathbb{R}^s} \left( \frac{v_n(C^*_x)}{n} + P_f(C^*_x) \right) dx.$$

$$\text{Como } \lambda(\{x \in \mathbb{R}^s \mid B \subset x - h_n P\}) \leq h_n^s \lambda(P) \text{ e } \int_{\mathbb{R}^s} \left( \frac{v_n(C^*_x)}{n} + P_f(C^*_x) \right) dx = 2\lambda(h_n(P - P^*))$$

$$\leq 4h_n^s \frac{\lambda(P)}{N} \sum_{i=1}^s b_i^{-1} \leq \varepsilon h_n^s, \text{ pela escolha de } N, \text{ vem que}$$

$$\Delta_{n,2} \leq \lambda(P) \sum_{B \in \Pi_n} \left| \frac{v_n(B)}{n} - P_f(B) \right| + \varepsilon.$$

Como  $h_n \rightarrow 0$ ,  $n \rightarrow \infty$  e  $nh_n^s \rightarrow \infty$ ,  $n \rightarrow \infty$ , então pelo teorema 1.B concluímos o que pretendemos.  $\square$

Tendo em conta os teoremas que foram demonstrados ao longo deste parágrafo, podemos enunciar o teorema de resumo seguinte, que nos estabelece condições n/s para a convergência  $L_1$  do estimador do núcleo :

**Teorema 2.4 :** *Seja  $K \in \mathcal{F}$ , qualquer. As seguintes condições são equivalentes :*

$$i) h_n \rightarrow 0, n \rightarrow \infty; nh_n^s \rightarrow \infty, n \rightarrow \infty.$$

$$ii) \forall f \in \mathcal{F} \Delta_n \rightarrow 0, n \rightarrow \infty, \text{ q.co. (p., q.c.)}$$

$$iii) \exists f \in \mathcal{F} \Delta_n \rightarrow 0, n \rightarrow \infty, \text{ p.}$$

Este teorema fornece-nos ainda a indicação de que sendo  $K \in \mathcal{F}$ , o estimador do núcleo converge para  $f$  (no sentido  $L_1$ ) q.co.(p., q.c.) para todo o  $f \in \mathcal{F}$ , ou então não converge em probabilidade para nenhum  $f \in \mathcal{F}$ .

Não existem portanto casos intermédios, o que não acontecia no estimador do histograma. Como exemplo, podemos considerar em  $\mathbb{R}^s$  definida a sucessão de partições

$$\Pi_n = \left( \prod_{i=1}^s [a_i h_n, (a_i + 1) h_n] \mid (a_1, a_2, \dots, a_s) \in \mathbb{Z}^s \right), n \in \mathbb{N}$$

com  $h_n = 1$ , se  $n$  par e  $h_n = n^{-1/2s}$ , se  $n$  impar, que não satisfaz a condição (a). Assim, pelo teorema 1.A existe  $f \in \mathcal{F}$  tal que  $\Delta_n \not\rightarrow 0, n \rightarrow \infty, \text{ p.}$  No entanto para a densidade de probabilidade  $\mathbb{1}_{[0,1]^s}$  tem-se imediatamente que  $\Delta_n \rightarrow 0 \text{ q.co., } n \rightarrow \infty.$

## 2.4. Convergência uniforme do estimador do núcleo

Este parágrafo é, como já referimos, destinado ao estudo da convergência uniforme do estimador do núcleo. Ao longo dele chamamos núcleo em  $\mathbb{R}^s$  a toda a função que satisfaz as condições da definição 2.1.1 e além disso é limitada.

Pretendemos assim apresentar o principal resultado que nos estabelece condições para que

$$\forall f \in D \quad \Gamma_n = \sup_{x \in \mathbb{R}^s} |f_n(x) - f(x)| \rightarrow 0, n \rightarrow \infty, \mathbb{M}. \quad (2.4.1)$$

com  $D$  um subconjunto conveniente de  $\mathcal{F}$  e onde supomos que  $\lim_n h_n = 0$ .

Acerca do conjunto de densidades que estamos a considerar, é de referir que tal como no caso do estimador do histograma (ver Abou-Jaoude, 1974)

$$\exists f \in \mathcal{F} \mathcal{C}(\mathbb{R}^s) : \Gamma_n \not\rightarrow 0, n \rightarrow \infty, \text{ p.}$$

logo, para se ter (2.4.1), é necessário que  $\mathcal{F} \mathcal{C}(\mathbb{R}^s) \not\subset D$ .

No caso real, com  $h_n$  satisfazendo a condição adicional  $\sum_{n=1}^{\infty} \exp(-cnh_n^2) < \infty$ , para to-

do o  $c > 0$  e sendo  $K$  de variação limitada em  $\mathbb{R}$ , uma condição n/s para que  $\Gamma_n \rightarrow 0, n \rightarrow \infty$ , q.c. é que  $f$  seja u-continua (Schuster, 1969 e 1970). Surge assim de forma "natural" considerarmos em (2.4.1) DC  $\mathcal{D}$ .

**Observação :** A condição imposta à sucessão  $h_n$  ( $\lim_n h_n = 0$ ) é fundamental para se ter (2.4.1). Para o mostrar basta considerar, no caso de  $s=1$ ,  $h_n = \frac{1}{k(n)}$  onde  $k(n) \in \mathbb{N}$ ,  $K = \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}$

e  $f(x) = \begin{cases} 4x + 2, & \text{se } x \in [-\frac{1}{2}, 0] \\ -4x + 2, & \text{se } x \in [0, \frac{1}{2}] \\ 0, & \text{senão} \end{cases}$ . É imediato provar que se  $\lim_n h_n \neq 0$  então, por exemplo,  $f_n(\frac{1}{2}) \not\rightarrow f(\frac{1}{2}), n \rightarrow \infty$ , p..

Vamos agora iniciar o estudo atrás exposto. Começaremos por tratar o caso particular importante do núcleo  $K = \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]^s}$ . Denotemos por  $f_n^{\sim}$  o estimador do núcleo associado a  $\mathbb{1}_B$  com  $B = [-\frac{1}{2}, \frac{1}{2}]^s$  e à sucessão  $(h_n)_{n \in \mathbb{N}}$  de números reais positivos (estimador da "janela Móvel"). Assim para  $x \in \mathbb{R}^s$  e  $n \in \mathbb{N}$

$$f_n^{\sim}(x) = \frac{1}{nh_n^s} \sum_{i=1}^n \mathbb{1}_{B_{n,x}}(X_i) = \frac{1}{nh_n^s} v_{n,x} \text{ onde } B_{n,x} = x - h_n B \text{ e } v_{n,x} = \sum_{i=1}^n \mathbb{1}_{B_{n,x}}(X_i).$$

As duas proposições seguintes dão-nos condições sobre a sucessão  $(h_n)$  para a convergência de  $\Gamma_n^{\sim} = \sup_{x \in \mathbb{R}^s} |f_n^{\sim}(x) - f(x)|$  para zero quando  $n \rightarrow \infty$ , segundo os modos de convergência em p., q.c. e q.co., para todo o  $f \in \mathcal{D}$ . Sob essas condições, concluímos ainda que os três modos de convergência são equivalentes.

**Proposição 2.4.1 :** Se  $\Gamma_n^{\sim} \rightarrow 0, n \rightarrow \infty$ , p., para todo o  $f \in \mathcal{D}$ , então  $h_n^{-s} = o(\frac{n}{\log n})$ .

**dem :**

Consideremos a sucessão  $(\mathcal{P}_n)_{n \in \mathbb{N}}$  de partições de  $\mathbb{R}^s$  definida no parágrafo 1.1.. Associe-mos a esta sucessão o estimador do histograma. Portanto para todo o  $n \in \mathbb{N}$ ,  $(q_1, q_2, \dots, q_s) \in \mathbb{Z}^s$ ,  $x \in \Delta_{n, q_1, q_2, \dots, q_s}$

$$\hat{f}_n(x) = \tilde{f}_n(x^*) \text{ onde } x^* = ((q_1 + \frac{1}{2})h_n, (q_2 + \frac{1}{2})h_n, \dots, (q_s + \frac{1}{2})h_n).$$

Assim como por hipótese  $\Gamma_n \xrightarrow{\sim} 0, n \rightarrow \infty, p.$ , para todo o  $f \in \mathcal{D}$ , conclui-se que  $\hat{\Gamma}_n \xrightarrow{\sim} 0, n \rightarrow \infty, p.$ , para todo o  $f \in \mathcal{D}$ , pois dado  $\varepsilon > 0$ , qualquer e tomando  $\varepsilon_1 > 0$  de modo que  $\varepsilon_1 < \varepsilon$ , tem-se pela continuidade uniforme de  $f$ , que para  $n$  s.g.  $P_f(\hat{\Gamma}_n \geq \varepsilon) \leq P_f(\tilde{\Gamma}_n \geq \varepsilon_1)$ .

Do teorema 1.C concluímos o que pretendemos.  $\square$

**Proposição 2.4.2:** Se  $h_n^{-s} = o(\frac{n}{\log n})$  então  $\Gamma_n \xrightarrow{\sim} 0, n \rightarrow \infty, q. co.,$  para todo o  $f \in \mathcal{D}$ .

*dem:*

Seja  $(k_n)$  uma sucessão de números reais positivos tal que  $\lim_n k_n = \infty, k_n^{-s} = o(\frac{n}{\log n})$  e  $\lim_n h_n k_n = \infty$  (basta tomar por exemplo  $k_n = \frac{1}{h_n (\varepsilon_n)^{1/2s}}$  onde  $h_n^{-s} = \frac{n \varepsilon_n}{\log n}$ ).

Consideremos a sucessão de partições  $(B_n)_{n \in \mathbb{N}}$  de  $\mathbb{R}^s$  definida por

$$B_n = (B_{n,r_1, r_2, \dots, r_s} \mid (r_1, r_2, \dots, r_s) \in \mathbb{Z}^s), \text{ onde } B_{n,r_1, r_2, \dots, r_s} = \prod_{i=1}^s [\frac{r_i-1}{k_n}, \frac{r_i}{k_n}].$$

Dado  $x = (x_1, x_2, \dots, x_s) \in \mathbb{R}^s$  ponhamos para  $i=1, 2, \dots, s$

$$\rho_i = \rho_i(x, n) = \min \{ r \in \mathbb{Z} \mid \frac{r-1}{k_n} \geq x_i - \frac{h_n}{2} \}$$

$$\rho'_i = \rho'_i(x, n) = \max \{ r \in \mathbb{Z} \mid \frac{r}{k_n} \leq x_i + \frac{h_n}{2} \}.$$

Como  $h_n k_n \rightarrow \infty, n \rightarrow \infty$ , existe uma ordem  $n_0$  a partir da qual  $h_n k_n \geq 3$  e portanto a partir de certa ordem para todo o  $x \in \mathbb{R}^s$  e  $i=1, 2, \dots, s$   $\rho'_i(x, n) > \rho_i(x, n)$ . Assim para todo o  $x \in \mathbb{R}^s$ ,

$$\bigcup_{r_1=\rho_1}^{\rho'_1} \dots \bigcup_{r_s=\rho_s}^{\rho'_s} B_{n,r_1, \dots, r_s} \subset B_{n,x} \subset \bigcup_{r_1=\rho_1-1}^{\rho'_1+1} \dots \bigcup_{r_s=\rho_s-1}^{\rho'_s+1} B_{n,r_1, \dots, r_s}$$

e ainda, sendo  $v_{n,r_1, \dots, r_s} = \sum_{i=1}^n \mathbb{1}_{B_{n,r_1, \dots, r_s}}(X_i)$ ,

$$\sum_{r_1=\rho_1}^{\rho'_1} \dots \sum_{r_s=\rho_s}^{\rho'_s} v_{n,r_1, \dots, r_s} \leq v_{n,x} \leq \sum_{r_1=\rho_1-1}^{\rho'_1+1} \dots \sum_{r_s=\rho_s-1}^{\rho'_s+1} v_{n,r_1, \dots, r_s}$$

Por outro lado como para todo o  $n \in \mathbb{N}, x \in \mathbb{R}^s$  e  $i=1, 2, \dots, s$   $\rho'_i - \rho_i + 1 \leq h_n k_n$  e  $\rho'_i - \rho_i + 3 \geq h_n k_n$

(2.4.2) concluímos que  $k_n^{-s} \prod_{i=1}^s (\rho'_i - \rho_i + 1) \leq h_n^s \leq k_n^{-s} \prod_{i=1}^s (\rho'_i - \rho_i + 3)$ .

Assim, para  $n$  s.g.

$$\frac{\sum_{r_1=\rho_1}^{\rho'_1} \dots \sum_{r_s=\rho_s}^{\rho'_s} \frac{k_n^s v_{n,r_1,\dots,r_s}}{n}}{\prod_{i=1}^s (\rho'_i - \rho_i + 3)} \leq f_n^{\sim}(x) \leq \frac{\sum_{r_1=\rho_1-1}^{\rho'_1+1} \dots \sum_{r_s=\rho_s-1}^{\rho'_s+1} \frac{k_n^s v_{n,r_1,\dots,r_s}}{n}}{\prod_{i=1}^s (\rho'_i - \rho_i + 1)}$$

para todo o  $x \in \mathbb{R}^s$  (2.4.3).

Seja  $f_n$  o estimador do histograma associado à sucessão de partições  $(B_n)_{n \in \mathbb{N}}$  de  $\mathbb{R}^s$ . Dado  $\eta > 0$ , qualquer, designemos, por  $E_{n,\eta}$  o acontecimento  $(\Gamma_n \leq \eta)$  e suponhamos que para  $n$  s.g.,  $E_{n,\eta}$  se realiza. Assim  $|f_n(u) - f(u)| \leq \eta$ , para todo o  $u \in \mathbb{R}^s$ .

Para  $x \in \mathbb{R}^s$  e  $u \in B_{n,r_1,\dots,r_s}$  com  $r_i = \rho_i - 1, \dots, \rho_i + 1$ ,  $i=1,2,\dots,s$ ,  $\|x - u\| \leq \sqrt{s}(h_n + k_n^{-1})$  e portanto atendendo à  $u$ -continuidade de  $f$  concluímos que para  $n$  s.g.  $|f_n(u) - f(x)| < 2\eta$ .

Tendo em conta (2.4.3) e a desigualdade anterior temos que para  $n$  s.g.

$$\frac{\prod_{i=1}^s (\rho'_i - \rho_i + 1)}{\prod_{i=1}^s (\rho'_i - \rho_i + 3)} (f(x) - 2\eta) \leq f_n^{\sim}(x) \leq \frac{\prod_{i=1}^s (\rho'_i - \rho_i + 3)}{\prod_{i=1}^s (\rho'_i - \rho_i + 1)} (f(x) + 2\eta)$$

para todo o  $x \in \mathbb{R}^s$ .

Seja  $\frac{\prod_{i=1}^s (\rho'_i - \rho_i + 3)}{\prod_{i=1}^s (\rho'_i - \rho_i + 1)} \leq 2^s$  e tomando  $H > \sup_{x \in \mathbb{R}^s} f(x)$  vem que

$$-2\eta + H \frac{\prod_{i=1}^s (\rho'_i - \rho_i + 1) - \prod_{i=1}^s (\rho'_i - \rho_i + 3)}{\prod_{i=1}^s (\rho'_i - \rho_i + 3)} \leq f_n^{\sim}(x) - f(x) \leq 2^{s+1}\eta + H \frac{\prod_{i=1}^s (\rho'_i - \rho_i + 3) - \prod_{i=1}^s (\rho'_i - \rho_i + 1)}{\prod_{i=1}^s (\rho'_i - \rho_i + 1)}$$

para todo o  $x \in \mathbb{R}^s$ .



Mas 
$$\frac{\prod_{i=1}^s (\rho'_i - \rho_{i+3}) - \prod_{i=1}^s (\rho'_i - \rho_{i+1})}{\prod_{i=1}^s (\rho'_i - \rho_{i+1})} = \prod_{i=1}^s \left(1 + \frac{2}{\rho'_i - \rho_{i+1}}\right) - 1$$
, 
$$\frac{\prod_{i=1}^s (\rho'_i - \rho_{i+3}) - \prod_{i=1}^s (\rho'_i - \rho_{i+1})}{\prod_{i=1}^s (\rho'_i - \rho_{i+3})} =$$
  

$$= 1 - \prod_{i=1}^s \left(1 - \frac{2}{\rho'_i - \rho_{i+3}}\right)$$
 e por (2.4.2) existe uma ordem a partir da qual  $H\left[\prod_{i=1}^s \left(1 + \frac{2}{\rho'_i - \rho_{i+1}}\right) - 1\right]$   

$$< \eta$$
 e  $H\left[1 - \prod_{i=1}^s \left(1 - \frac{2}{\rho'_i - \rho_{i+3}}\right)\right] < \eta$  para todo o  $x \in \mathbb{R}^s$ .

Das expressões obtidas acima concluímos finalmente que para  $n$  s.g.

$$E_n \eta \subset \{\Gamma_n^{\sim} \leq (2^{s+1} + 1)\eta\}.$$

Portanto para  $n$  s.g.  $P_f(\{\Gamma_n^{\sim} > (2^{s+1} + 1)\eta\}) \leq P_f(\{\Gamma_n^{\sim} > \eta\})$  e como  $k_n^s = o\left(\frac{n}{\log n}\right)$  temos pelo

teorema 1.C o resultado pretendido.  $\square$

Passemos agora ao estudo do caso geral em que  $K$  é um qualquer núcleo em  $\mathbb{R}^s$  (ao qual imporemos mais tarde certas condições), começando por demonstrar condições necessárias de convergência uniforme em probabilidade de  $f_n$  para  $f$ , com  $f \in \mathcal{D}$ , expressas no lema (a mais fraca) e teorema seguintes :

**Lema 2.4.1:** *Se para todo o  $f \in \mathcal{D}$ ,  $\Gamma_n \rightarrow 0$ ,  $n \rightarrow \infty$ , p. então  $h_n^{-s} = o(n)$ .*

**dem:**

Suponhamos que existe  $\alpha \in ]0,1[$  e um subconjunto infinito  $N_0$  de  $\mathbb{N}$ , tal que para todo o

$$n \in N_0, \frac{h_n^{-s}}{n} > \alpha \text{ e seja } \Sigma > 0 \text{ satisfazendo } \int_{\mathbb{R}^s - \left[-\frac{\Sigma}{2}, \frac{\Sigma}{2}\right]^s} |K(x)| dx \leq \frac{1}{2} \quad (2.4.4).$$

Fazendo  $\rho_n = \Sigma h_n$  vem para todo o  $n \in N_0$   $\frac{1}{\rho_n^s} > \frac{\alpha n}{\Sigma^s}$ . Seja  $Q_n = [0, \rho_n]^s$  e  $f \in \mathcal{D}$ , satisfazendo

$$f(x) = \alpha, \text{ para } x \in [0,1]^s \text{ e } \sup_{x \in \mathbb{R}^s} f(x) = \alpha.$$

Sendo  $(X_1, X_2, \dots, X_n)$  uma  $n$ -amostra de lei  $P_f$ , definamos o acontecimento  $H_n = \bigcap_{i=1}^n (X_i \notin Q_n)$ .

Como  $p_n = P(H_n) = (1 - \alpha \rho_n^s)^n$ , pois a partir de certa ordem  $[0, \rho_n]^s \subset [0,1]^s$ , vem que para

todo o  $n \in \mathbb{N}_0$ ,  $p_n > (1 - \frac{\sum^s}{n})^n$  e portanto  $\overline{\lim} p_n \geq \exp(-\sum^s)$  (2.4.5).

Determinemos a densidade da lei condicional de cada  $X_j$ , com  $j=1,2,\dots,n$ , relativamente ao acontecimento  $H_n$ :

$$X_j : (\Omega, \mathcal{A}, P^{H_n}) \longrightarrow (\mathbb{R}^s, \mathfrak{B}_{\mathbb{R}^s}, (P^{H_n})_{X_j}),$$

como para todo o  $A \in \mathfrak{B}_{\mathbb{R}^s}$  e  $j=1,2,\dots,n$ ,  $P_{X_j}^{H_n}(A) \leq \frac{P_f(A)}{P(H_n)}$  então  $(P^{H_n})_{X_j} \ll \lambda$ , sendo como facil-

mente se conclui  $f_{H_n}(x) = \frac{f(x)}{1 - \alpha p_n^s} (1 - \mathbb{1}_{Q_n}(x))$  uma versão da derivada de Radon-Nikodym de

$(P^{H_n})_{X_j}$  relativamente a  $\lambda$ .

$$\begin{aligned} \text{Sendo } \varepsilon_n = (\frac{\rho_n}{2}, \frac{\rho_n}{2}, \dots, \frac{\rho_n}{2}) \text{ então } E(|f_n(\varepsilon_n)| | H_n) &\leq \frac{1}{1 - \alpha p_n^s} \int_{\mathbb{R}^s - [-\frac{\sum^s}{2}, \frac{\sum^s}{2}]^s} |K(v)| f(\varepsilon_n - h_n v) dv \\ &\leq \frac{\alpha}{2(1 - \alpha p_n^s)} \text{ por (2.4.4).} \end{aligned}$$

Usando a desigualdade de Markov, obtemos que

$$P_f(|f_n(\varepsilon_n)| > \frac{3\alpha}{4(1 - \alpha p_n^s)} | H_n) \leq \frac{4(1 - \alpha p_n^s)}{3\alpha} E(|f_n(\varepsilon_n)| | H_n) \leq \frac{2}{3}$$

e então  $P_f(\Gamma_n < \alpha - \frac{3\alpha}{4(1 - \alpha p_n^s)} | H_n) \leq \frac{2}{3}$ . Como para  $n$  s.g.  $1 - \alpha p_n^s > \frac{7}{8}$ ,  $P_f(\Gamma_n < \frac{\alpha}{7} | H_n) \leq \frac{2}{3}$

e por (2.4.5)  $\overline{\lim} P_f(\Gamma_n \geq \frac{\alpha}{7}) \geq \frac{1}{3} \exp(-\sum^s)$ , não convergindo assim  $f_n$  para  $f$ , uniformemente em probabilidade.  $\square$

Na demonstração do teorema que se segue incluímos os enunciados de lemas auxiliares cuja prova apresentamos no fim da demonstração do teorema.

**Teorema 2.4.1:** *Se para todo o  $f \in \mathcal{D}$ ,  $\Gamma_n \rightarrow 0$ ,  $n \rightarrow \infty$ , p. então  $h_n^{-s} = o(\frac{n}{\log n})$ .*

*Dem:*

Do lema 2.4.1 sabemos que nas condições da hipótese  $h_n^{-s} = o(n)$ . Suponhamos que existe  $\beta > 0$

e um subconjunto infinito  $N_0$  de  $\mathbb{N}$ , tais que para todo o  $n \in N_0$ ,  $\frac{h_n^{-s} \log n}{n} > \beta$ .

Sejam  $\sum > 0$  satisfazendo (2.4.4),  $k_n = \lfloor \frac{1}{\sum h_n} \rfloor$  e  $\rho_n = \frac{1}{k_n}$ . Temos que para  $n$  s.g.  $\sum h_n \leq \rho_n \leq \frac{\sum h_n}{1 - \sum h_n}$  (2.4.6) e existem  $\beta' > 0$  e  $N \subset \mathbb{N}_0$ , infinito, tais que  $k_n^s > \frac{\beta' n}{\log n}$  para todo o  $n \in N$  (2.4.7).

Seja  $\alpha \in ]0, \frac{\beta'}{2}[ \cap ]0, 1[$  e  $f \in \mathcal{D}$  tal que  $f(x) = \alpha$ , para  $x \in ]0, 1[$  e  $\sup_{x \in \mathbb{R}^s} f(x) = \alpha$  (2.4.8).

Consideremos a sucessão de partições de  $]0, 1[$  em borelianos da forma

$$Q_{n,t} = \prod_{i=1}^s [(r_i-1)\rho_n, r_i\rho_n[ \text{ onde } t = \psi((r_1, r_2, \dots, r_s)),$$

sendo  $\psi$  uma aplicação bijectiva de  $\{1, 2, \dots, k_n\}^s$  em  $\{1, 2, \dots, k_n^s\}$ .

Seja  $(X_1, X_2, \dots, X_n)$  uma  $n$ -amostra de lei  $P_f$ , designemos por  $V_n = \sum_{i=1}^n \mathbf{1}_{]0, 1[}^s(X_i)$ .

Sejam  $\alpha'$  e  $\alpha''$  números reais tais que  $0 < \alpha' < \alpha < \alpha'' < 1$  e designemos por  $A_n$  o acontecimento  $\{ \alpha' < \frac{V_n}{n} < \alpha'' \}$ .

**Lema 2.4.2:**  $\lim_n P_f(A_n) = 1$ .

Vamos designar por  $N_n$  a v.a.  $\sum_{t=1}^{k_n^s} \mathbf{1}_{\bigcap_{i=1}^n (X_i \notin Q_{n,t})}$  ( $N_n$  dá-nos o número de borelianos

$Q_{n,t}$  ( $t=1, 2, \dots, k_n^s$ ) que não contêm nenhum ponto da  $n$ -amostra).

**Lema 2.4.3:**  $\forall \varepsilon > 0 \quad \overline{\lim} P_f(1 \leq N_n \leq \varepsilon k_n^s) = 1$ .

Seja  $j \in \{1, 2, \dots, k_n^s\}$  qualquer. Tomemos  $t_1, t_2, \dots, t_j$  inteiros tais que  $1 \leq t_1 < t_2 < \dots < t_j \leq k_n^s$  e seja

$V_n(t_1, t_2, \dots, t_j)$  o acontecimento :

Cada um dos borelianos  $Q_{n,t}$  com  $t=t_1, t_2, \dots, t_j$  não contém nenhum ponto da  $n$ -amostra e cada  $Q_{n,t}$  com  $t \neq t_1, t_2, \dots, t_j$  contém pelo menos um ponto da  $n$ -amostra.

Temos então que  $(N_n = j) = \bigcup_{\substack{t_1, t_2, \dots, t_j \\ 1 \leq t_1 < t_2 < \dots < t_j \leq k_n^s}} V_n(t_1, t_2, \dots, t_j)$ .

Seja  $\varepsilon > 0$ , qualquer, vamos denotar por  $B_n$  o acontecimento  $\{1 \leq N_n \leq \varepsilon k_n^s\}$ . Assim

$$B_n = \bigcup_{j=1}^{\lfloor \varepsilon k_n^s \rfloor} \bigcup_{\substack{t_1, t_2, \dots, t_j \\ 1 \leq t_1 < t_2 < \dots < t_j \leq k_n^s}} V_n(t_1, t_2, \dots, t_j).$$

Supomos  $n$  s.g. de modo que  $k_n^s \geq \alpha' n$  e  $\alpha'' n - \alpha' n \geq 1$ . Tomemos  $\nu$  um inteiro satisfazendo  $\lfloor \alpha' n \rfloor + 1 \leq \nu \leq \alpha'' n$  (2.4.9).

**Lema 2.4.4:** A lei condicional de  $X_i$  ( $i=1, 2, \dots, n$ ) relativamente ao acontecimento

$$e_n = e_n(\nu; t_1, t_2, \dots, t_j) = \{v_n = \nu\} \cap V_n(t_1, t_2, \dots, t_j)$$

é absolutamente contínua e tem por densidade

$$f_n^{e_n}(x) = \begin{cases} \frac{n-\nu}{n} \frac{f(x)}{1-\alpha} & \text{se } x \in [0, 1]^s \\ \frac{\nu}{n} \frac{1}{1-j\rho_n^s} & \text{se } x \in [0, 1]^s - \bigcup_{r=1}^j Q_{n,t_r} \\ 0 & \text{se } x \in \bigcup_{r=1}^j Q_{n,t_r} \end{cases}$$

Concluamos agora a demonstração provando que para  $f$  definida em (2.4.8)  $f_n$  não converge uniformemente em probabilidade para  $f$ .

Seja  $\varepsilon \in ]0, \frac{1}{2}[$  e tomemos  $\alpha'$  e  $\alpha''$  suficientemente próximos de  $\alpha$  de modo que

$$\frac{1-\alpha'}{1-\alpha} < (1+2\varepsilon) \quad \text{e} \quad \frac{\alpha''}{\alpha} < (1+2\varepsilon)(1-\varepsilon).$$

Seja  $n$  s.g. de modo que se tenha (2.4.6) e (2.4.9). Então pelo lema 2.4.4 para todo o  $x \in \mathbb{R}^s$

$$E(|f_n(x)| | e_n) \leq \frac{1}{h_n^s} \int_I |K(\frac{x-u}{h_n})| \frac{n-\nu}{n(1-\alpha)} f(u) du + \frac{1}{h_n^s} \int_{I'} |K(\frac{x-u}{h_n})| \frac{\nu}{n\alpha(1-j\rho_n^s)} f(u) du$$

onde  $I = \mathbb{R}^s - [0, 1]^s$  e  $I' = [0, 1]^s - \bigcup_{r=1}^j Q_{n,t_r}$ .

Seja  $j$  um inteiro tal que  $1 \leq j \leq \varepsilon k_n^s$  (o que é possível para  $n$  s.g.).

De (2.4.9)  $\frac{n-v}{n} \leq (1-\alpha')$ . Por outro lado  $1 - j\rho_n^s \geq 1-\varepsilon$ , então, como  $\frac{v}{n} \leq \alpha''$ , vem que

$$\frac{v}{n\alpha(1-j\rho_n^s)} \leq \frac{\alpha''}{\alpha(1-\varepsilon)}$$

Assim  $E(|f_n(x)| | e_n) \leq \frac{1}{h_n^s} \int_{|U|} |K(\frac{x-u}{h_n})| (1+2\varepsilon) f(u) du$ . Tomando para  $x$  o ponto médio.

$x_n$  de  $Q_{n,t_1}$  temos que  $E(|f_n(x_n)| | e_n) \leq (1+2\varepsilon) \int_{I''} |K(v)| f(x_n-vh_n) dv$  onde  $I'' = T(|U|)$

com  $T(u) = \frac{x_n-u}{h_n}$  para  $u \in \mathbb{R}^s$ . Como  $(|U|) \cap Q_{n,t_1} = \emptyset$ , também  $I'' \cap T(Q_{n,t_1}) = \emptyset$ . Sendo  $T(Q_{n,t_1}) =$

$[-\frac{\rho_n}{2h_n}, \frac{\rho_n}{2h_n}]^s$  concluímos por (2.4.6) que  $I'' \cap [-\frac{\rho_n}{2}, \frac{\rho_n}{2}]^s = \emptyset$  e portanto

$$E(|f_n(x_n)| | e_n) \leq \frac{1+2\varepsilon}{2} \alpha \quad \text{tendo em conta (2.4.4) e (2.4.8).}$$

Tomando  $\varepsilon = \frac{1}{6}$ ,  $E(|f_n(x_n)| | e_n) \leq \frac{2\alpha}{3}$ . Tendo em conta a desigualdade de Markov

$$P_f(\Gamma_n > \frac{\alpha}{4} | e_n) \geq \frac{1}{9}$$

Finalmente

$$\begin{aligned} P_f((\Gamma_n > \frac{\alpha}{4}) \cap A_n \cap B_n) &= \sum_{v=[\alpha'n]+1}^{[\alpha'n]} \sum_{j=1}^{[k_n^s]} \sum_{t_1, t_2, \dots, t_j} P_f(\Gamma_n > \frac{\alpha}{4} | e_n) P_f(e_n(v; t_1, t_2, \dots, t_j)) \\ &\quad [1 \leq t_1 < t_2 < \dots < t_j < k_n^s] \\ &\geq \frac{1}{9} P_f(A_n \cap B_n). \text{ Por maioria de razão } P_f(\Gamma_n > \frac{\alpha}{4}) \geq \frac{1}{9} P_f(A_n \cap B_n) \end{aligned}$$

e portanto  $\overline{\lim} P_f(\Gamma_n > \frac{\alpha}{4}) \geq \frac{1}{9}$  tendo em conta os lemas 2.4.2 e 2.4.3.  $\square$

dem: (do lema 2.4.2)

Pela lei fraca dos grandes números  $\frac{v_n}{n} \rightarrow \alpha$ ,  $n \rightarrow \infty$ , p., ou seja,

$$\forall \varepsilon > 0 \quad P_f(|\frac{v_n}{n} - \alpha| < \varepsilon) \rightarrow 1, n \rightarrow \infty.$$

Sendo  $\varepsilon_0 > 0$  tal que  $\alpha' < \alpha - \varepsilon_0$  e  $\alpha + \varepsilon_0 < \alpha''$  então

$$P_f(\alpha' < \frac{v_n}{n} < \alpha'') \geq P_f(|\frac{v_n}{n} - \alpha| < \varepsilon_0). \square$$

dem : (do lema 2.4.3)

Provemos que  $\overline{\lim} \text{Pf}(N_n \geq 1) = 1$ .

Para  $t = 1, 2, \dots, k_n^s$  seja  $V_n(t) = \bigcap_{i=1}^n (X_i \notin Q_{n,t})$ .  $\text{Pf}(N_n \geq 1) = \text{Pf}(\bigcup_{j=1}^{k_n^s} V_n(t))$ . Assim por

razões análogas às apresentadas na demonstração do Teorema 1.4.1

$$\text{Pf}(N_n=0) = \prod_{t=1}^{k_n^s} \text{Pf}(V_n(t)^c) = \prod_{t=1}^{k_n^s} \text{Pf}\left(\bigcap_{i=1}^{t-1} V_n(i)^c\right) \leq \prod_{t=1}^{k_n^s} \text{Pf}(V_n(t)^c).$$

Como pela definição de  $f$ ,  $\text{Pf}(V_n(t)^c) = 1 - (1 - \alpha \rho_n^s)^n$ , para todo o  $t = 1, 2, \dots, k_n^s$ , concluímos da expressão anterior que  $\log(-\log \text{Pf}(N_n=0)) \geq \log k_n^s + n \log(1 - \alpha \rho_n^s)$ .

Para  $x$  suficientemente próximo de zero,  $\log(1-x) \geq -2x$  e portanto para  $n \in \mathbb{N}$  s.g. temos por (2.4.7) que

$$\log(-\log \text{Pf}(N_n=0)) \geq \log k_n^s - \frac{2n\alpha}{k_n^s} > \left(1 - \frac{2\alpha}{\beta'} + \frac{\log \beta' - \log \log n}{\log n}\right) \log n.$$

Pela escolha de  $\alpha$  em (2.4.8)  $\overline{\lim} \log(-\log \text{Pf}(N_n=0)) = +\infty$ , ou seja,  $\overline{\lim} \text{Pf}(N_n \geq 1) = 1$ .

Provemos agora que  $\forall \varepsilon > 0 \quad \lim \text{Pf}(N_n \leq \varepsilon k_n^s) = 1$ .

$E(N_n) = k_n^s (1 - \alpha \rho_n^s)^n$ . Como  $h_n^{-s} = o(n)$ ,  $k_n^s = \xi_n n$ , com  $\xi_n \rightarrow 0$ ,  $n \rightarrow \infty$ , (por (2.4.6)) e portanto  $\lim_n E\left(\frac{N_n}{k_n^s}\right) = 0$ . Sendo limitada a v.a.  $\frac{N_n}{k_n^s}$ , temos que  $\frac{N_n}{k_n^s} \rightarrow 0$ ,  $n \rightarrow \infty$ , p.□

dem : (do lema 2.4.4)

Seja  $A$  um boreliano de  $\mathbb{R}^s = [0,1]^s$  temos que  $\text{Pf}(\{X_1 \in A\} \cap \{e_n\}) = \text{Pf}(A) \text{Pf}(e_n | X_1 \in A)$ .

A realização do acontecimento  $\{e_n | X_1 \in A\}$  significa que

$$\left\{ \begin{array}{l} \forall \text{ pontos de } X_2, X_3, \dots, X_n \text{ pertencem a } [0,1]^s \\ \text{nenhum ponto pertence a } \bigcup_{r=1}^j Q_{n,t_r} \\ \text{pelo menos um ponto pertence a } Q_{n,t} \text{ para } t \in \{t_1, t_2, \dots, t_j\} \end{array} \right.$$

sendo portanto independente do conjunto A tomado em  $\mathbb{R}^s - [0,1]^s$ . Assim  $P_f((X_1 \in A) | e_n) = CP_f(A)$ , com C independente de A. Como  $P_f((X_1 \in \mathbb{R}^s - [0,1]^s) | e_n) = \frac{n-\nu}{n}$  concluímos da

expressão anterior que  $P_f((X_1 \in A) | e_n) = \int_A \frac{n-\nu}{n(1-\alpha)} f(x) dx$ .

Seja agora  $t \in \{1, 2, \dots, k_n^s\} - \{t_1, t_2, \dots, t_j\}$  e A um boreliano contido em  $Q_{n,t}$ . O acontecimento  $(e_n | X_1 \in A)$  realiza-se quando

$$\left\{ \begin{array}{l} \nu-1 \text{ pontos de } X_2, X_3, \dots, X_n \text{ pertencem a } [0,1]^s \\ \text{nenhum ponto pertence a } \bigcup_{r=1}^j Q_{n,t_r} \\ \text{pelo menos um ponto pertence a } Q_{n,t} \text{ para } t \notin \{t_1, t_2, \dots, t_j\} \end{array} \right.$$

e por razões análogas às anteriores  $P_f((X_1 \in A) | e_n) = C' P_f(A)$  com  $C'$  independente de A. Desta igualdade concluímos que para  $t \notin \{t_1, t_2, \dots, t_j\}$ ,  $P_f((X_1 \in Q_{n,t}) | e_n) = C' \alpha \rho_n^s$ , logo

$$P_f((X_1 \in [0,1]^s - \bigcup_{r=1}^j Q_{n,t_r}) | e_n) = \sum_{t \notin \{t_1, t_2, \dots, t_j\}} P_f((X_1 \in Q_{n,t}) | e_n) = (1 - j\rho_n^s) \alpha C'.$$

Mas directa-

mente tem-se que  $P_f((X_1 \in [0,1]^s - \bigcup_{r=1}^j Q_{n,t_r}) | e_n) = \frac{\nu}{n}$  e portanto

$$P_f((X_1 \in A) | e_n) = \int_A \frac{\nu}{n(1-j\rho_n^s)} dx. \square$$

Para a determinação de condições suficientes de convergência uniforme quase completa, vamos necessitar de fazer algumas hipóteses suplementares sobre o núcleo K. Tomemos a seguinte

**Definição 2.4.2:** Dizemos que um núcleo K em  $\mathbb{R}^s$  é de Bertrand-Retali - Geffroy se:

- i) O conjunto dos seus pontos de descontinuidade é de medida nula
- ii)  $K_1$  é integrável em  $\mathbb{R}^s$ .

Sendo K um núcleo em  $\mathbb{R}^s$  cujo conjunto dos seus pontos de descontinuidade é de medida nula, as seguintes classes de núcleos são de Bertrand-Retali - Geffroy :

a)  $K$  possui suporte limitado.

b)  $K$  é de variação limitada nas vizinhanças de infinito ( $s=1$ ).

c)  $K(x) = \prod_{i=1}^s K_i(x_i)$ , para  $x \in \mathbb{R}^s$ , com  $K_i$  núcleo de Bertrand-Retali - Geffroy em  $\mathbb{R}$ .

Designemos por  $(\mathcal{P}_m)_{m \in \mathbb{N}}$  a sucessão de partições de  $\mathbb{R}^s$ , como definida no parágrafo 1.1, tal que  $h_m \rightarrow 0$ ,  $m \rightarrow \infty$  e para todo o  $m \in \mathbb{N}$ ,  $\frac{h_m}{h_{m+1}} \in \mathbb{N}$ . Como consequência imediata da proposição 1.5.2 podemos obter o resultado seguinte:

**Lema 2.4.5:** *Seja  $K$  um núcleo de Bertrand-Retali - Geffroy em  $\mathbb{R}^s$  existem duas sucessões  $(\varphi_m^+)$  e  $(\varphi_m^-)$  ( $m \geq m_0$ ) de funções  $\mathcal{P}_m$ -escalonadas e integráveis em  $\mathbb{R}^s$  tais que*

$$\forall m \geq m_0 \quad \varphi_m^- \leq \varphi_{m+1}^- \leq K \leq \varphi_{m+1}^+ \leq \varphi_m^+$$

$$\text{e} \quad \lim_m \int_{\mathbb{R}^s} \varphi_m^+(x) dx = \lim_m \int_{\mathbb{R}^s} \varphi_m^-(x) dx = 1.$$

Passemos então a demonstrar o

**Teorema 2.4.2:** *Seja  $K$  um núcleo de Bertrand-Retali - Geffroy. Se  $h_n^{-s} = o(\frac{n}{\log n})$  então*

$$\Gamma_n \rightarrow 0, n \rightarrow \infty, \text{ q.co. para todo o } f \in \mathcal{D}.$$

*dem:*

Seja  $(Q_m)_{m \in \mathbb{N}}$  a sucessão de partições de  $\mathbb{R}^s$  cujo elemento genérico de  $Q_m$  é da forma

$$Q_{m,r_1,r_2,\dots,r_s} = \prod_{i=1}^s \left[ \frac{r_i}{2^m}, \frac{r_i+1}{2^m} \right], \text{ com } (r_1, r_2, \dots, r_s) \in \mathbb{Z}^s.$$

Tendo em conta o lema anterior existem duas sucessões  $\varphi_m^-$  e  $\varphi_m^+$  de funções  $Q_m$ -escalonadas tais que para  $m \geq m_0$   $\varphi_m^- \leq \varphi_{m+1}^- \leq K \leq \varphi_{m+1}^+ \leq \varphi_m^+$  e

$$\lim_m \int_{\mathbb{R}^s} \varphi_m^+(x) dx = \lim_m \int_{\mathbb{R}^s} \varphi_m^-(x) dx = 1.$$

$$\text{Sejam então, para } m \geq m_0, \quad \varphi_m^+ = \sum_{r=1}^{\infty} \alpha_{m,r} \mathbb{1}_{B_{m,r}} \quad \text{e} \quad \varphi_m^- = \sum_{r=1}^{\infty} \alpha'_{m,r} \mathbb{1}_{B_{m,r}}$$

onde  $\{B_{m,r} \mid r \in \mathbb{N}\} = \{Q_{m,r_1,r_2,\dots,r_s} \mid (r_1,r_2,\dots,r_s) \in \mathbb{Z}^s\}$  e  $\alpha_{m,r}, \alpha'_{m,r} \in \mathbb{R}$  para  $r \in \mathbb{N}$ .



Sejam  $f \in \mathcal{D}$  e  $\varepsilon > 0$ , quaisquer. Tomemos  $H > \sup_{x \in \mathbb{R}^s} f(x)$  e  $m \geq m_0$  fixo de modo que

$$1 - \varepsilon \leq \int_{\mathbb{R}^s} \varphi_m^-(x) dx \leq 1 \quad \text{e} \quad 1 \leq \int_{\mathbb{R}^s} \varphi_m^+(x) dx \leq 1 + \varepsilon \quad (2.4.10).$$

Como  $\varphi_m^+$  e  $\varphi_m^-$  são integráveis, existe  $R$ , parte finita de  $\mathbb{N}$ , tal que  $\sum_{r \in R} |\alpha_{m,r}| \lambda(B_m)$ ,

$\sum_{r \in R} |\alpha'_{m,r}| \lambda(B_m) < \varepsilon$  (2.4.11), onde  $B_m = [-\frac{1}{2^{m+1}}, \frac{1}{2^{m+1}}]^s$  e consideremos o estimador da

janela móvel associado à sucessão  $h_n = h_n/2^m$ , ou seja,

$$\forall x \in \mathbb{R}^s, n \in \mathbb{N} \quad f_n^{\sim}(x) = \frac{1}{n(h_n/2^m)^s} \sum_{i=1}^n \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]^s} \left( \frac{x - X_i}{h_n/2^m} \right).$$

Como por hipótese,  $(h_n)^{-s} = o(\frac{n}{\log n})$ , sabemos da proposição 2.4.2 que  $f_n^{\sim}$  converge uniformemente quase completamente para  $f$ . Por definição de  $B_m$ , para todo  $x \in \mathbb{R}^s$  e  $n \in \mathbb{N}$ ,

$$f_n^{\sim}(x) = \frac{1}{nh_n^s \lambda(B_m)} \sum_{i=1}^n \mathbb{1}_{B_m} \left( \frac{x - X_i}{h_n} \right).$$

Utilizando as propriedades de  $\varphi_m^+$ , temos que  $f_n^{\sim}(x) \leq \frac{1}{nh_n^s} \sum_{i=1}^n \varphi_m^+ \left( \frac{x - X_i}{h_n} \right)$

$$= \sum_{r=1}^{\infty} \alpha_{m,r} \lambda(B_m) f_n^{\sim}(x - h_n a_r), \text{ onde } a_r$$

designa o ponto médio de  $B_{m,r}$ . Assim

$$f_n^{\sim}(x) - f(x) \leq \sum_{r=1}^{\infty} \alpha_{m,r} \lambda(B_m) [f_n^{\sim}(x - h_n a_r) - f(x)] + \left[ \sum_{r=1}^{\infty} \alpha_{m,r} \lambda(B_m) - 1 \right] f(x) \quad (2.4.12).$$

Designemos para  $m \geq m_0$   $t_m = \sum_{r: \alpha_{m,r} > 0} \alpha_{m,r} \lambda(B_m)$ . Como  $\varphi_m^+ \geq \varphi_{m+1}^+$  obtemos que  $t_{m+1} \leq t_m$

e sendo  $t_m \geq 1$  tem-se que  $t_m \rightarrow L, m \rightarrow \infty$ .

Seja  $t'_m = \sum_{r=1}^{\infty} |\alpha_{m,r}| \lambda(B_m) \rightarrow 2L - 1, m \rightarrow \infty$ . Assim  $t'_m \leq A$ , para  $m \geq m_0$  (2.4.13).

Definamos então os acontecimentos

$$E_n = \{ \forall x \in \mathbb{R}^s \ f_n^{\sim}(x) < H \}$$

$$E'_n = \bigcap_{r \in \mathbb{R}} \{ \forall x \in \mathbb{R}^s \ |f_n^{\sim}(x - h_n a_r) - f(x)| < \varepsilon H \},$$

e sejam  $u_n = P_f((E_n)^c)$  e  $v_n = P_f((E'_n)^c)$ . Da  $u$ -continuidade de  $f$  e da convergência uniforme

q.co. de  $f_n^{\sim}$  para  $f$  tem-se que  $\sum_{n=1}^{\infty} u_n, \sum_{n=1}^{\infty} v_n < \infty$ .

Supondo então que  $E_n$  e  $E'_n$  se realizam temos de (2.4.12) que

$$\begin{aligned} f_n(x) - f(x) &\leq \sum_{r \in \mathbb{R}} |\alpha_{m,r}| \lambda(B_m) |f_n^{\sim}(x - h_n a_r) - f(x)| + \sum_{r \in \mathbb{R}} |\alpha_{m,r}| \lambda(B_m) [f_n^{\sim}(x - h_n a_r) - f(x)] + \\ &+ \left[ \sum_{r=1}^{\infty} \alpha_{m,r} \lambda(B_m) - 1 \right] f(x) \end{aligned}$$

$\leq \varepsilon H(A+3)$  por (2.4.10), (2.4.11), (2.4.13).

Assim  $P_f(E_n \cap E'_n) \leq P_f(\sup_{x \in \mathbb{R}^s} (f_n(x) - f(x)) \leq \varepsilon H(A+3))$ , ou seja,

$$P_f(\sup_{x \in \mathbb{R}^s} (f_n(x) - f(x)) > \varepsilon H(A+3)) \leq u_n + v_n.$$

Utilizando as propriedades da sucessão  $\varphi_m^{\sim}$  e de forma análoga ao que atrás se fez obtemos que se  $E_n$  e  $E'_n$  se realizam então para todo o  $x \in \mathbb{R}^s$   $f(x) - f_n(x) \leq \varepsilon H(B+3)$  sendo  $B$  um majo-

rante da sucessão  $t'_m = \sum_{r=1}^{\infty} |\alpha'_{m,r}| \lambda(B_m)$  para  $m \geq m_0$  e tendo em conta (2.4.10) e (2.4.11).

Portanto  $P_f(\sup_{x \in \mathbb{R}^s} (f(x) - f_n(x)) > \varepsilon H(B+3)) \leq u_n + v_n$ .

Da convergência das séries  $\sum_{n=1}^{\infty} u_n, \sum_{n=1}^{\infty} v_n$ , vem que  $\Gamma_n \rightarrow 0, n \rightarrow \infty, q.co. \square$

Dos teoremas 2.4.1 e 2.4.2 podemos obter o teorema de equivalência seguinte:

**Teorema 2.4:** Se  $K$  é um núcleo de Bertrand-Retali - Geffroy em  $\mathbb{R}^s$  e  $\lim_n h_n = 0$  são equivalentes as seguintes afirmações:

$$i) h_n^{-s} = o\left(\frac{n}{\log n}\right)$$

$$ii) \Gamma_n \rightarrow 0, n \rightarrow \infty, \text{ q.co. (p., q.c.) para todo } f \in \mathcal{D}.$$

## 2.5. Escolha assintoticamente óptima de $h_n$ e de $K$

Vamos em primeiro lugar obter, para uma determinada classe de densidades de probabilidade e de núcleos em  $\mathbb{R}^s$ , uma expressão para o erro quadrático médio integrado (EQMI).

Em consequência deste resultado deduzimos a forma da sucessão  $(h_n)_n$  que minimiza assintoticamente o EQMI e também o núcleo óptimo num determinado sentido e dentro de uma certa classe de núcleos.

Vamos designar por  $W^s$  o conjunto dos núcleos em  $\mathbb{R}^s$  que satisfazem as seguintes condições :

$$i) \int_{\mathbb{R}} y_1 K(y) dy_1 < \infty \text{ e } \int_{\mathbb{R}} y_1 K(y) dy_1 = 0 \text{ para } i=1,2,\dots,s$$

$$ii) \int_{\mathbb{R}^s} \|y\|^2 K(y) dy < \infty$$

onde  $y = (y_1, y_2, \dots, y_s)$ .

Para que a segunda condição de i) seja satisfeita (supondo satisfeita a primeira) basta exigir que  $K$  seja par.

Como exemplos de núcleos pertencentes a  $W^1$  temos o núcleo normal, o núcleo  $K(x) = \mathbb{1}_{[-1/2, 1/2]}(x)$ , o núcleo triangular  $K(x) = (1 - |x|) \mathbb{1}_{[-1, 1]}(x)$  e o núcleo de Picard  $K(x) = \frac{1}{2} \exp(-|x|)$ .

Facilmente se conclui que sendo  $K_i \in W^1, i=1,2,\dots,s$ , então  $K(x) = \prod_{i=1}^s K_i(x_i)$  com  $x=(x_1, x_2, \dots, x_s)$  é um núcleo da classe  $W^s$ .

**Teorema 2.5.1:** *Seja  $K \in W^s \cap L_2(\mathbb{R}^s)$ . Suponhamos que  $f \in L_2(\mathbb{R}^s)$  é duas vezes continuamente diferenciável em  $\mathbb{R}^s$  e que as suas derivadas parciais de 2ª ordem pertencem a  $L_2(\mathbb{R}^s)$  e são limitadas. Se  $h_n \rightarrow 0, n \rightarrow \infty$  então*

$$EQMI(h_n) = \frac{1}{nh_n^s} \int_{\mathbb{R}^s} K^2(x) dx + \frac{h_n^4}{4} \int_{\mathbb{R}^s} \Delta_s^2(x) dx + O\left(\frac{1}{n}\right) + o(h_n^4)$$

onde  $A_S(x) = \sum_{i=1}^s \frac{\partial^2 f}{\partial x_i^2}(x) \int_{\mathbb{R}^s} y_i^2 K(y) dy$

dem :

Temos que  $EQMI(h_n) = \frac{1}{nh_n^s} \int_{\mathbb{R}^s} K^2(x) dx - \frac{1}{n} \int_{\mathbb{R}^s} E^2[f_n(x)] dx + \int_{\mathbb{R}^s} (E[f_n(x)] - f(x))^2 dx$ .

Atendendo a que  $f \in L_2(\mathbb{R}^s)$ ,  $\int_{\mathbb{R}^s} E^2[f_n(x)] dx \leq \int_{\mathbb{R}^s} f^2(x) dx$ ,  $n \rightarrow \infty$  e portanto

$\frac{1}{n} \int_{\mathbb{R}^s} E^2[f_n(x)] dx = O(\frac{1}{n})$ . Preocupemo-nos agora com a expressão

$$\int_{\mathbb{R}^s} (E[f_n(x)] - f(x))^2 dx = \int_{\mathbb{R}^s} \left( \int_{\mathbb{R}^s} K(z)(f(x-h_n z) - f(x)) dz \right)^2 dx.$$

Das condições sobre  $f$  e  $K$  concluímos que  $\left( \int_{\mathbb{R}^s} K(z)(f(x-h_n z) - f(x)) dz \right)^2 =$

$$= h_n^4 \sum_{i,j,k,l=1}^s \int_{\mathbb{R}^s} \int_0^1 (1-t) \frac{\partial^2 f}{\partial x_i \partial x_j}(x-h_n t z) K(z) z_i z_j dt dz \cdot \int_{\mathbb{R}^s} \int_0^1 (1-t') \frac{\partial^2 f}{\partial x_k \partial x_l}(x-h_n t' z') K(z') z'_k z'_l dt' dz'.$$

Assim  $\int_{\mathbb{R}^s} (E[f_n(x)] - f(x))^2 dx = h_n^4 \sum_{i,j,k,l=1}^s c_{i,j,k,l}(n)$  onde  $c_{i,j,k,l}(n) =$

$$= \int_{\mathbb{R}^s} \int_{\mathbb{R}^s} \int_0^1 \int_0^1 (1-t)(1-t') K(z) z_i z_j K(z') z'_k z'_l \int_{\mathbb{R}^s} \frac{\partial^2 f}{\partial x_i \partial x_j}(x-h_n t z) \frac{\partial^2 f}{\partial x_k \partial x_l}(x-h_n t' z') dx dt' dt dz dz'.$$

Denotando por  $I_{n,i,j,k,l}(t,t',z,z')$  a função integranda anterior temos que

$$I_{n,i,j,k,l}(t,t',z,z') \rightarrow (1-t)(1-t') K(z) z_i z_j K(z') z'_k z'_l \int_{\mathbb{R}^s} \frac{\partial^2 f}{\partial x_i \partial x_j}(y) \frac{\partial^2 f}{\partial x_k \partial x_l}(y) dy, n \rightarrow \infty,$$

para todo o ponto  $(t,t',z,z')$ , atendendo à continuidade do produto de convolução de funções de quadrado integrável. Como

$$|I_{n,i,j,k,l}(t,t',z,z')| \leq$$

$$\leq (1-t)(1-t') |K(z) z_i z_j| |K(z') z'_k z'_l| \left( \int_{\mathbb{R}^s} \left( \frac{\partial^2 f}{\partial x_i \partial x_j}(y) \right)^2 dy \right)^{1/2} \left( \int_{\mathbb{R}^s} \left( \frac{\partial^2 f}{\partial x_k \partial x_l}(y) \right)^2 dy \right)^{1/2}$$

então pelo teorema da convergência dominada

$$c_{i,j,k,l}(n) \rightarrow \frac{1}{4} \int_{\mathbb{R}^s} K(z)z_i z_j dz \int_{\mathbb{R}^s} K(z')z'_k z'_l dz' \cdot \int_{\mathbb{R}^s} \frac{\partial^2 f}{\partial x_i \partial x_j}(y) \frac{\partial^2 f}{\partial x_k \partial x_l}(y) dy, n \rightarrow \infty.$$

Das condições sobre  $K$  vem finalmente o que pretendemos.  $\square$

Para cada  $n \in \mathbb{N}$ , designemos por *erro quadrático médio integrado assintótico em*  $h_n$   $> 0$ , EQMIA( $h_n$ ), a quantidade  $\frac{1}{nh_n^s} \int_{\mathbb{R}^s} K^2(x) dx + \frac{h_n^4}{4} \int_{\mathbb{R}^s} A_s^2(x) dx$ , para  $f \in \mathcal{F}$  nas condições

do teorema anterior.

**Corolário 2.5.1** : Nas condições do teorema anterior e supondo que  $\int_{\mathbb{R}^s} A_s^2(x) dx \neq 0$  então

para cada  $n \in \mathbb{N}$ , o valor  $h_n^* > 0$ , que minimiza o EQMIA( $h_n$ ) é

$$h_n^* = \left( \frac{\int_{\mathbb{R}^s} K^2(x) dx}{\int_{\mathbb{R}^s} A_s^2(x) dx} \right)^{\frac{1}{s+4}} n^{-\frac{1}{s+4}}$$

$$e \text{EQMI}(h_n^*) = (s+4) \left[ \left( \int_{\mathbb{R}^s} A_s^2(x) dx \right)^s \left( \int_{\mathbb{R}^s} K^2(x) dx \right)^4 \right]^{\frac{1}{s+4}} n^{-\frac{4}{s+4}} + o\left(n^{-\frac{4}{s+4}}\right)$$

Uma conclusão importante pode ser tirada deste corolário e do corolário 1.5.1 acerca da comparação entre as ordens de convergência, no sentido do erro quadrático médio integrado, dos estimadores do histograma e do núcleo.

Supondo então que  $f \in \mathcal{F}$  está nas condições quer do teorema 1.5.1, quer do teorema 2.5.1, podemos afirmar que as ordens de convergência assintoticamente óptimas são no caso do estimador do histograma  $n^{-\frac{2}{s+2}}$  e no caso do estimador do núcleo  $n^{-\frac{4}{s+4}}$  (para  $K \in \mathcal{W}^s$  de

quadrado integrável), daí uma convergência mais rápida deste último.

Reparar no entanto que por exemplo no caso  $s=1$ , estas ordens são semelhantes para  $n$  pequeno :

n	25	50	100	500
$n^{-2/3}$	0.117	0.0736	0.0464	0.0159
$n^{-4/5}$	0.0761	0.0437	0.025	0.0069

Quadro 2.5.1

Assim o que realmente vai ter grande influência no EQMI são as constantes que interferem nas expressões obtidas para  $h_n^*$  e que na prática são desconhecidas pois dependem da densidade  $f$  que desconhecemos.

Tal como fizemos quando do estudo do histograma, analisemos a influência que pode ter sobre o EQMIA( $h_n$ ) o facto de escolhermos para  $h_n$  um valor que não é o óptimo. Seja  $h_n = ch_n^*$  com  $c > 0$ . Concluimos então que  $\text{EQMIA}(h_n) = \left( \frac{4 + sc^{s+4}}{c^s(s+4)} \right) \text{EQMIA}(h_n^*)$ .

A função de  $s$ ,  $\frac{4 + sc^{s+4}}{c^s(s+4)}$  é estritamente crescente para  $c \neq 1$  e as conclusões que se podem tirar de tal facto são semelhantes às tiradas para o estimador do histograma.

O quadro seguinte exprime o aumento do EQMIA( $h_n$ ) relativamente ao EQMIA( $h_n^*$ ) para alguns valores de  $s$  e  $c$ :

s \ c	0.5	0.75	1.25	2
1	61%	13%	13%	260%
2	169%	29%	24%	450%
3	360%	49%	34%	593%

Quadro 2.5.2

Também aqui podemos estabelecer uma comparação com o estimador do histograma. Recordando o Quadro 1.5.1 concluimos que o estimador do núcleo é mais sensível a variações de  $h_n$  relativamente ao valor óptimo  $h_n^*$ .

Notar também, o grande crescimento, no caso do estimador do núcleo, do EQMIA( $h_n$ ) quando se escolhe  $h_n$  superior a  $h_n^*$ .

Quando usamos o método do núcleo para estimar a densidade de probabilidade, além da escolha da sucessão  $(h_n)_n$  a utilizar temos também de tomar um núcleo  $K$ . No que se segue vamos provar que existe um núcleo óptimo num determinado sentido mas que a utilização de outro núcleo "razoável" dá-nos resultados quase óptimos.

Suponhamos então que a densidade de probabilidade  $f$  está nas condições do teorema 2.5.1. Designemos por  $\mathcal{C}$  a classe dos núcleos em  $\mathbb{R}^s$  da forma

$$K(x) = \prod_{i=1}^s K_0(x_i), \text{ com } x = (x_i)_{i=1}^s \in \mathbb{R}^s$$

onde  $K_0$  é um núcleo positivo sobre  $\mathbb{R}$ , par, tal que  $\int_{\mathbb{R}} t^2 K_0(t) dt = 1$  e  $K_0$  é de quadrado integrável.

Pode parecer que a condição  $\int_{\mathbb{R}} t^2 K_0(t) dt = 1$  sobre o núcleo  $K$  restringe a generalidade da questão. Tal não acontece pois se  $K$  é um núcleo que satisfaz todas as condições impostas aos núcleos da classe  $\mathcal{C}$  excepto o facto de  $\int_{\mathbb{R}} t^2 K_0(t) dt = \frac{1}{\alpha^2} \neq 1$ , basta reparar que

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

onde  $h'_n = \frac{h_n}{\alpha}$  e  $K'(x) = \prod_{i=1}^s K_1(x_i)$  (com  $K_1(t) = \frac{1}{\alpha} K_0\left(\frac{t}{\alpha}\right)$ ) pertence à classe  $\mathcal{C}$ .

A imposição de que  $K_0$  seja par surge naturalmente pois sabemos que no caso de  $s=1$  a função  $K$  de quadrado integrável que minimiza o EQMI é necessariamente par (ver Leadbetter, p.482).

Nas condições anteriores temos que se  $h_n \rightarrow 0, n \rightarrow \infty$ , então

$$EQMIA(h_n) = \frac{1}{n h_n^s} \left( \int_{\mathbb{R}} K_0^2(x) dx \right)^s + \frac{h_n^4}{4} \int_{\mathbb{R}^s} \left| \sum_{i=1}^s \frac{\partial^2 f}{\partial x_i^2}(x) \right|^2 dx$$

Vamos tomar então como critério para a escolha de  $K$ , o núcleo que minimiza para todo o  $n \in \mathbb{N}$  o EQMIA( $h_n$ ).

Pretendemos assim determinar  $K_0: \mathbb{R} \rightarrow \mathbb{R}^+$  que minimize  $\int_{\mathbb{R}} K_0^2(t) dt$  onde  $K$  satisfaz as condições :

faz as condições :

$$- K_0(t) = K_0(-t) \text{ para todo } t \in \mathbb{R}$$

$$- \int_{\mathbb{R}} K_0(t) dt = 1$$

$$- \int_{\mathbb{R}} K_0(t) t^2 dt = 1$$

$$- K_0 \in L_2(\mathbb{R}).$$

Provemos em primeiro lugar que o problema assim posto possui solução única no caso desta existir. Designemos por  $\mathfrak{K}$  o conjunto das funções  $K$  que satisfazem as condições anteriores. Sejam então  $K_1$  e  $K_2$  elementos de  $\mathfrak{K}$  tais que

$$\int_{\mathbb{R}} K_1^2(t) dt = \int_{\mathbb{R}} K_2^2(t) dt \leq \int_{\mathbb{R}} K^2(t) dt \text{ para todo } K \in \mathfrak{K}.$$

Como  $\mathfrak{K}$  é um subconjunto convexo de  $L_2(\mathbb{R})$  então para todo o  $\alpha \in ]0,1[$ ,

$$\alpha \int_{\mathbb{R}} (K_1(t) - K_2(t))^2 dt \geq 2 \left( \int_{\mathbb{R}} K_2^2(t) dt - \int_{\mathbb{R}} K_1(t)K_2(t) dt \right).$$

Assim da desigualdade de Hölder e das expressões anteriores concluímos que  $\int_{\mathbb{R}} K_2^2(t) dt =$

$$= \int_{\mathbb{R}} K_1(t)K_2(t) dt = \int_{\mathbb{R}} K_1^2(t) dt \text{ e portanto } \int_{\mathbb{R}} (K_1(t) - K_2(t))^2 dt = 0.$$

Esqueçamos por enquanto a condição de positividade imposta a  $K_0$ . Pretendemos minimizar  $\int_{\mathbb{R}} K^2(t) dt$  onde  $K \in \mathfrak{K}_1 \cap \mathfrak{K}_2$  sendo  $\mathfrak{K}_1 = L_2(\mathbb{R}) \cap \{K \in L_1(\mathbb{R}) \mid \int_{\mathbb{R}} K(t)t^2 dt < \infty\}$  e

$$\mathfrak{K}_2 = \{K : \mathbb{R} \rightarrow \mathbb{R} \mid \int_{\mathbb{R}} K(t) dt = \int_{\mathbb{R}} K(t)t^2 dt = 1\}.$$

Este problema engloba-se num problema de minimização de uma funcional  $F(K) = \int_{\mathbb{R}} K^2(t) dt$  com domínio  $\mathfrak{K}_1$  sujeita às condições de ligação

$$G_1(K) = \int_{\mathbb{R}} K(t) dt - 1 = 0 \text{ e } G_2(K) = \int_{\mathbb{R}} K(t)t^2 dt - 1 = 0 \text{ (Luenberger, p.185-186).}$$



Facilmente se conclui que  $F, G_1$  e  $G_2$  são diferenciáveis à Gateaux em  $K \in \mathfrak{B}_1$  (Luenberger, p.171) e se  $h \in \mathfrak{B}_1$

$$\delta F(K;h) = 2 \int_{\mathbb{R}} h(t)K(t)dt, \quad \delta G_1(K;h) = \int_{\mathbb{R}} h(t)dt \quad \text{e} \quad \delta G_2(K;h) = \int_{\mathbb{R}} h(t)t^2 dt.$$

Se  $F, G_1$  e  $G_2$  fossem diferenciáveis à Fréchet (Luenberger, p.172) e se  $K: \mathbb{R} \rightarrow \mathbb{R}^+$  fosse uma solução para o problema anterior existiriam reais  $\lambda_1$  e  $\lambda_2$  tais que

$$\delta(F + \lambda_1 G_1 + \lambda_2 G_2)(K, h) = 0 \text{ para todo o } h \in \mathfrak{B}_1 \text{ (Luenberger, p.187-189), ou}$$

seja,  $2 \int_{\mathbb{R}} h(t)K(t) dt + \lambda_1 \int_{\mathbb{R}} h(t) dt + \lambda_2 \int_{\mathbb{R}} h(t)t^2 dt = 0$  para todo o  $h \in \mathfrak{B}_1$ . Assim para todo o  $t \in \mathbb{R}$ ,  $2K(t) + \lambda_1 + \lambda_2 t^2 = 0$  ( $K$  é necessariamente par). Sendo  $K \in \mathfrak{B}_1$  e positiva tomemos tendo em conta a expressão anterior

$$K(t) = \frac{3}{4\sqrt{5}} \left(1 - \frac{t^2}{5}\right) \mathbb{1}_{[-\sqrt{5}, \sqrt{5}]}(t).$$

Provemos que  $K$  assim definida é a solução pretendida :

Sejam  $h \in \mathfrak{B}$  e  $g = h - K$ . Temos que  $\int_{\mathbb{R}} g(t) dt = \int_{\mathbb{R}} t^2 g(t) dt = 0$  e  $g(t) \geq 0$  para  $|t| \geq \sqrt{5}$ .

$$\text{Assim } \int_{\mathbb{R}} h^2(t) dt \geq \int_{\mathbb{R}} K^2(t) dt + \int_{\mathbb{R}} g^2(t) dt \geq \int_{\mathbb{R}} K^2(t) dt.$$

Podemos então enunciar o

**Teorema 2.5.2 :** *O núcleo de Epanechnikov definido por*

$$K_{E,s}(x) = \prod_{i=1}^s \frac{3}{4\sqrt{5}} \left(1 - \frac{x_i^2}{5}\right) \mathbb{1}_{[-\sqrt{5}, \sqrt{5}]}(x_i),$$

é de EQMIA mínimo na classe  $\mathfrak{C}$  onde supomos que  $f \in \mathfrak{F}$  está nas condições do teorema 2.5.1 e que  $h_n \rightarrow 0, n \rightarrow \infty$ .

Na tabela seguinte estão indicados os quocientes  $\int_{\mathbb{R}} K^2(t) dt / \int_{\mathbb{R}} K_{E,1}^2(t) dt$  para alguns núcleos de  $\mathbb{R}$  pertencentes à classe  $\mathfrak{C}$ .

Podemos concluir que o quociente atrás referido é pouco sensível a alterações do núcleo  $K$ , portanto mesmo não utilizando o núcleo de Epanechnikov os resultados obtidos com os núcleos aí indicados são quase óptimais no sentido do erro quadrático médio integrado assintótico.

$K(x)$	$\int_{\mathbb{R}} K^2(t) dt / \int_{\mathbb{R}} K_{E,1}^2(t) dt$
$\frac{1}{2\sqrt{3}} \mathbb{1}_{[-\sqrt{3}, \sqrt{3}]}(x)$	1.076
$\frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$	1.051
$(\frac{1}{\sqrt{6}} - \frac{ x }{6}) \mathbb{1}_{[-\sqrt{6}, \sqrt{6}]}(x)$	1.014
$\frac{1}{\sqrt{2}} \exp(-\sqrt{2} x )$	1.318
$\frac{2\pi}{(1+x^2)^2}$	1.483
$\frac{\sqrt{\pi^2-8}}{4} \cos(\frac{\sqrt{\pi^2-8}}{4}x) \mathbb{1}_{[-\frac{\pi}{\sqrt{\pi^2-8}}, \frac{\pi}{\sqrt{\pi^2-8}}]}(x)$	1.0005

Quadro 2.5.3

Nos quadros seguintes são apresentadas para quatro densidades de probabilidade sobre  $\mathbb{R}$  os valores do EQMIA( $h_n^*$ ) e do EQMI $\tilde{}$ ( $h_n^*$ ) (para o núcleo normal e para o núcleo de Epanechnikov) que é calculado de forma análoga ao que fizemos para o estimador do histograma e as conclusões aí tiradas sobre a comparação destes dois valores são aqui adaptáveis.

$n$	$h_n^* \times 10$	EQMIA( $h_n^*$ ) $\times 10^3$	EQMI $\tilde{}$ ( $h_n^*$ ) $\times 10^3$
10	6,68	52,8	27,0
50	4,84	14,6	10,1
100	4,22	8,36	7,56

Quadro 2.5.4 - Lei normal centrada e reduzida. Núcleo normal

$n$	$h_n^* \times 10$	EQMIA( $h_n^*$ ) $\times 10^3$	EQMI $\tilde{}$ ( $h_n^*$ ) $\times 10^3$
10	6,62	50,7	25,8
50	4,8	14,0	9,6
100	4,18	8,03	7,28
500	3,03	2,22	1,58

Quadro 2.5.5 - Lei normal centrada e reduzida. Núcleo de Epanechnikov.

n	$h_n^* \times 10$	$EQMIA(h_n^*) \times 10^2$	$EQMI^-(h_n^*) \times 10^2$
10	3,75	9,41	5,5
50	2,72	2,6	1,81
100	2,36	1,49	1,23

Quadro 2.5.6 - Lei de Cauchy. Núcleo normal.

n	$h_n^* \times 10$	$EQMIA(h_n^*) \times 10^3$	$EQMI^-(h_n^*) \times 10^3$
10	3,71	90,4	52,0
50	2,69	25,0	17,3
100	2,34	14,3	11,6
500	1,69	3,95	3,27

Quadro 2.5.7 - Lei de Cauchy. Núcleo de Epanechnikov.

n	$h_n^* \times 10$	$EQMIA(h_n^*) \times 10^3$	$EQMI^-(h_n^*) \times 10^3$
10	6,84	51,6	31,7
50	4,96	14,2	9,43
100	4,32	8,17	5,21

Quadro 2.5.8 - Mistura de leis normais. Núcleo normal.

n	$h_n^* \times 10$	$EQMIA(h_n^*) \times 10^3$	$EQMI^-(h_n^*) \times 10^3$
10	6,77	49,5	30,5
50	4,91	13,7	9,02
100	4,27	7,85	4,93
500	3,1	2,17	1,94

Quadro 2.5.9 - Mistura de leis normais. Núcleo de Epanechnikov.

n	$h_n^* \times 10$	$EQMIA(h_n^*) \times 10^3$	$EQMI^-(h_n^*) \times 10^3$
10	4,26	82,8	43,0
50	2,69	22,8	13,9
100	1,95	13,1	8,56

Quadro 2.5.10 - Lei t(5) de Student. Núcleo normal

n	$h_n^* \times 10$	$EQMIA(h_n^*) \times 10^3$	$EQMI^-(h_n^*) \times 10^3$
10	4,22	79,5	40,5
50	3,06	21,9	10,9
100	2,66	12,6	8,19
500	1,93	3,48	2,2

Quadro 2.5.11 - Lei t(5) de Student. Núcleo de Epanechnikov.

Para as quatro densidades de probabilidade atrás referidas, efectuámos também um estudo de simulação sobre a variação do  $EQMI^-(h_n)$  relativamente ao  $EQMI^-(h_n^*)$  quando tomamos  $h_n = 2h_n^*$  e  $h_n = h_n^*/2$ , utilizando os dois núcleos atrás referidos. Como os resultados são semelhantes para um e outro núcleo, fornecemos apenas resultados relativos ao núcleo de Epanechnikov.

Para valores de  $n=10$ ,  $n=50$  e  $n=500$  é generalizado verificar-se que a escolha de  $h_n = h_n^*/2$  provoca aumentos maiores do que os dados pelo quadro 2.5.2. Assim por exemplo para a lei normal os aumentos são de 124%, 110% e 82%. Para a lei de Cauchy de 139%, 112% e 89%, para a lei t(5) de Student de 140%, 146% e 119% e para a mistura de leis normais de 113%, 117% e 70%.

Por outro lado para  $h_n = 2h_n^*$  temos na lei normal um aumento de 109%, 146% e 268% para  $n=10$ ,  $n=50$  e  $n=500$ , enquanto que nas leis de Cauchy e Student temos diminuições de 54% e 41% para  $n=10$ , de 38% e 25% para  $n=50$  e de 23% e 11% para  $n=500$ .

Nota-se assim uma gradual aproximação aos resultados obtidos para o EQMIA (o que seria de esperar), no entanto mesmo para amostras de tamanho 500 os resultados estão em alguns casos longe dos dados pelo quadro 2.5.2.

A seguir o método do núcleo é aplicado a  $n$ -amostras simuladas.

Os gráficos seguintes representam, para alguns valores de  $n$ , estimativas das densidades atrás referidas. Os valores de  $h_n$  nelas usadas são os dados pelos quadros anteriores.

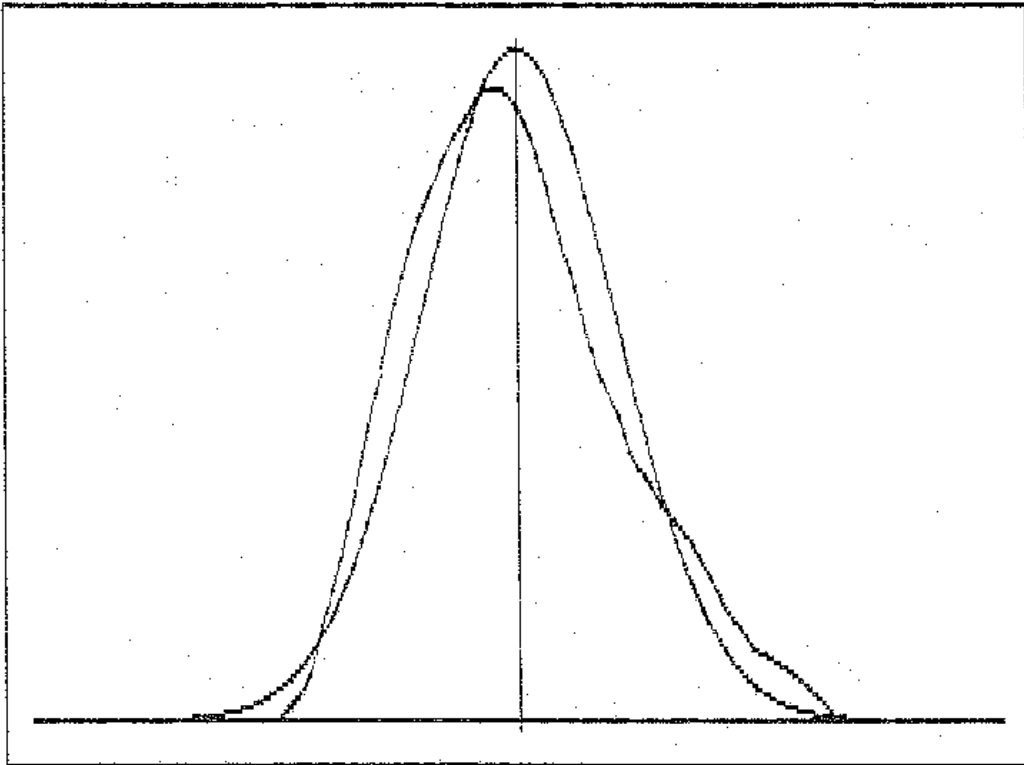


Gráfico 2.5.1 - Lei normal centrada e reduzida. Núcleo de Epanechnikov.  
 $n = 10$ ,  $EQI = 0.0099$ ,  $(-5,5) \times (0,0.399)$ .

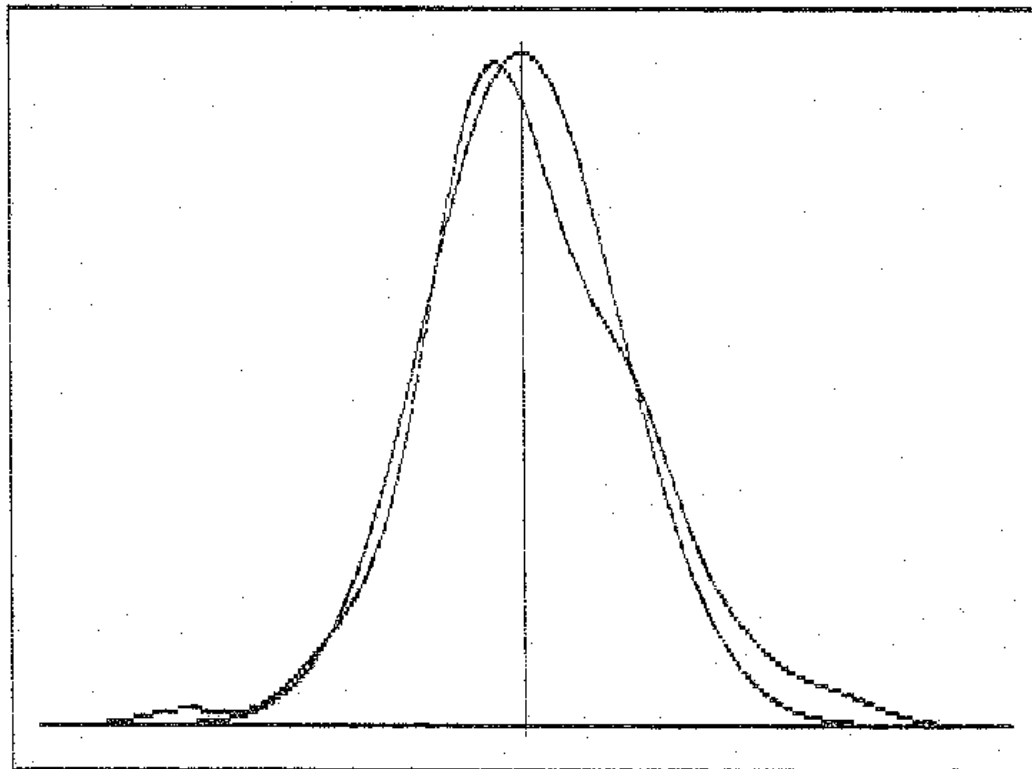


Gráfico 2.5.2 - Lei normal centrada e reduzida. Núcleo normal.  
 $n = 100$ ,  $EQI = 0.0049$ ,  $(-5,5) \times (0,0.399)$ .

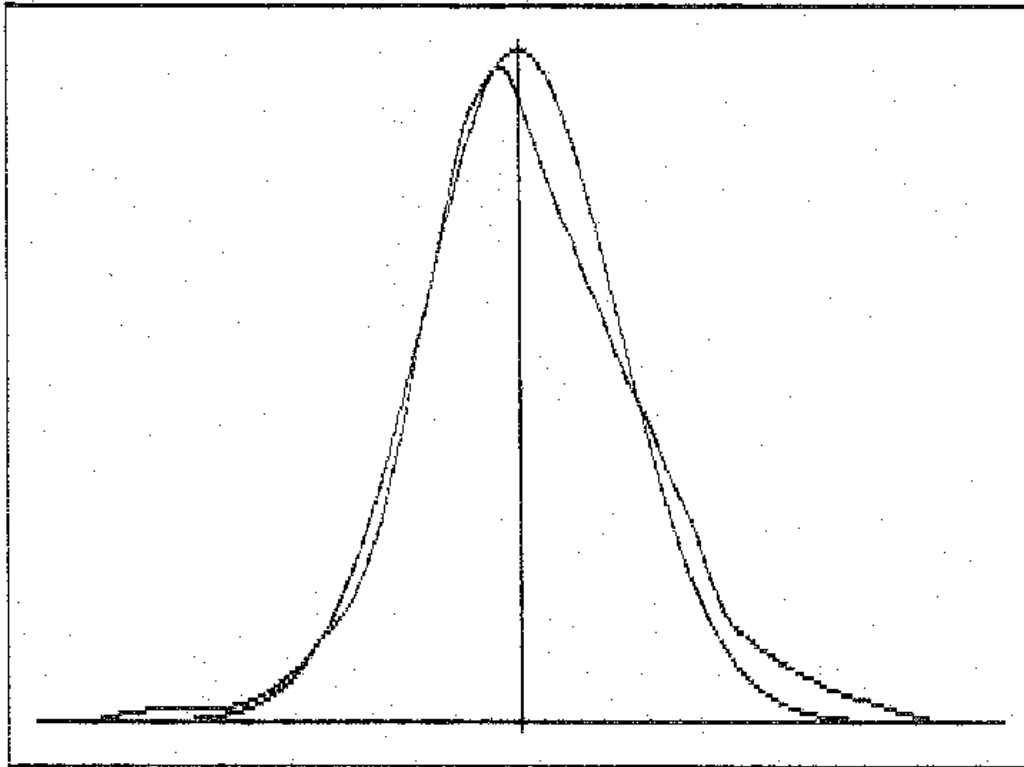


Gráfico 2.5.3 - Lei normal centrada e reduzida. Núcleo de Epanechnikov.  
 $n = 100$ ,  $EQI = 0.0043$ ,  $(-5,5) \times (0,0.399)$ .

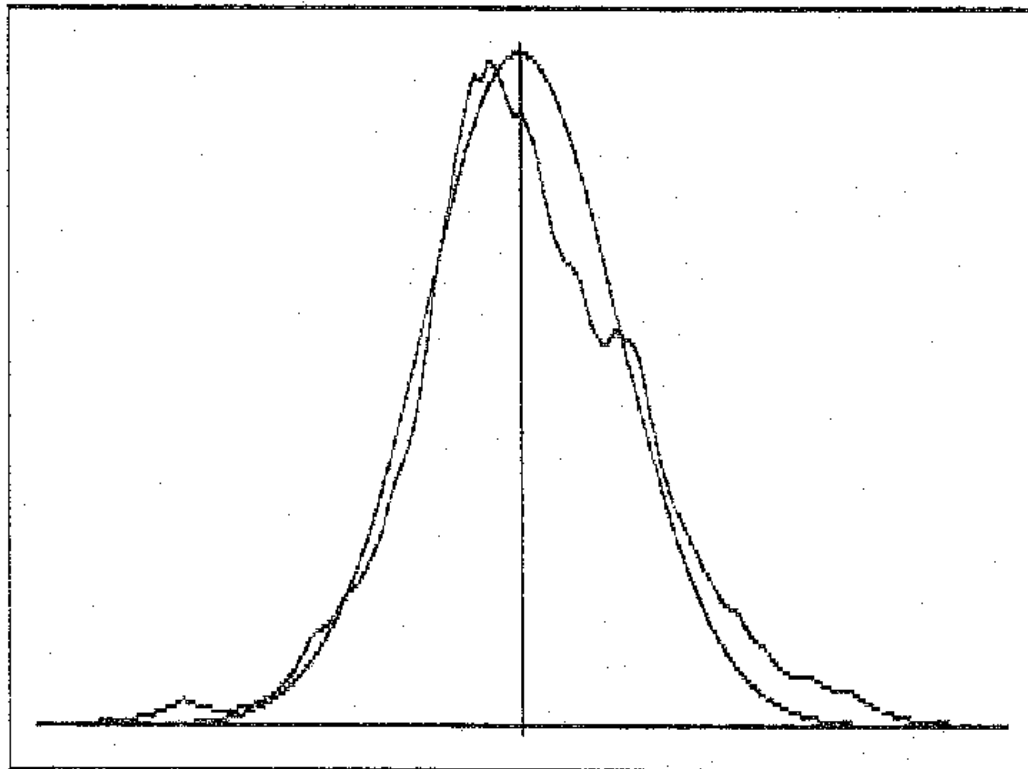


Gráfico 2.5.4 - Lei normal centrada e reduzida. Núcleo de Picard.  
 $n = 100$ ,  $EQI = 0.0064$ ,  $(-5,5) \times (0,0.399)$ .

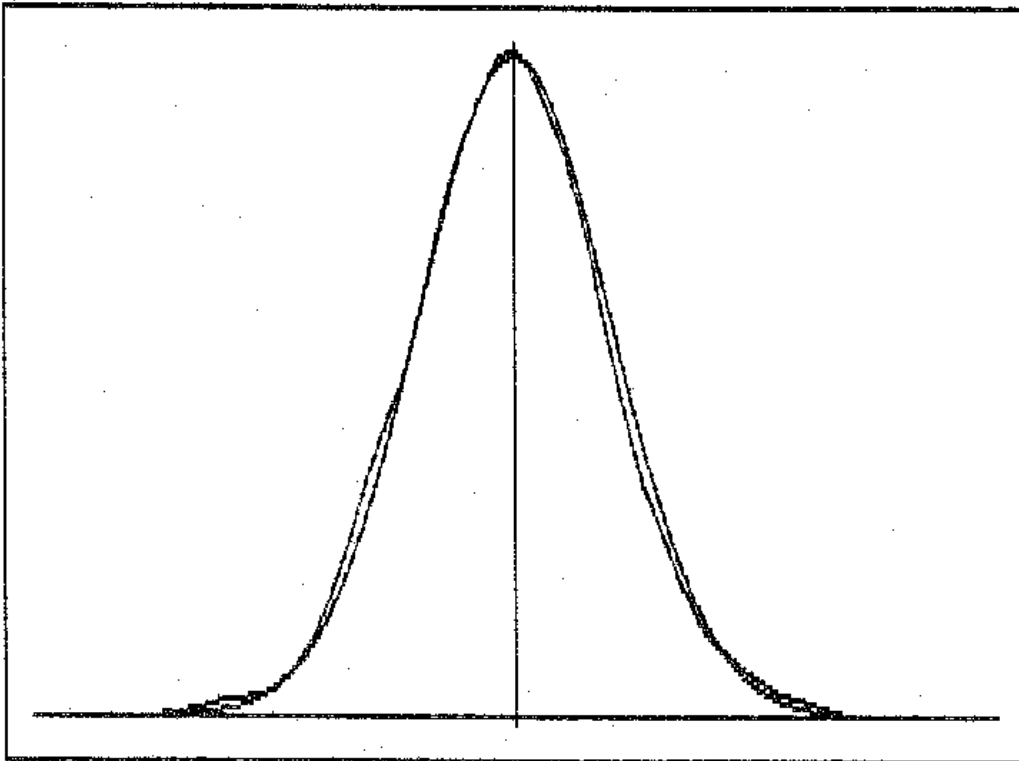


Gráfico 2.5.5 - Lei normal centrada e reduzida. Núcleo de Epanechnikov.  
 $n = 500$ ,  $EQI = 0.00065$ ,  $(-5;5) \times (0;0.402)$ .

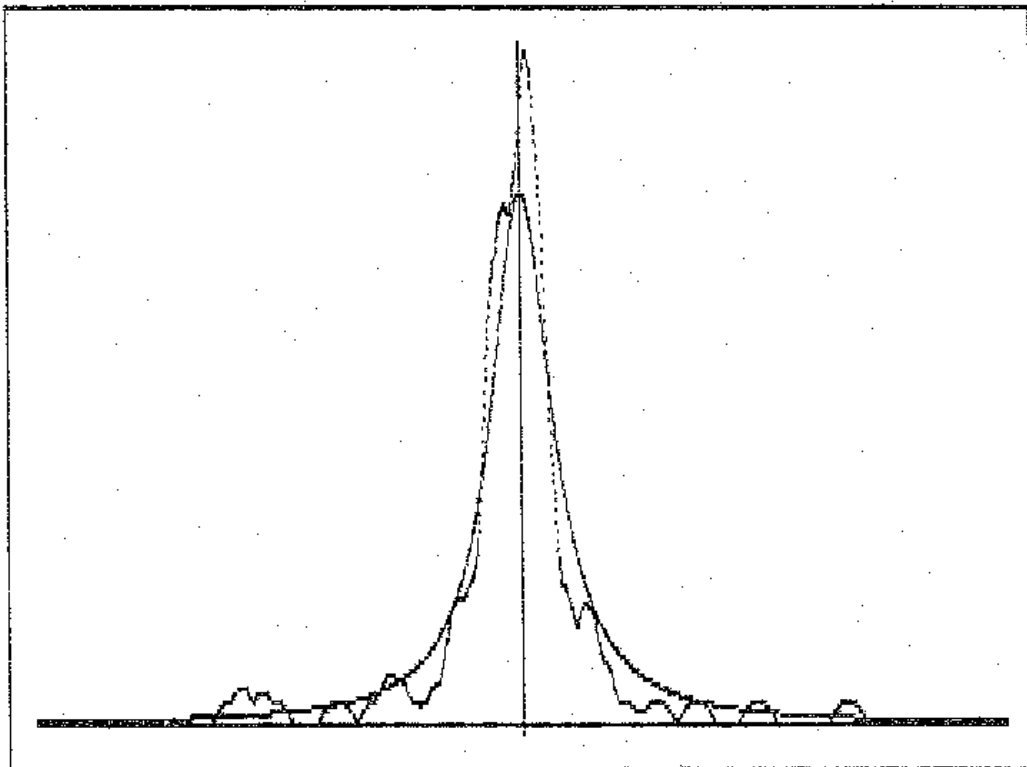


Gráfico 2.5.6 - Lei de Cauchy. Núcleo de Epanechnikov.  
 $n = 100$ ,  $EQI = 0.0103$ ,  $(-13;13) \times (0;0.403)$ .

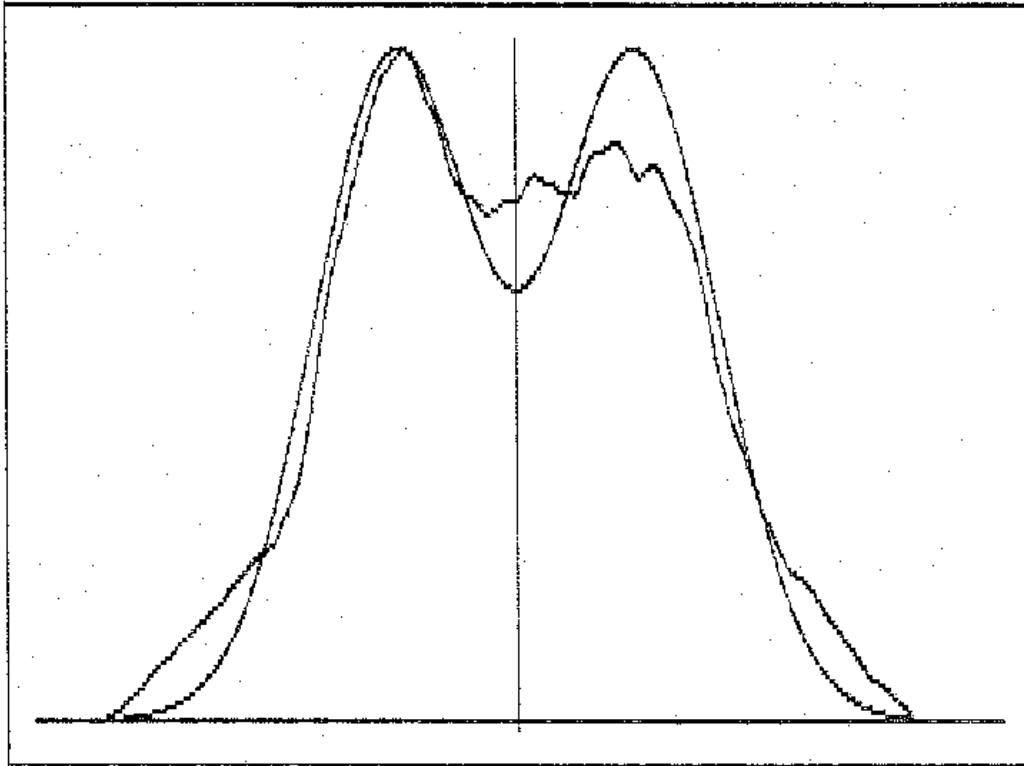


Gráfico 2.5.7 - Mistura de leis normais. Núcleo de Epanechnikov.  
 $n = 100$ ,  $EQI = 0.0025$ ,  $(-6,6) \times (0,0.202)$ .

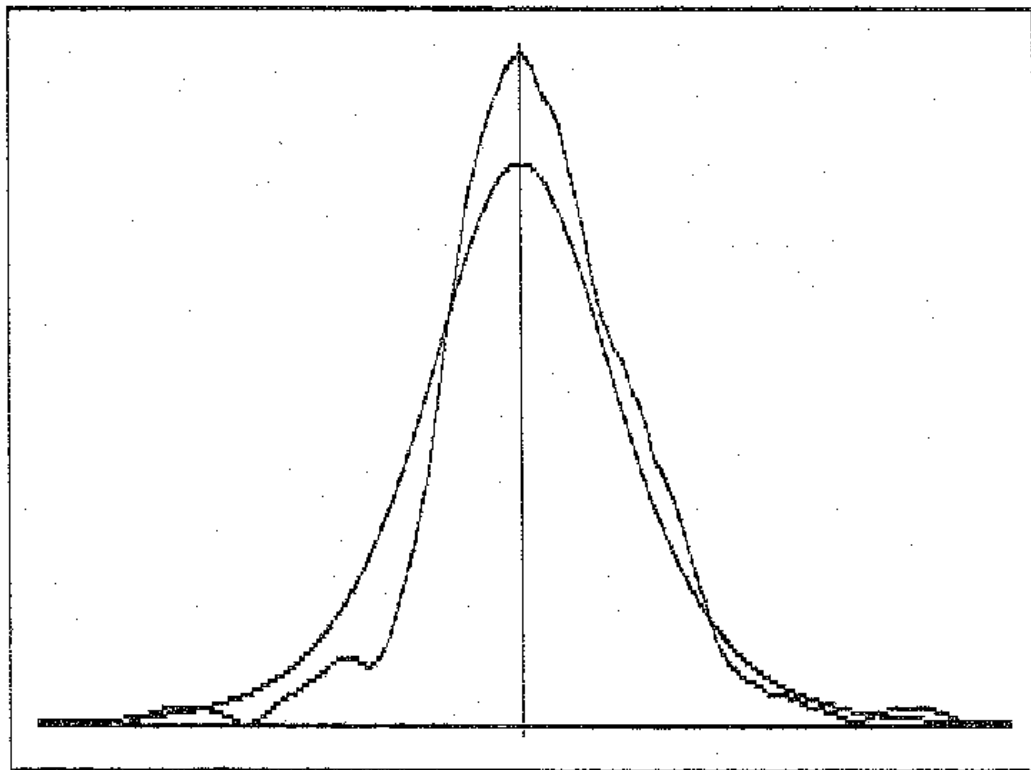


Gráfico 2.5.8 - Lei  $t(5)$  de Student. Núcleo de Epanechnikov.  
 $n = 100$ ,  $EQI = 0.0106$ ,  $(-5,5) \times (0,0.454)$ .



**CAPÍTULO III : O ESTIMADOR AUTOMÁTICO DO NÚCLEO  
BREVE ANÁLISE**

Vimos já ao longo deste trabalho, o quanto é importante a escolha de  $h_n$ . É o problema da escolha de  $h_n$ , para  $n \in \mathbb{N}$  fixo e no caso real, que vamos abordar neste último capítulo. Fixar-nos-emos simplesmente no estimador do núcleo; no entanto um estudo análogo pode ser feito para o caso do estimador do histograma.

Na parte final do capítulo anterior vimos que, utilizando a minimização do erro quadrático médio integrado assintótico como critério de escolha para  $h_n$ , esta sucessão depende da densidade a estimar. Do ponto de vista prático, os resultados obtidos têm assim um interesse pouco relevante.

Para  $n \in \mathbb{N}$ , fixo,  $f_n(\cdot | h)$  (denotaremos também  $f_n(x)$  por  $f_n(x | h)$  para melhor expressar a dependência de  $f_n(x)$  do parâmetro  $h$ ) é extremamente sensível à escolha de  $h$  (ver parágrafo 2.5.), não existindo contudo nenhum método, que seja de acordo geral, para a determinação deste parâmetro.

Sendo a amostra o nosso único dado, é natural querer usar as observações feitas para ajudar na escolha de  $h$ , ou melhor, que ele seja perfeitamente determinado por estas observações.

Assim consideraremos  $h_n$  não como uma constante real positiva, mas sim como uma aplicação mensurável da amostra

$$h_n : \mathbb{R}^n \longrightarrow \mathbb{R}^+ \\ (X_1, X_2, \dots, X_n) \longrightarrow h_n(X_1, X_2, \dots, X_n) \quad (1).$$

Obtemos assim, não o estimador do núcleo no sentido que lhe atribuímos no capítulo 2, mas sim o chamado estimador automático do núcleo, que é definido, para  $x \in \mathbb{R}$ , por

$$f_n(x; X^n) = f_n(x) = \frac{1}{nh_n(X_1, X_2, \dots, X_n)} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n(X_1, X_2, \dots, X_n)}\right),$$

onde  $K$  é um núcleo em  $\mathbb{R}$  e  $X^n$ , para  $n \in \mathbb{N}$ , uma  $n$ -amostra de lei  $P_f$ .

Facilmente se conclui que para cada  $x^n \in \mathbb{R}^n$ ,  $f_n(\cdot; x^n)$  é uma densidade de probabili-

---

(1) Especialmente no caso da estimação pontual da densidade considera-se também  $h_n(x, X_1, X_2, \dots, X_n)$ , o que aqui não vamos fazer.

dade sobre  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  desde que  $K \in \mathcal{F}$ .

A finalidade deste capítulo não é estudar o estimador automático do núcleo no sentido em que o fizemos para o estimador do núcleo (para tal ver por exemplo Wagner (1975) e Deheuvels (1980) para a convergência uniforme e Devroye (1984) e Devroye (1984a, p.148-190) para a convergência pontual e  $L_1$ ) mas sim fazer um estudo de simulação sobre a eficiência deste estimador, usando como medida dessa eficiência o erro quadrático médio integrado, para quatro possíveis escolhas da aplicação  $h_n$  e usando os núcleos normal e de Epanechnikov.

Não deixaremos contudo de fazer referência aos resultados teóricos mais importantes sobre a consistência de cada um desses estimadores.

A primeira das escolhas que vamos considerar é inspirada no resultado que obtivemos no corolário 2.5.1. As outras escolhas têm por base a ideia comum de medir, através duma função do parâmetro a estimar,  $h$ , a capacidade do estimador para interpretar ou se ajustar às observações realizadas. A escolha terá como objectivo a maximização dessa medida de interpretação.

### 3.1. Escolha "paramétrica" de $h_n$

O primeiro método, para a escolha de  $h_n$ , que vamos abordar, é como já referimos, baseado no resultado obtido no corolário 2.5.1 e foi proposto em 1977 por Deheuvels. No resultado referido concluímos que para  $f$  pertencente a uma determinada classe de densidades de probabilidade se tem que o valor que minimiza, para cada  $n \in \mathbb{N}$ , o erro

quadrático médio integrado assintótico é : 
$$h_n = c_k \left( \int_{\mathbb{R}} f''^2(x) dx \right)^{-1/5} n^{-1/5} \quad (3.1.1)$$

onde  $c_k = \left( \frac{\int_{\mathbb{R}} K^2(x) dx}{\left( \int_{\mathbb{R}} x^2 K(x) dx \right)^2} \right)^{1/5}$  e sendo  $K$  um núcleo nas condições aí expressas.

Para todas as densidades  $g$  do tipo de  $f$ , ou seja, tais que  $g(x) = bf(bx+a)$  onde  $a \in \mathbb{R}$  e  $b \in \mathbb{R}^+$ , tem-se que  $\left( \int_{\mathbb{R}} g''^2(x) dx \right)^{-1/5} = b^{-1} \left( \int_{\mathbb{R}} f''^2(x) dx \right)^{-1/5}$ . Assim conhecendo-se  $h_n$  para

uma densidade reduzida obtem-se facilmente  $h_n$  para uma qualquer densidade do mesmo tipo. Sendo então  $g$  reduzida do mesmo tipo de  $f$  temos por (3.1.1) que

$$h_n = c_k b \left( \int_{\mathbb{R}} g^2(x) dx \right)^{-1/5} n^{-1/5} \quad (3.1.2),$$

onde  $b^2$  pode ser estimada a partir de  $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ , desde que  $f$  possua momentos de 2ª ordem finitos.

Em (3.1.2) continuamos a desconhecer  $\int_{\mathbb{R}} g^2(x) dx$ . Vamos assim privilegiar a lei normal ao tomarmos para  $g$  a densidade normal centrada e reduzida. A escolha de outras densidades para  $g$  pode não afectar muito o valor de  $\left( \int_{\mathbb{R}} g^2(x) dx \right)^{-1/5}$ , devido ao expoente

que surge nesta expressão. No entanto e como exemplo, para a lei  $t(5)$  de Student a quantidade anterior é aproximadamente igual a 0.3 enquanto que para a lei normal standart o seu valor é  $\cong 1.36$ .

Escrevamos então a forma que  $h_n$  vai tomar :  $h_n(X_1, X_2, \dots, X_n) = 1.36 c_k S_n n^{-1/5}$ .

Realcemos finalmente, acerca do estimador obtido pela escolha anterior de  $h_n$ , alguns resultados sobre a sua consistência. Designemos por  $\mathbb{W}$  o conjunto dos núcleos  $K$  da classe  $\mathbb{W}^1$  que verificam  $\int_{\mathbb{R}} x^2 K(x) dx \neq 0$ .

**Teorema 3.1.1:** (Devroye, 1984a, p.148) *Seja  $K \in \mathcal{F} \cap \mathbb{W} \cap L_2(\mathbb{R})$ . Para todo o  $f \in \mathcal{F}$  que possua momentos de 2ª ordem finitos, tem-se que*

$$\int_{\mathbb{R}} |f_n(x) - f(x)| dx \longrightarrow 0, n \rightarrow \infty, \text{ q.c..}$$

**Teorema 3.1.2:** (Deheuvels, 1980, p.31) *Seja  $K \in \mathcal{F} \cap \mathbb{W} \cap L_2(\mathbb{R})$ , integrável à Riemann. Para todo o  $f \in \mathcal{D}$  que possua momentos de 2ª ordem finitos, tem-se que*

$$\sup_{x \in \mathbb{R}} |f_n(x) - f(x)| \longrightarrow 0, n \rightarrow \infty, \text{ q.c..}$$

**Teorema 3.1.3:** (Devroye, 1984, p.1232) *Seja  $K \in \mathcal{F} \cap \mathbb{W} \cap L_2(\mathbb{R})$  de suporte limitado e integrável à Riemann. Para todo o  $f \in \mathcal{F}$  que possua momentos de 2ª ordem finitos, tem-se que*

$$f_n(x) \longrightarrow f(x), n \rightarrow \infty, \text{ q.c., para quase todo o } x \in \mathbb{R}.$$

### 3.2. Método do tipo "máximo de verosimilhança" para a escolha de $h_n$

A ideia para a construção da função do parâmetro  $h$  que vamos passar a expôr, foi proposta separadamente por Habbema, Hermans, e Vandebroek em 1974 e por Duin em 1973 (ver Devroye, 1984a, p.152 e Chow, 1983, p.26). A aparência fortemente intuitiva e a sua relativa simplicidade talvez sejam as razões para ser estudado por diversos autores.

Sendo  $(X_1, X_2, \dots, X_n)$  uma  $n$ -amostra de lei  $P_f$  e se, para um valor  $h > 0$ ,  $f_n(X_j | h)$  ( $j \in \{1, 2, \dots, n\}$ ) é grande, é porque o parâmetro  $h$  é mais apropriado do que o parâmetro que torna  $f_n(X_j | h)$  pequeno. Assim para cada observação da  $n$ -amostra obtemos a expressão de

verosimilhança  $L^*(h) = \prod_{i=1}^n f_n(X_i | h)$  e  $h > 0$  será escolhido de modo a maximizar  $L^*$ .

Como para os núcleos usados habitualmente  $L^*(h) \rightarrow \infty$ , quando  $h \rightarrow 0$ , foi introduzida uma modificação no critério anterior utilizando um processo conhecido como "validação cruzada". Consideremos o estimador construído através da  $(n-1)$ -amostra  $(X_1, X_2, \dots, X_n) \setminus (X_i)$  (para cada  $i \in \{1, 2, \dots, n\}$ ) que vamos designar por  $f_{n,i}$ . Para  $n \in \mathbb{N}$ ,  $i \in \{1, 2, \dots, n\}$  e  $x \in \mathbb{R}$

$$f_{n,i}(x) = f_{n,i}(x | h) = \frac{1}{(n-1)h} \sum_{j \neq i} K\left(\frac{x - X_j}{h}\right).$$

Tomemos assim como critério de escolha de  $h$  :

$$" \text{Um dos } h > 0 \text{ que maximiza } L(h) = \prod_{i=1}^n f_{n,i}(X_i) ".$$

Nem sempre podemos garantir a existência do  $\sup_{h > 0} L(h)$  para qualquer núcleo positivo. No entanto no caso dos núcleos normal e de Epanechnikov, que são os que nos vão interessar mais, concluímos que  $\sup_{h > 0} L(h) \in \mathbb{R}^+$ , para quase todo o  $x^n \in \mathbb{R}^n$  (conclusão esta que pode ser generalizada para todo o núcleo positivo e limitado de suporte limitado).

Em 1983, Chow, Geman e Wu obtêm um resultado que estabelece condições para a convergência  $L_1$  do estimador atrás descrito. O enunciado que apresentamos a seguir é um melhoramento desse resultado (ver Devroye, 1984a, p.153).

**Teorema 3.2.1 :** Escolhamos  $h$  tal que  $L(h) \geq a \sup_{h > 0} L(h)$  para algum  $a \in ]0, 1[$ . Seja  $K \in \mathcal{F}$ , limitada, de suporte limitado, integrável à Riemann e satisfazendo  $K \geq c \mathbb{1}_{]-r, r[}$  para  $c$  e  $r$  positivos e fixos. Para todo o  $f \in \mathcal{F}$  de suporte limitado, temos que :

$$i) \int_{\mathbb{R}} |f_n(x) - f(x)| dx \longrightarrow 0, n \rightarrow \infty, \text{ q.c.}$$

$$ii) f_n(x) \longrightarrow f(x), n \rightarrow \infty, \text{ q.c.}, \text{ para quase todo o } x \in \mathbb{R}.$$

Apesar deste resultado de consistência, conhecemos alguns resultados acerca deste estimador que nos indicam que o seu comportamento a nível prático pode não ser o melhor.

Assim Shuster e Gregory provaram em 1981 (ver Devroye, 1984a, p.154), que o uso de núcleo limitados de suporte limitado na estimação de densidades em que  $X_{(n)} - X_{(n-1)} \rightarrow \infty, n \rightarrow \infty, p.$  (onde  $X_{(k)}$ , para  $k \in \mathbb{N}$ , representa a  $k$ -ésima estatística de ordem) é inconsistente para a convergência  $L_1$ . Como exemplo de densidades em que  $X_{(n)} - X_{(n-1)} \rightarrow \infty, n \rightarrow \infty, p.$  temos as densidades de Student, as de Pareto e todas as densidades estáveis excepto a normal.

Por outro lado, Hall em 1982, concluiu que se a densidade a estimar é concava no intervalo limitado  $[a,b]$ , que contém o seu suporte e aí possui derivadas de segunda ordem, então sob certas condições de regularidade em  $K$  o método anterior fornece-nos constantes  $h_n$  da ordem de  $n^{-1/3}$ , o que é, como vimos no parágrafo 2.5., suboptimal para uma larga classe de densidades de probabilidade (no sentido do EQMI).

Uma observação ainda acerca da utilização prática deste método. No caso da amostra extraída possuir valores repetidos (caso, por exemplo, da amostra dividida por classes) a aplicação deste método não é aconselhada pois ter-se-á, no caso da utilização de núcleos positivos satisfazendo  $K(0) \neq 0$ , que  $\lim L(h) = +\infty, h \rightarrow 0$ .

### 3.3. Minimização do EQI como critério para a escolha de $h_n$

O método que vamos expôr foi estudado e proposto por Stone em 1984 e a ideia da sua construção é a seguinte : Pretendemos determinar  $h > 0$  que, para  $n$  fixo, minimize a função de perda  $EQI(n,h) = \int_{\mathbb{R}} (f_n(x|h) - f(x))^2 dx$ . Tal não é possível pois desconhecemos

$\int_{\mathbb{R}} f_n(x|h)f(x) dx$ . A ideia será substituir na expressão do  $EQI(n,h)$  a quantidade anterior por

outra que possamos manejar e de tal modo que o processo de determinação de  $h$  assim obtido se aproxime da ideia inicial que seria determinar  $h > 0$  de modo a minimizar o  $EQI(n,h)$ .

Observe-se que, sem especificar condições em  $K$  ou em  $f$ ,

$$\int_{\mathbb{R}} f_n(x|h)f(x) dx = \frac{1}{nh} \sum_{i=1}^n E(K(\frac{X-X_i}{h})) \cong \frac{1}{n(n-1)h} \sum_{j=1}^n \sum_{i \neq j} K(\frac{X_j-X_i}{h})$$

utilizando o processo de validação cruzada já atrás referido. Não vamos considerar a aproximação assim obtida mas sim  $\frac{1}{n^2} \sum_{j=1}^n \sum_{i \neq j} K_h(X_j-X_i)$  onde  $K_h = \frac{1}{h} K(\frac{\cdot}{h})$ .

$$\text{Como } \int_{\mathbb{R}} f_n^2(x|h) dx = \frac{1}{n^2} \sum_{i,j=1}^n (K_h * K_h)(X_j-X_i) \text{ vamos designar por}$$

$$M_{n,h} = \frac{1}{n^2} \sum_{i,j=1}^n (K_h * K_h)(X_j-X_i) - \frac{2}{n^2} \sum_{j=1}^n \sum_{i \neq j} K_h(X_j-X_i)$$

e consideremos o seguinte procedimento para a escolha de  $h$  :

" Escolher  $h > 0$  de modo a minimizar  $M_{n,h}$  ".

Para a boa formulação de tudo o que foi exposto e do teorema de consistência que a seguir enunciaremos, Stone impõe as seguintes condições a  $K$  :

- i)  $\int_{\mathbb{R}} K(x) dx = 1$
- ii)  $K$  é simétrica relativamente à origem
- iii)  $K$  possui suporte limitado
- iv)  $\exists \alpha, \beta > 0 \forall x, y \in \mathbb{R} \quad |K(x) - K(y)| \leq \alpha |x - y|^\beta$
- v)  $(K * K)(0) < 2K(0)$

O núcleo de Epanechnikov satisfaz as condições anteriores.

Nas condições anteriores existe  $h_0 > 0$  tal que  $h_0$  minimiza  $M_{n,h}$  e além disso também existe  $h > 0$  que minimiza o  $EQI(n, h)$  (estas duas últimas observações são também válidas para o núcleo Normal).

**Teorema 3.3.1:** *Seja  $K$  satisfazendo as condições de i) a v) atrás citadas. Para  $h_n$  escolhido como acima foi referido então*

$$\lim_n \left( \frac{EQI(n, h_n)}{\min_{h > 0} EQI(n, h)} \right) = 1 \text{ q.c.}$$

para todo o  $f \in \mathcal{F}$  limitado.

De forma análoga ao que fizemos notar no fim do parágrafo anterior, também neste método podem surgir problemas quanto à sua utilização no caso da amostra extraída conter valores repetidos, pois poder-se-á ter, no caso da utilização de núcleos satisfazendo as condições i)-v), que  $\lim_{n,h} M_{n,h} = -\infty, h \rightarrow 0$ .

### 3.4. Escolha de $h_n$ "segundo Kappenmann"

O quarto e último método que apresentamos é devido a Kappenmann (1987) e ao contrário dos métodos referidos nos parágrafos anteriores não temos conhecimento, de qualquer resultado sobre a sua consistência.

O método consiste na obtenção de dois estimadores para  $E(f(X)) = \int_{\mathbb{R}} f^2(x) dx$  cada um deles dependente do parâmetro  $h$ . Tendo-se duas aproximações da mesma quantidade, será natural que uma escolha razoável de  $h$ , seja tomar  $h > 0$  de modo a tornar iguais estas duas funções de  $h$ .

Se  $f_n(x|h)$  é estimador de  $f(x)$  então o primeiro estimador de  $E(f(X))$  que vamos considerar é  $\int_{\mathbb{R}} f_n^2(x|h)(x) dx$ . Por outro lado, como  $E(f(X)) \cong \frac{1}{n} \sum_{j=1}^n f(X_j)$ , onde  $(X_1, X_2, \dots, X_n)$  é uma  $n$ -amostra de lei  $P_f$ , e sendo  $f_{n,j}(X_j)$  um estimador de  $f(X_j)$  baseado nas observações  $\{X_1, X_2, \dots, X_n\} \setminus \{X_j\}$ , vamos tomar como segundo estimador de  $E(f(X))$  a função de  $h$   $\frac{1}{n} \sum_{j=1}^n f_{n,j}(X_j)$ .

Consideremos então o seguinte procedimento para a escolha de  $h$  :

"Escolher para  $h$ , uma das raízes positivas da equação  $\int_{\mathbb{R}} f_n^2(x|h)(x) dx = \frac{1}{n} \sum_{j=1}^n f_{n,j}(X_j)$ "

No caso dos núcleos normal e de Epanechnikov imediatamente se prova que a equação anterior tem pelo menos uma raiz positiva.

### 3.5. Estudo de simulação. Resultados.

No estudo de simulação que fizemos e do qual apresentamos seguidamente os resultados, utilizámos as seguintes densidades de probabilidade para efectuar a comparação



entre os métodos que atrás apresentámos : normal centrada e reduzida, Cauchy, exponencial bilateral, t de Student com 5 graus de liberdade, mistura de normais ( $\frac{1}{2}N(-1.5,1) + \frac{1}{2}N(1.5,1)$ ), triangular ( $f(x) = (1 - |x|)\mathbb{1}_{[-1,1]}(x)$ ), gama de parâmetro 2 e exponencial de parâmetros 1 e 5.

Algumas destas densidades são habitualmente usadas em estudos do género como podemos encontrar em Davis (1981), Scott (1981), Bowman (1984), e Kappenmann (1987). Tentamos com esta escolha considerar densidades quer simétricas quer não, contínuas ou não, de suporte limitado e não limitado e também densidades com "diferentes pesos nas caudas" (esta última consideração aparece pois diversos autores são de opinião que a consistência de determinados estimadores automáticos da densidade tem a ver com o grande ou pequeno peso nas caudas da densidade a estimar, como já vimos por exemplo no parágrafo 3.2. Ver também Chow, 1983, p.27).

Refiro também que na aplicação prática que fizemos dos métodos expostos nos parágrafos 3.2. e 3.3. determinámos aproximadamente os pontos extremantes das funções  $L(h)$  e  $M_{n,h}$  por pesquisa num conjunto finito da forma  $\{\frac{2i}{100} \mid i \in \mathbb{N}\}$  onde  $N \subset \mathbb{N}$  é escolhido consoante a densidade de probabilidade. No caso das densidades exponenciais de parâmetros 1 e 5 utilizámos passo de pesquisa igual a  $\frac{1}{100}$ .

#### \* Eficiência

Sendo os resultados enunciados nos parágrafos anteriores de carácter assintótico, achámos que teria interesse analisar o comportamento dos estimadores aí apresentados, no caso de termos amostras de tamanho pequeno. Assim considerámos amostras de tamanho 10 e 50 tomando nas estimativas do EQMI 50 e 40 amostras, respectivamente.

Os resultados obtidos sobre a eficiência dos estimadores obtidos pelos métodos de escolha de  $h$  que referimos e que como já dissemos foram comparados tendo em conta o EQMI, são apresentados nos quadros 3.5.1. e 3.5.2.. Denotaremos por E.i.N. e E.i.E. os estimadores automáticos do núcleo abordados respectivamente nos parágrafo 3.i. ( $i=1,2,3,4$ ) quando usamos o núcleo normal ou o núcleo de Epanechnikov. Nos quadros referidos e por baixo do valor estimado para o EQMI surge entre parênteses o valor médio obtido para  $h_n$ .

Acerca da utilização dos núcleos normal e de Epanechnikov é de realçar o seguinte :

- No primeiro dos métodos abordado, devido à sua definição, não era de esperar grande diferença entre os dois estimadores, o que realmente se confirma para as densidades consideradas.

- Para o segundo tipo de estimadores considerado, os resultados dados pelo núcleo normal são melhores do que os do estimador E.2.E., excepção feita para  $n=50$  e no caso da

densidade triangular, que é afinal a única das densidades consideradas em que podemos garantir a consistência  $L_1$  e pontual de E.2.E..

Notar também a confirmação dos fracos resultados do estimador E.2.E. em densidades como a de Cauchy, a de Student e as exponenciais e ainda as alternativas dadas por E.4.N. e E.4.E. apesar de nada sabermos acerca da consistência destes últimos estimadores.

n=10	E.1.N.	E.1.E.	E.2.N.	E.2.E.	E.3.N.	E.3.E.	E.4.N.	E.4.E.
Normal	0.0339 (0.671)	0.0328 (0.664)	0.0431 (0.705)	0.0454 (0.668)	0.0606 (0.884)	0.0631 (0.773)	0.0423 (0.696)	0.0407 (0.708)
Cauchy	0.0676 (5.968)	0.0731 (5.85)	0.0536 (3.118)	0.0632 (3.311)	0.0534 (1.174)	0.058 (1.047)	0.0366 (1.222)	0.0369 (1.311)
Exponencial Bilateral	0.0454 (1.042)	0.0488 (1.030)	0.0504 (1.203)	0.0542 (1.16)	0.0742 (0.893)	0.0707 (0.798)	0.0471 (0.796)	0.0481 (0.827)
t(5) de Student	0.0337 (0.776)	0.0335 (0.769)	0.0531 (0.893)	0.0567 (0.852)	0.05 (1.011)	0.054 (0.906)	0.0408 (0.8)	0.0398 (0.803)
Mistura de Normais	0.0249 (1.184)	0.0266 (1.172)	0.0303 (0.979)	0.0317 (0.905)	0.0353 (1.574)	0.0379 (1.387)	0.0309 (1.258)	0.031 (1.35)
Triangular	0.0718 (0.274)	0.0699 (0.271)	0.0925 (0.287)	0.1031 (0.267)	0.1443 (0.371)	0.1326 (0.325)	0.0855 (0.301)	0.078 (0.31)
Gama (2)	0.0471 (0.966)	0.0191 (0.957)	0.0571 (0.945)	0.0607 (0.882)	0.0614 (0.898)	0.0668 (0.801)	0.0464 (0.804)	0.0476 (0.826)
Exponencial(1)	0.1571 (0.685)	0.1702 (0.678)	0.1566 (0.677)	0.176 (0.734)	0.1788 (0.457)	0.1876 (0.397)	0.1471 (0.429)	0.1562 (0.459)
Exponencial(5)	0.8173 (0.136)	0.8841 (0.135)	0.8357 (0.149)	0.919 (0.149)	0.845 (0.094)	0.8293 (0.082)	0.7579 (0.087)	0.7582 (0.093)

Quadro 3.5.1.

- Referindo-nos aos estimadores E.3.N. e E.3.E. (este último que sabemos consistente no sentido do teorema 3.3.1 para todas as densidades aqui consideradas) é de notar os significativos melhores resultados obtidos por E.3.N. que são bem notórios para  $n=50$ , apesar de nada ainda estar feito sobre a sua consistência.

- Sobre os estimadores E.4.N. e E.4.E., os resultados dos quadros mostram um grande equilíbrio nos resultados obtidos por um e outro, havendo contudo para  $n=50$ , uma ligeira vantagem para a utilização do núcleo de Epanechnikov.

De um modo geral, existe para  $n=10$  uma supremacia dos estimadores  $E_{i,j}$ . ( $i=1,4$ ;  $j=NE$ ) relativamente aos outros quatro estimadores, excepção feita ao caso da lei de Cauchy em que

qualquer dos estimadores apresentados tem vantagem sobre E.1.j.(j=E,N).

Para  $n=50$  e para a maioria das densidades consideradas E.4.E. obtêm os melhores resultados. Ressalvo contudo os casos das densidades normal e triangular onde E.1.j.(j=E,N) são significativamente melhores do que E.4.E..

n=50	E.1.N.	E.1.E.	E.2.N.	E.2.E.	E.3.N.	E.3.E.	E.4.N.	E.4.E.
Normal	0.0115 (0.472)	0.011 (0.468)	0.0146 (0.485)	0.0161 (0.482)	0.0137 (0.526)	0.0419 (0.481)	0.0149 (0.384)	0.0141 (0.384)
Cauchy	0.0808 (2.219)	0.0869 (2.197)	0.0731 (5.102)	0.0781 (5.241)	0.0136 (0.708)	0.0141 (0.657)	0.0335 (0.594)	0.0128 (0.6)
Exponencial Bilateral	0.02 (0.701)	0.0217 (0.694)	0.0282 (0.904)	0.0366 (1.004)	0.0199 (0.528)	0.019 (0.487)	0.0136 (0.424)	0.0179 (0.429)
t(5) de Student	0.0092 (0.616)	0.0093 (0.61)	0.0112 (0.689)	0.014 (0.686)	0.0124 (0.59)	0.0369 (0.544)	0.0096 (0.447)	0.0088 (0.453)
Mistura de Normais	0.0085 (0.885)	0.0093 (0.876)	0.0106 (0.597)	0.0111 (0.561)	0.0117 (0.739)	0.0477 (0.66)	0.0094 (0.649)	0.0088 (0.658)
Triangular	0.0218 (0.2)	0.0212 (0.198)	0.0339 (0.156)	0.0306 (0.151)	0.0318 (0.233)	0.0969 (0.22)	0.0261 (0.174)	0.0249 (0.177)
Gama (2)	0.0175 (0.695)	0.0196 (0.688)	0.017 (0.611)	0.0224 (0.684)	0.0149 (0.458)	0.052 (0.414)	0.0136 (0.403)	0.0136 (0.409)
Exponencial(1)	0.1138 (0.489)	0.1253 (0.484)	0.1107 (0.475)	0.1462 (0.654)	0.0778 (0.197)	0.1908 (0.177)	0.0755 (0.196)	0.0777 (0.201)
Exponencial(5)	0.5964 (0.098)	0.6549 (0.097)	0.5785 (0.096)	0.7983 (0.149)	0.3906 (0.042)	0.8574 (0.038)	0.3924 (0.039)	0.4033 (0.04)

Quadro 3.5.2.

Outros estudos de simulação podem ser encontrados em Scott (1981) (onde E.2.N. é comparado com dois outros estimadores automáticos), em Bowman (1984) (onde E.2.N. é comparado com um método cuja ideia de determinação de  $h$  é semelhante ao método apresentado por Stone) ou em Kapenmann (1987) (onde E.4.N. é comparado com os dois estimadores analisados por Bowman em 1984).

#### \* Sensibilidade

Continuando na análise das propriedades a nível prático dos métodos apresentados efectuámos uma análise breve sobre a sensibilidade dos métodos M.1.N., M.2.N., M.3.N. e M.4.E. (métodos que dão origem aos estimadores E.1.N., E.2.N., E.3.N. e E.4.E. respectivamente) que expusemos, ou seja, queremos saber se existe grande variação no parâmetro  $h$  com a varia-

ção na amostra extraída, de um dos seus elementos.

Assim obtivemos uma amostra de tamanho 24 de uma lei normal centrada e reduzida e acrescentámos a essa amostra um vigésimo quinto elemento  $x \in \mathbb{R}$ . A tabela seguinte apresenta-nos os valores obtidos para o parâmetro  $h$  por cada um dos métodos e para alguns valores de  $x$ .

$x$	M.1.N.	M.2.N.	M.3.N.	M.4.E.
-10	1.182	2.205	0.340	0.429
-8	0.976	1.845	0.340	0.429
-6	0.779	1.480	0.340	0.429
-4	0.600	1.065	0.340	0.429
-2	0.461	0.550	0.365	0.457
0	0.406	0.525	0.345	0.492
2	0.464	0.335	0.345	0.493
4	0.605	0.775	0.340	0.430
6	0.784	1.345	0.340	0.429
8	0.981	1.735	0.340	0.429
10	1.187	2.110	0.340	0.429

Quadro 3.5.3.

É notória a insensibilidade dos dois últimos métodos sendo portanto mais um factor em abono dos dois últimos tipos de estimadores que expusemos (análises semelhantes podem ser encontradas em Scott (1981) e Kapenmann (1987)).

Esta abordagem sucinta sobre o estimador automático do núcleo deixa-nos, no entanto, várias hipóteses para uma possível investigação.

Recordando a definição do primeiro estimador automático que considerámos, podemos concluir, que se o processo de cálculo de  $h_n$  for dado por  $h_n(X_1, X_2, \dots, X_n) = M c_k S_n n^{-1/5}$ , (para  $M \in \mathbb{R}$ , fixo, mas significativamente maior do que o valor  $a$  utilizado) o estimador resultante continua a verificar as propriedades assintóticas que enunciámos nos teoremas 3.1.1, 3.1.2, 3.1.3. Não será no entanto difícil verificar que o estimador resultante de tal escolha tem péssimas propriedades de eficiência. Isto levanta-nos o problema de que para o estudo destes estimadores, não nos interessarão apenas processos de estudo que caracterizam as suas propriedades assintóticas, mas também critérios que nos permitam inferir as suas propriedades de eficiência.

Para o segundo método que abordamos, é conhecido o teorema de consistência 3.2.1. No entanto ele é bastante limitativo, quer nos núcleos a usar, quer principalmente, nas densidades para as quais podemos garantir a sua consistência. Para esses mesmos núcleos temos a inconsistência  $L_1$  para densidades com "grande peso nas caudas". Os resultados de simulação obtidos são favoráveis ao uso do núcleo normal aquando da estimação de densidades de suporte não limitado, o que nos leva a crer na possibilidade de que se venham a estabelecer resultados sobre a sua consistência.

Um caso semelhante ao anterior é-nos dado pelos estimadores considerados no parágrafo 3.3. Ai é notório os melhores resultados obtidos com o núcleo normal, apesar de não conhecermos resultados sobre a sua consistência.

O quarto método que analisamos, é o único acerca do qual não são conhecidas quaisquer propriedades de consistência, apesar de que, com tão bons resultados de eficiência, é natural que estas se venham a estabelecer.

Muitos foram os problemas, para os quais não existem ainda respostas satisfatórias, que não chegámos a abordar e cuja importância é determinante na estimação da densidade de probabilidade, a nível de aplicações práticas. Referimo-nos por exemplo ao estabelecimento de regiões de confiança para  $f$  e à construção de testes. São fundamentalmente estes dois instrumentos que nos indicam em que medida é que podemos "confiar" na estimativa construída para  $f$ .

## BIBLIOGRAFIA

- ABOU-JAOUDE,S. (1974) Sur un Théorème de non existence d'estimateurs convergent en probabilité.  
C.R.Acad. Sc. Paris , t.278 , série A , p.1445-1448.
- ABOU-JAOUDE,S. (1976) Condition nécessaires et suffisantes de convergence  $L_1$  en probabilité de l'histogramme pour une densité.  
Ann. Inst. Henri Poincaré , 12 , 3 , p.213-231.
- ABOU-JAOUDE,S. (1976a) 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. Sc. Paris , t.283 , série A , p.1107-1110.
- ABOU-JAOUDE,S. (1977) La convergence  $L_1$  et  $L_\infty$  de certains estimateurs d'une densité de probabilité.  
These de Doctorat D'Etat , Université Pierre et Marie Curie.
- BERTRAND-RETALI,M. (1974) Convergence uniforme stochastique d'un estimateur d'une densité de probabilité.  
C.R.Acad. Sc. Paris , t.278 , série-A , p.451-453.
- BERTRAND-RETALI,M. (1978) Convergence uniforme d'un estimateur de la densité par la méthode du noyau .  
Revue. Roum. Math. Pures et Appl. , 13 , 3 , p.361-385.
- BOSQ,D. ; LEOUTRE,J.-P. (1987) Théorie de l'estimation Fonctional.  
Economica.
- BOWMAN,A.W. (1984) An alternative methods of cross-validation for the smoothing of density estimates.  
Biometrika , 71 , 2 , p.353-360.

- CACOULOS,T. (1966) Estimation of a multivariate density.  
Ann.Inst.Stat.Math. , 18 , p.179-189.
- CHOW,Y.-S. ;GEMAN,S. ; WU,L.-D. (1983) Consistent cross-validated density estimation.  
Ann. Statist. , 11 , 1 , p.25-38.
- DAVIS,K.B. (1981) Estimation of the scaling parameter for Kernel-type density estimates.  
Jour. Amer. Statist. Assoc. , 76 , 375 , p.632-636.
- DEHEUVELS,P. (1977) Estimation non paramétrique de la densité par histogrammes généralisés.  
Revue Statist. Appliq. , 25 , 3 , p.5-42.
- DEHEUVELS,P. ; HOMINAL,P.(1980) Estimation automatique de la densité.  
Revue Statist. Appliq. , 28 , 1 , p.25-55.
- DEVROYE,L. ; WAGNER,T.J. (1979) The  $L_1$  convergence of Kernel density estimates.  
Ann. Statist. , 7 , 5 , p.1136-1139.
- DEVROYE,L. (1983) The equivalence of weak, strong and complete convergence in  $L_1$  for Kernel density estimates.  
Ann. Statist. , 12 , 4 , p.1231-1244.
- DEVROYE,L. ; PENROD,C.S. (1984) The consistency of automatic Kernel density Estimates.  
Ann. Statist. , 11 , 3 , p.896-904.
- DEVROYE,L. (1984a) Nonparametric Density Estimation.  
The  $L_1$  view . Wiley , New York.
- EPANECHNIKOV,V.A. (1968) Non-parametric estimation of a multivariate probability density.  
Theory Prob.Appl. , 10 , p.153-158.
- FRECHET,M. (1937) Généralités sur les Probabilités. Variables Aléatoires.  
Paris. Gauthier-Villars, Éditeur.
- FREEDMAN,D. ;DIACONIS,P.(1981) On the histogram as a density estimator :  $L_2$  theory.  
Z. Wahrscheinlichkeitstheorie verw. Gebiete , 57 , p.453-476.

- GEFFROY, J. (1974) Sur l'estimation d'une densité dans un espace métrique.  
C.R.Acad. Sc. Paris , t.278 , série-A , p.1449-1452.
- GMURMAN, V.E. (1977) Teoria das Probabilidades e Estatística Matemática.  
Editora Mir, Moscou.
- HALL, P. (1982) Cross-validation in density estimation.  
Biometrika , 69 , 2 , p.383-390.
- HALL, P. (1983) Large sample optimality of least squares cross-validation in density estimation.  
Ann. Statist. , 11 , 4 , p.1156-1174.
- KAPPENMANN, R.F. (1987) A nonparametric data based univariate density function estimate.  
Computational Statistics & Data Analysis , 5 , p.1-7.
- LEADBETTER, M.R.; WATSON, G.S. (1963) On the estimation of probability density, I.  
Inst. Ann. Math. Statist. , 43 , p.480-491.
- LECOUTRE, J.-P. (1985) The  $L_2$ -optimal cell width for the histogram.  
Statistics & Probability Letters , 3 , p.303-306.
- LUENBERGER, D.G. (1969) Optimization By Vector Space Methods.  
Wiley, New York.
- MÉTIVIER, M. (1972) Notions fondamentales de la Théorie des Probabilités.  
Paris. Dunod.
- NADARAYA, E.A. (1965) On non-parametric estimates of density functions and regression curves.  
Theory Prob. Appl. , 19 , p.133-141.
- NADARAYA, E.A. (1974) On the integral mean square error of some nonparametric estimates for the density function.  
Theory Prob. Appl. , 10 , p.186-190.
- PARZEN, E. (1962) On estimation of a probability density function and mode.  
Ann. Math. Statist. , 33 , p.1065-1076.



- PRAKASA RAO, B.L.S. (1983) *Nonparametric Functional Estimation*.  
Academic Press, New York.
- RÉVESZ, P. (1972) *On empirical density function*.  
*Periodica Mathematica Hungarica*, 2, p.85-110.
- ROSENBLATT, M. (1956) *Remarks on some nonparametric estimates of a density function*.  
*Ann. Statist.*, 27, p.832-837.
- SCHUSTER, E.F. (1969) *Estimation of a probability density function and its derivatives*.  
*Ann. Mathe. Statist.*, 40, p.1187-1195.
- SCHUSTER, E.F. (1970) *Note on the uniform convergence of density estimates*.  
*Ann. Mathe. Statist.*, 41, p.1347-1348.
- SCHUSTER, E.F. ; GREGORY, G.G. (1981) *On the non-consistency of maximum likelihood nonparametric density estimators*.  
*I.M.S. Bull.*, 10, 215.
- SCOTT, D.W. (1979) *On optimal and data-based histograms*.  
*Biometrika*, 66, 3, p.605-610.
- SCOTT, D.W. ; FACTOR, L.E. (1981) *Monte Carlo study of three data-based nonparametric probability density estimators*.  
*Jour. Amer. Statist. Assoc.*, 76, 373, p.9-15
- STEIN, E.M. (1970) *Singular Integrals and Differentiability Properties of Functions*.  
University Press, Princeton, New Jersey.
- STONE, C.J. (1984) *An asymptotically optimal window selection rule for Kernel density estimates*.  
*Ann. Statist.*, 12, 4, p.1285-1297.
- WAGNER, T.J. (1975) *Nonparametric estimates of probability densities*.  
*IEEE Trans. Infor. Theory*, IT-21, 4, p.438-440.
- WEGMAN, E.J. (1972) *Nonparametric probability density estimation : I. A summary of available methods*.  
*Technometrics*, 14, 3, p.533-546.

## ERRATA

página (linha)	onde está	deve estar
9 (6ª)	A é limitado e	e A limitado,
14 (2ª fim)	1.3.2	1.3.5
15 (9ª)	$\sqrt{2} \propto \exp(-\frac{\varepsilon^2 n}{8})$	$\frac{\alpha}{\sqrt{n}}$
21 (2ª)	$\frac{\eta}{3} + \varepsilon$	$\eta + \varepsilon$
(4ª fim)	$\beta$	$\beta_n$
22 (6ª)	$[0, \mu(A)]$	$[0, \mu(A)]$
(7ª)	$A_\varepsilon$	$(a)_\varepsilon$
(3ª fim)	$A_\varepsilon$	$(\Delta^*_{n,r})_\varepsilon$
28 (7ª fim)	$\mathbb{R}^s - ]q, q[^s$	$\mathbb{R}^s - ]-q, q[^s$
31 (7ª fim e 1ª fim)	EQMIA'	EQMI
42 (3ª)	1976	1977
43 (7ª fim)	ii)	i)
47 (11ª)	B um boreliano de $\mathbb{R}^s$ tal que $0 < \lambda(B) < +\infty$	B um boreliano limitado de $\mathbb{R}^s$ tal que $\lambda(B) > 0$
50 (6ª)		suprimir
(7ª)	$\mathbb{1}_{(K(x) \subseteq M)}$	$\mathbb{1}_{(K(x) \subseteq M)}(x)$
(6ª fim)	$\int_{\mathbb{R}^s} K^*(y) dy$	$\int_{B(0,L)} K^*(y) dy$
(5ª fim)	$\geq \int_{\mathbb{R}^s} K^*(y) dy$	$= \int_{B(0,L)} K^*(y) dy$
53 (9ª)	deve estar : $\prod_{n=1}^s \{ \prod_{i=1}^s [a_i h_n, (a_i+1) h_n] \mid (a_1, a_2, \dots, a_s) \in \mathbb{Z}^s \}, n \in \mathbb{N}$ com $h_n = 1$ , se n par e $h_n = n^{-1/2s}$ , se n impar	
(10ª e 11ª)	$\Delta_n$	$\hat{\Delta}_n$
54 (7ª)	$4x - 2$	$4x + 2$
56 (9ª)		suprimir (2.4.5)
59 (5ª fim e 1ª fim)	$1 \leq t_1 \leq t_2 \leq \dots \leq t_j \leq k_n^s$	$1 \leq t_1 < t_2 < \dots < t_j \leq k_n^s$
60 (2ª)	"	"
61 (8ª fim)	"	"

(5 <sup>a</sup> )	$\frac{x-u}{h_n}$	$\frac{x_n-u}{h_n}$
(7 <sup>a</sup> )	$E(f_n(x_n)   e_n)$	$E( f_n(x_n)    e_n)$
63 (10 <sup>a</sup> )	$t$	$t'$
65 (5 <sup>a</sup> e 6 <sup>a</sup> )	$h_n 2^m$	$h_n / 2^m$
(7 <sup>a</sup> )	$h'_n$	$(h'_n)^{-s}$
66 (6 <sup>a</sup> )	$ f_n^{\sim}(x-h_n a_r) - f(x) $	$ f_n^{\sim}(x-h_n a_r) - f(x) $
67 (4 <sup>a</sup> fim)	$f$	$f \in L_2(\mathbb{R}^s)$
	diferenciável	continuamente diferenciável
(3 <sup>a</sup> fim)	$L_1(\mathbb{R}^s) \cap L_2(\mathbb{R}^s)$	$L_2(\mathbb{R}^s)$ e sejam limitadas
68 (1 <sup>a</sup> )	$dy_1$	$dy$
(4 <sup>a</sup> )	deve estar : Atendendo a que $f \in L_2(\mathbb{R}^s)$ , $\int_{\mathbb{R}^s} E^2[f_n(x)] dx \leq \int_{\mathbb{R}^s} f^2(x) dx$ e	
(6 <sup>a</sup> fim)	$\frac{\partial^2 f}{\partial x_i \partial x_j} * \frac{\partial^2 f}{\partial x_k \partial x_l} (0)$	$\int_{\mathbb{R}^s} \frac{\partial^2 f}{\partial x_i \partial x_j} (y) \frac{\partial^2 f}{\partial x_k \partial x_l} (y) dy$
69 (1 <sup>a</sup> )	"	"
72 (12 <sup>a</sup> )	$K_2(t) dt$	$K_2(t) dt$
73 (3 <sup>a</sup> )	"	"
(12 <sup>a</sup> )	$K_0$	$K$
76 (14 <sup>a</sup> )	referido	referidos
82 (5 <sup>a</sup> fim)	$\frac{1}{h_n(X_1, X_2, \dots, X_n)}$	$\frac{1}{nh_n(X_1, X_2, \dots, X_n)}$
83 (2 <sup>a</sup> fim)	$b=1$	$\sigma=1$
85 (7 <sup>a</sup> fim)	positivo	positivo e limitado
(4 <sup>a</sup> fim)	183	153
87 (6 <sup>a</sup> )	$\frac{1}{n^2 h}$	$\frac{2}{n^2}$
88 (3 <sup>a</sup> )	ter-se-á	poder-se-á ter
89 (6 <sup>a</sup> fim)	Acerca	Acerca