

**Splines na estimação não paramétrica da
densidade de probabilidade de uma
variável aleatória não negativa**

Alexandra Teles Resende



Splines na estimação não paramétrica da densidade de probabilidade de uma variável aleatória não negativa

Alexandra Teles Resende

Dissertação para a obtenção do Grau de **Mestre em Matemática**
Área de Especialização em **Estatística, Optimização e Matemática Financeira**

Júri

Presidente: Maria de Nazaré Simões Quadros Mendes Lopes
Orientador: João Luís Cardoso Soares
Vogal: Maria Emília de Mesquita Nogueira

Data: 22 de Junho de 2012

Resumo

Este trabalho, escrito em estilo Latex, consiste numa dissertação de Mestrado em Matemática para obtenção do grau de Mestre em Estatística, Optimização e Matemática Financeira.

O objectivo desta dissertação consiste em encontrar uma função suave e de *área unitária* no seu intervalo de definição que aproxime funções seccionalmente constantes em subintervalos do mesmo domínio, também estas de área um, como descrevemos na secção 1.1. Na secção 1.2 encontra-se um exemplo de aplicação deste problema. Na secção 1.3 Figura a pesquisa bibliográfica e na secção 1.4 encontra-se a nossa ideia para abordar o problema.

No capítulo 2 descrevemos dois métodos não paramétricos para a estimação da função densidade de probabilidade de uma variável aleatória real, o histograma e o método do núcleo.

A nossa abordagem tem como apoio a aproximação de uma *função quadrática*, assim como o uso de *splines cúbicos* para encontrar esta mesma função de aproximação. Este desenvolvimento encontra-se no capítulo 3. Aí são definidas as funções spline quadrático e *spline cúbico* $H(t)$, $t \in [x_0, x_n]$, tais são definidas por troços no conjunto de nós $\{x_0, x_1, \dots, x_n\}$, satisfazendo $x_0 < x_1 < \dots < x_n$.

Neste mesmo capítulo caracterizamos a função $H(t)$ com o sistema $B'M = C'H$, em que as componentes do vector H são dadas por $H(x_i)$, $i = 1, \dots, n - 1$. Para que este sistema tenha solução única temos que impor condições em x_0 e x_n e por isso são abordados três casos, o primeiro caso consiste num sistema que “arranca” e “termina” a uma velocidade constante. O segundo caso é referente a um sistema que já está em funcionamento no instante inicial x_0 , e não termina abruptamente no instante final x_n (os nós x_0 e x_n funcionam como nós internos). E por fim o terceiro caso, os nós extremos x_0 e x_n identificam o mesmo ponto.

Ainda no mesmo capítulo enunciamos e demonstramos que o spline quadrático interpolador com área fixa em cada subintervalo reduz-se ao problema de encontrar um determinado *spline cúbico* interpolador (sem restrição de área). De seguida consideramos o problema de encontrar um *spline cúbico* com área previamente fixada em cada subintervalo, para diferentes condições iniciais, acompanhado de uma ilustração.

No fim deste capítulo encontra-se a descrição em notação matricial do problema original como um problema no qual a sua função objectivo consite numa *função quadrática* e com restrições de *área unitária* e de *splines cúbicos*. Aplicamos a nossa formulação a um pequeno exemplo para diferentes condições iniciais.

Por fim, no capítulo 4 testamos computacionalmente a nossa abordagem num conjunto de dados reais e em observações da lei exponencial geradas aleatoriamente. Para testar a abordagem descrita foi utilizado o programa que se encontra na secção B.9, o seu funcionamento é descrito na secção 4.2.

Palavras Chave: área unitária, splines cúbicos, função quadrática

Abstract

This text, written in Latex, corresponds to a dissertation in Master in Mathematics formation, to obtain the Master Degree in Statistics, Optimization and Financial Mathematics.

The goal of this thesis consists in finding a smooth function of area one on its domain that approximates piecewise constant functions where each of these constant sections has area one as well, as we will describe in section 1.1. In section 1.2 we present an example of an application of this problem. Section 1.3 contains the bibliographical research and in section 1.4 we present our idea to approach the problem.

In chapter 2 we describe two non parametric methods to estimate the probability density function of a real random variable, the histogram and the kernel estimator.

Our approach is supported by the approximation of a quadratic function and the utilization of *cubic splines* to find an approximate function. We develop this in chapter 3. We define there the quadratic spline and the *cubic spline* $H(t)$, $t \in [x_0, x_n]$, defined piecewise in a set of knots $\{x_0, x_1, \dots, x_n\}$, satisfying $x_0 < x_1 < \dots < x_n$.

In this same chapter we characterize the function $H(t)$ with the system $B'M = C'H$, where the components of the vector H are given by $h(x_i)$, $i = 1, \dots, n - 1$. In order for this system to have a unique solution we must impose some conditions in x_0 and x_n therefore we separate the problem in three cases. In the first case the system "starts" and "finishes" at a constant speed. The second one refers to a system that is already functioning at the starting time x_0 and doesn't finish abruptly at the final instant x_n (the knots x_0 and x_n can be treated as internal knots). Finally, in the third case, the extreme knots identify the same point.

Still in the same chapter we state and show that the interpolation quadratic spline with fixed area in each sub-interval can be reduced to the problem of finding a certain interpolation cubic spline (without area restrictions). After that we shall consider the problem of finding a cubic spline with previously fixed area in each sub-interval, for different initial conditions, with a picture to illustrate.

In the end of the chapter we present the description in matrix notation of the original problem as a problem where the goal function consists of a quadratic function with area one and cubic splines restrictions.

Finally, in chapter 4 we test computationally our approach with real data and observations of the exponential law, generated randomly. To test the described approach we used the program from section B.9 and the way it works it described in section 4.2.

Keywords: area one, cubic spline, quadratic function

Agradecimentos

Gostaria de agradecer ao Prof. Dr. João Luís Soares por toda a sua disponibilidade, correcção, orientação e sugestões ao longo de todo o trabalho assim como por tudo o que me ensinou.

Devo também os meus agradecimentos ao Prof. Dr. Carlos Tenreiro pela sua ajuda fundamental na bibliografia e programação do estimador do núcleo.

Agradeço aos meus pais pelo seu apoio incondicional e fundamental.

Agradeço também ao Departamento de Matemática da Universidade de Coimbra, professores e colegas, por todos os conhecimentos adquiridos e pelo convívio proporcionado.

Conteúdo

1	Introdução	1
1.1	O problema	1
1.2	A Motivação	2
1.3	Pesquisa bibliográfica	3
1.4	Ideia	4
2	Estimadores não paramétricos da densidade de probabilidade	5
2.1	Histograma	5
2.2	Estimador do Núcleo	6
3	Desenvolvimento	11
3.1	Spline quadráticos e cúbicos	11
3.2	A nossa formulação	21
3.2.1	A função objectivo	21
3.2.2	As restrições	23
3.2.3	O problema de optimização	24
4	Resultados Numéricos	29
4.1	Dados	29
4.2	Programa	30
4.3	Resultados preliminares com dados reais	31
4.4	Tamanho do <i>bin</i>	34
4.5	Método do núcleo	35
4.5.1	Programa	35
4.5.2	Resultados	35
4.6	Resultados com a lei exponencial	37
5	Conclusões	41
A	Histogramas	43
B	Listagem de programas	45
B.1	dados.m	45
B.2	Programa programa_abcissas_R.m	48
B.3	Programa estimadorNucleo.txt	50
B.4	Programa programa_abcissas_R_exemplo_histogramas.m	50
B.5	Função cubicsplinearea.m	52
B.6	Programa dados_problema_dummy_areafixa.m	56
B.7	Programa dados_problema_dummy2_areafixa.m	56
B.8	dados_leiexponencial.m	57
B.9	Programa tese.m	57
B.10	Programa densidade_lei_exponencial_media1.txt	72

Capítulo 1

Introdução

1.1. O problema

O problema que aqui abordamos é um problema de teoria das probabilidades e estatística, o de encontrar a densidade de probabilidade a partir do conhecimento de m histogramas independentes, com os mesmos *bins*, previamente fixados, provenientes de uma mesma distribuição de probabilidade.

O problema pode descrever-se em termos matemáticos do seguinte modo. Consideremos uma partição de um intervalo $[a, b)$ em n subintervalos,

$$[x_0, x_1), [x_1, x_2), \dots, [x_{n-1}, x_n),$$

onde $a \equiv x_0 < x_1 < \dots < x_n \equiv b$. Em cada subintervalo $[x_{i-1}, x_i)$ conhecem-se m escalares

$$f_{i,1}, f_{i,2}, \dots, f_{i,m} \quad (1.1)$$

- escalares esses que caracterizam o i -ésimo histograma - que satisfazem as condições

$$\sum_{i=1}^n f_{ij} = 1, \text{ para todo } j = 1, 2, \dots, m, \quad (1.2a)$$

$$f_{ij} \geq 0, \text{ para todo } j = 1, 2, \dots, m \text{ e } i = 1, 2, \dots, n. \quad (1.2b)$$

Pretendemos encontrar uma função *suave* $H: [a, b) \rightarrow \mathbb{R}$ que satisfaça as condições

$$\int_a^b H(t) dt = 1, \quad (1.3a)$$

$$H(t) \geq 0, \text{ para todo } t \in [a, b). \quad (1.3b)$$

De tal modo que

$$\int_{x_{i-1}}^{x_i} H(t) dt \approx f_{i,j},$$

para todo $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, m$.

1.2. A Motivação

Um processo de contagem $\{N(t), t \geq 0\}$ diz-se um processo de Poisson não estacionário com intensidade $\lambda(t)$ com $t \geq 0$ se:

1. $N(0) = 0$
2. $\{N(t), t \geq 0\}$ é de acréscimos independentes
3. $P\{N(t+h) - N(t) \geq 2\} = o(h)$
4. $P\{N(t+h) - N(t) = 1\} = \lambda(t)h + o(h)$

Com

$$m(t) = \int_0^t \lambda(s) ds$$

temos

$$P\{N(t+s) - N(t) = n\} = \frac{[m(t+s) - m(t)]^n}{n!} e^{-(m(t+s) - m(t))}, \quad n \geq 0$$

Assim, conhecido o histórico do número de chegadas em cada período de tempo (por exemplo, 1/4 de hora) num processo de Poisson não estacionário com intensidade $\lambda(t)$ (desconhecida) ao longo de um período longo pretendemos estimar a função $\lambda(t)$.

Num sistema que seja definido por uma função $\lambda(t)$ periódica a estimação de $\lambda(t)$ é semelhante a estimar a função densidade de uma variável aleatória não negativa por métodos não paramétricos (após adequado escalonamento).

Os dados que se encontram na secção B.1 do apêndice B correspondem ao número de chegadas (telefonemas) que chegaram a um serviço de call center em cada período de 1/4 de hora ao longo de um dia completo de 24h (desde as 0:00am até às 12:00pm) em oito segundas-feiras, consideradas normais ou típicas. Estes dados são reais.

1.3. Pesquisa bibliográfica

Segundo *Silverman* [2] os métodos não paramétricos para estimar a função densidade de uma variável aleatória não negativa são:

- Histograma;
- Estimador do núcleo;
- Método do vizinho mais próximo;
- Método do núcleo variável;
- Séries de Fourier.

No nosso trabalho utilizaremos apenas o histograma e o estimador do núcleo. A bibliografia fundamental que recorreremos para estes métodos foi [2] e [3].

Na tese de César (2009) [1] procurou-se um *spline cúbico* $H(t), t \in [a, b]$ que tornasse mínimo o valor de

$$\sum_{i=1}^n \sum_{j=1}^m \left(\int_{x_{i-1}}^{x_i} H(t) dt - f_i^j \right)^2$$

Ou seja, procurou-se a aproximação no sentido dos mínimos quadrados, com o uso de splines cúbicos. Esta ideia não funcionou, pois verificaram-se muitas oscilações nos resultados como podemos ver na Figura 1.1 seguinte, retirada de [1]

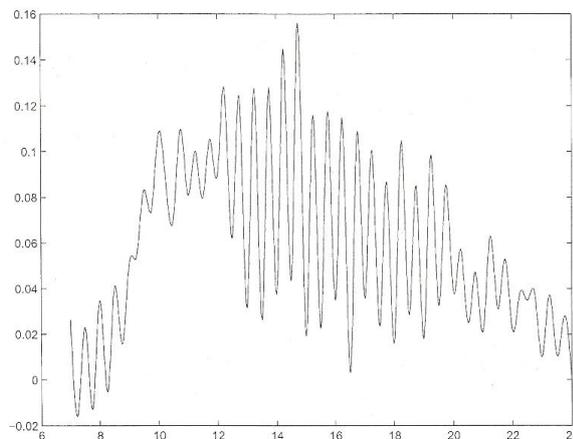


Figura 1.1: Call Center com observações de 1/4 em 1/4 hora durante 8 segundas-feiras

1.4. Ideia

Neste trabalho procuraremos obviar algumas limitações encontradas no trabalho de César [1]. Assim, procuraremos um *spline cúbico* $H(t), t \in [a, b]$ que torne mínimo o valor de

$$\sum_{i=1}^n \sum_{j=1}^m \int_{x_{i-1}}^{x_i} \left(H(t) - \frac{f_i^j}{h_i} \right)^2 dt.$$

Capítulo 2

Estimadores não paramétricos da densidade de probabilidade

Nesta secção vamos descrever dois métodos não paramétricos para a estimação da função densidade de probabilidade de uma variável aleatória real, a partir do conhecimento de um amostra de n observações, X_1, X_2, \dots, X_n . Os métodos considerados são o histograma e o método do núcleo.

2.1. Histograma

O histograma é o estimador da densidade de probabilidade mais simples. Para uma dada amostra X_1, X_2, \dots, X_n , a construção do histograma depende da escolha para os valores de dois parâmetros: a origem x_0 e a largura h (de cada *bin*). Com essa escolha fica determinada uma partição do intervalo $[x_0, +\infty)$ em subintervalos

$$[x_0, x_0 + h), [x_0 + h, x_0 + 2h), \dots, [x_0 + mh, x_0 + (m + 1)h), \dots$$

onde m denota um inteiro não negativo. A cada um destes subintervalos chamamos *bin* ou caixa. O estimador histograma é uma função de domínio $[x_0, +\infty)$ e definido por

$$\hat{f}(x) = \frac{1}{nh} \times (\text{Número de observações } X_i \text{ na mesma bin onde está } x)$$

A função $\hat{f}(x)$ é não negativa, seccionalmente constante, e o seu integral é unitário em $[x_0, +\infty)$.

Em geral, o estimador histograma, apesar de ser muito simples de obter, é sensível à escolha de x_0 e h . A largura da caixa ou *bin*, o parâmetro h , controla a suavidade da função $\hat{f}(x)$. Quanto maior h , menos patamares se tem e menos se observa um eventual comportamento oscilatório. A origem x_0 também tem influência no estimador pois uma má escolha pode conduzir a conclusões erradas.

Outra das eventuais desvantagens de uma utilização “crua” do histograma tem a ver com a ausência total de diferenciabilidade. Quando existe interesse em obter

informação diferencial sobre a densidade de probabilidade outra classe de métodos é mais adequada, como aquela que descrevemos a seguir.

2.2. Estimador do Núcleo

Seja K uma função não negativa, denominada *função núcleo*, que satisfaz a seguinte condição

$$\int_{-\infty}^{+\infty} K(x) dx = 1. \quad (2.1)$$

Esta função pode ser, por exemplo, a função densidade de probabilidade de uma lei normal centrada e reduzida,

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

A ideia será usar o valor $K(x)$ como uma medida da *distância* de um objecto x ao centro 0. Quanto maior for $K(x)$ mais próximo está x de 0. Então, o estimador do Núcleo é a função

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - X_i}{h}\right) \quad (2.2)$$

onde h é um escalar positivo denominado *janela*. Para cada observação X_i , a função *núcleo relativamente a X_i* ,

$$\frac{1}{h} K\left(\frac{x - X_i}{h}\right) \quad (2.3)$$

é uma densidade de probabilidade que avalia a dispersão de $x \in (-\infty, +\infty)$ em torno de X_i . Valores maiores de (2.3) correspondem a valor de x mais próximos de X_i . Posto isto, o estimador do núcleo (2.2) é a média de todos os núcleos relativamente a cada observação da amostra. Assim, em regiões onde houver muitas observações da amostra é esperado que $\hat{f}(x)$ assumia valores maiores.

A escolha da função K não é determinante para a qualidade do estimador [3]. Já o mesmo não se passa com a escolha da janela h . Vamos ilustrar com os dados da secção B.1 no apêndice B correspondentes à primeira segunda-feira.

Atendendo a que o método do núcleo requer os instantes em que efectivamente ocorreram as chegadas tivemos de proceder à sua geração (determinística) do seguinte modo. Se num dado período de 0.25 horas ocorreram s chamadas então, para $\delta = 0.25/s$, assumimos que a primeira chamada ocorreu no instante $\delta/2$, a segunda em $\delta/2 + \delta$, e por aí fora até à última no instante $\delta/2 + (s - 1)\delta (= 0.25 - \delta/2)$,

conforme programa B.2 (apêndice B) em MATLAB. Depois, utilizámos o programa B.3 (apêndice B) para obtermos assim os gráficos da estimação da densidade do núcleo. Para utilizar este programa foi utilizado o *software* R (R Development Core Team, 2009) com a *package* KernSmooth (S original by Matt Wand. R port by Brian Ripley, 2009)

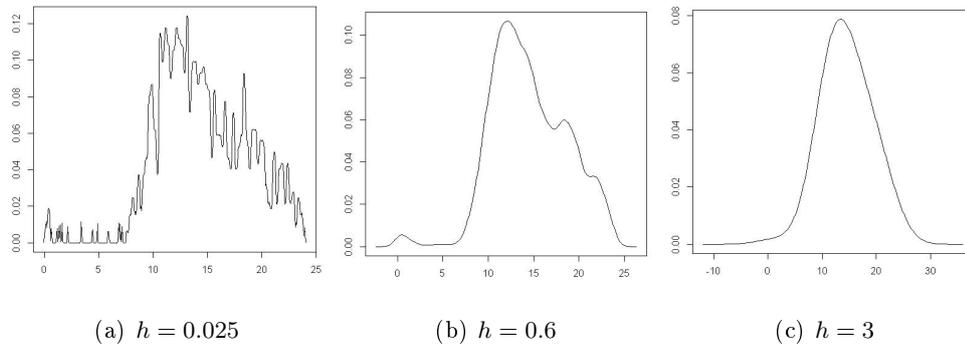


Figura 2.1: Estimador do Núcleo com os dados referentes da primeira segunda-feira dos dados (da secção B.1) com janela: (a) $h = 0.025$, (b) $h = 0.6$ e (c) $h = 3$.

Na Figura 2.1 observa-se que para diferentes h 's obtemos diferentes resultados. Para um h pequeno ($h = 0.025$) temos muitas oscilações. Para um h um pouco maior ($h = 0.6$) já temos menos oscilações, enquanto para um h grande ($h = 3$) já não temos o problema de oscilações mas temos o problema de obter valores negativos.

Em geral, $\hat{f}(x)$ é uma função diferenciável em todo $(-\infty, +\infty)$. Na Figura 2.1 nota-se que uma das desvantagens do estimador do núcleo é ser uma função que requer, em geral, todo o conjunto de números reais para ser uma densidade de probabilidade - e não apenas o conjunto dos reais não negativos ou um conjunto limitado como é do nosso interesse.

Ilustremos agora o método do núcleo para um dado exemplo correspondente aos dois histogramas das Figuras A.1 e A.2 (apêndice A). Trata-se de um exemplo muito simples que mais adiante também usaremos para testar em uma outra abordagem. Aqui através do programa B.4 (apêndice B) em código MATLAB gerámos os instantes em que ocorreram as chegadas, para de seguida serem utilizados no programa B.3 de forma a obtermos assim os gráficos da estimação da densidade do núcleo. Na Figura 2.2 temos a estimação da densidade correspondente ao histograma A.1.

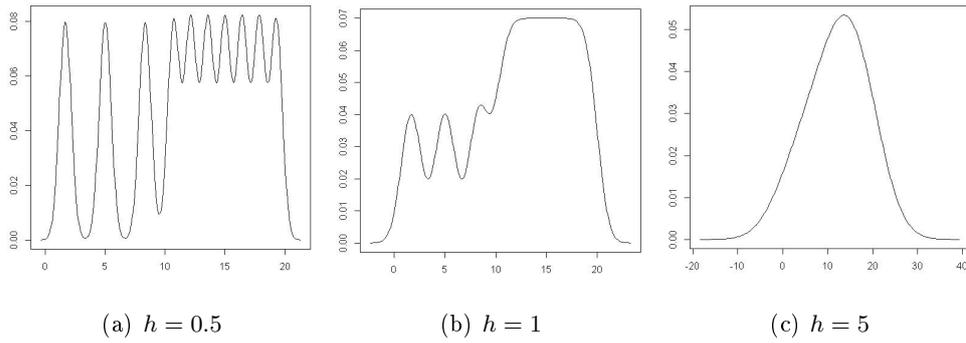


Figura 2.2: Estimador do Núcleo com os dados correspondentes ao histograma A.1 para (a) $h = 0.5$, (b) $h = 1$ e (c) $h = 5$.

Assim, relativamente à Figura 2.2, observa-se de novo que o h tem bastante influência nos resultados obtidos. Para h pequeno ($h = 0.05$) observamos muitas oscilações. Para um h não tão pequeno ($h = 1$) observamos poucas oscilações e observamos também pareças com o histograma correspondente, enquanto que para h grande ($h = 5$) temos o problema de obter valores negativos.

Na Figura 2.3 temos a estimação da densidade através do mesmo método, mas agora para o correspondente histograma A.2,

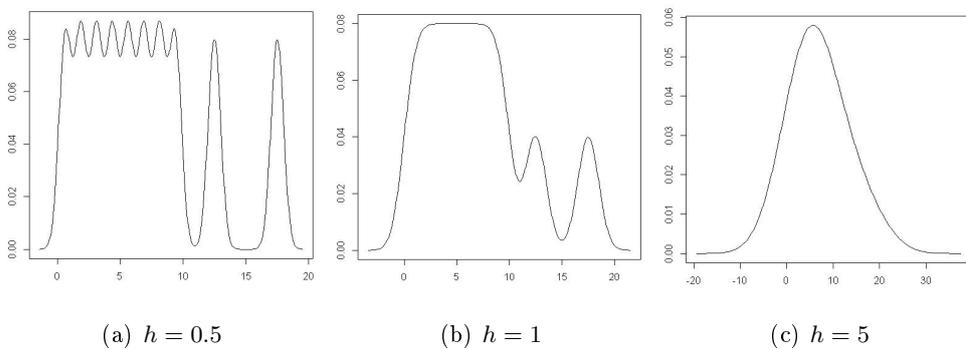


Figura 2.3: Estimador do Núcleo com os dados correspondentes ao histograma A.2 para (a) $h = 0.5$, (b) $h = 1$ e (c) $h = 5$.

As conclusões que se podem tirar desta Figura são exactamente as mesmas referidas para a Figura 2.2.

Fazendo a estimação da densidade usando conjuntamente estes grupos de dados, obtivemos a Figura 2.4.

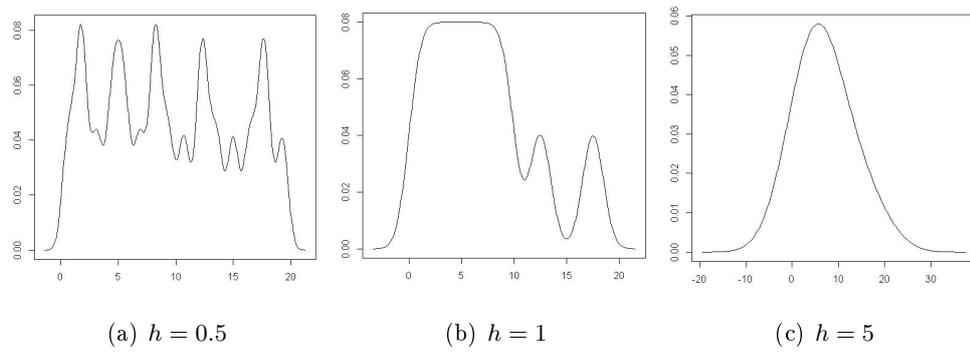


Figura 2.4: Estimador do Núcleo com os dois grupos de dados para (a) $h = 0.5$, (b) $h = 1$ e (c) $h = 5$.

Capítulo 3

Desenvolvimento

Neste capítulo definimos a função spline cúbico, enunciamos o problema de encontrar um spline cúbico com área fixa e a nossa formulação para o problema abordado.

Este capítulo está estruturado do seguinte modo, na secção 3.1 começamos por definir as funções spline quadrático e spline cúbico, $H(t)$, para de seguida caracterizármos a função $H(t)$ com o sistema linear $B'M = C'H$ e definirmos as condições iniciais em x_0 e x_n . Nesta mesma secção enunciamos e demonstramos que o spline quadrático interpolador com área fixa reduz-se ao problema de encontrar um determinado spline cúbico interpolador (sem restrição de área). Considerámos também o problema de encontrar um spline cúbico com área previamente fixada em cada subintervalo para diferentes condições iniciais.

Na secção 3.2 fazemos a descrição em notação matricial do problema original como um problema cuja a função objectivo consite numa função quadrática, com restrições de área unitária e de splines cúbicos. Aplicámos a nossa formulação a um pequeno exemplo para diferentes de condições iniciais.

3.1. Spline quadráticos e cúbicos

Consideremos uma partição de um intervalo $[a, b)$ em n subintervalos,

$$[x_0, x_1), \quad [x_1, x_2), \quad \dots, \quad [x_{n-1}, x_n),$$

onde $a \equiv x_0 < x_1 < \dots < x_n \equiv b$. As componentes de $x = (x_0, x_1, \dots, x_n)$ serão referidas por *nós*; as componentes x_1, x_2, \dots, x_{n-1} serão referidas como *nós internos*.

Denote-se por $P_{k,x}[a, b)$ o conjunto de todas as funções reais cujo domínio é o intervalo $[a, b)$ e cuja restrição a cada um dos subintervalos $[x_{i-1}, x_i)$ é um polinómio de ordem k , *i.e.*, um polinómio de grau inferior a k . Portanto, $H[s] \in P_{k,x}[a, b)$ se,

para cada $t \in [a, b)$,

$$H[s](t) = \begin{cases} s_0(t), & \text{se } t \in [x_0, x_1), \\ s_1(t), & \text{se } t \in [x_1, x_2), \\ \dots & \\ s_{n-1}(t), & \text{se } t \in [x_{n-1}, x_n), \end{cases} \quad (3.1)$$

sendo cada s_i um polinómio de ordem k . Portanto, a função $H[s]$ fica completamente caracterizada com o conhecimento dos k coeficientes de cada um daqueles n polinómios. Como estes coeficientes podem assumir qualquer valor real, o conjunto $P_{k,x}[a, b)$ é um espaço vectorial real de dimensão kn (pois existe uma correspondência um para um entre $P_{k,x}[a, b)$ e \mathbb{R}^{kn}).

Uma função $H[s] \in P_{k,x}[a, b)$ dir-se-á um *spline de ordem k* se possuir derivada contínua em (a, b) até à ordem $k - 2$, inclusivé. Em particular, um spline dir-se-á *parabólico* se $k = 3$ e dir-se-á *cúbico* se $k = 4$. Portanto, um spline é um elemento de $P_{k,x}[a, b)$ que satisfaz um conjunto de condições nos coeficientes dos polinómios, conforme Teoremas 3.1.1 e 3.1.2 enunciados abaixo para splines quadráticos e cúbicos, respectivamente.

O Teorema 3.1.1 caracteriza os splines quadráticos no conjunto $P_{3,x}[a, b)$ através de um sistema de n equações nas $2(n + 1)$ variáveis

$$\{(H_0, m_0), (H_1, m_1), \dots, (H_n, m_n)\}, \quad (3.2)$$

onde

$$\begin{aligned} H_{i-1} &\equiv s_{i-1}(x_{i-1}), & i &= 1, 2, \dots, n, \\ m_{i-1} &\equiv s'_{i-1}(x_{i-1}), & i &= 1, 2, \dots, n, \end{aligned}$$

denotam os valores de H e sua derivada nos nós x_0, x_1, \dots, x_{n-1} ; e

$$H_n \equiv s_{n-1}(x_n), \quad m_n \equiv s'_{n-1}(x_n),$$

denotam os valores limite de H e sua derivada em x_n . Seja ainda $h \in \mathbb{R}^n$ um vector-coluna cujas componentes são as amplitudes do intervalo entre os nós, nomeadamente, $h_i \equiv x_i - x_{i-1}$, $i = 1, 2, \dots, n$.

Teorema 3.1.1. *Seja $H[s] \in P_{3,x}[a, b)$, conforme (3.1). Então, $H[s]$ é um spline quadrático se e só se, para cada $i = 1, 2, \dots, n$,*

$$s_{i-1}(t) = \frac{m_i - m_{i-1}}{2h_i}(t - x_{i-1})^2 + m_{i-1}(t - x_{i-1}) + H_{i-1}, \quad (3.4)$$

com os escalares (3.2) a satisfazer

$$\frac{h_i}{2}(m_i + m_{i-1}) = H_i - H_{i-1}, \quad i = 1, 2, \dots, n, \quad (3.5)$$

Demonstração. Como a primeira derivada de cada s_i é uma função linear então, a continuidade da primeira derivada de H no intervalo (a, b) impõe que

$$\begin{aligned} s'_{i-1}(t) &= m_{i-1} \frac{x_i - t}{h_i} + m_i \frac{t - x_{i-1}}{h_i} \\ &= \frac{m_i - m_{i-1}}{h_i}(t - x_{i-1}) + m_{i-1}, \quad i = 1, 2, \dots, n, \end{aligned}$$

pelo que

$$s_{i-1}(t) = \frac{m_i - m_{i-1}}{2h_i}(t - x_{i-1})^2 + m_{i-1}(t - x_{i-1}) + H_{i-1}, \quad i = 1, 2, \dots, n.$$

A continuidade de H no intervalo (a, b) impõe que nos nós internos se verifique

$$\begin{aligned} \underbrace{\frac{m_i - m_{i-1}}{2h_i}h_i^2 + m_{i-1}h_i + H_{i-1}}_{s_{i-1}(x_i)} &= \underbrace{H_i}_{s_i(x_i)} \\ \iff \frac{h_i}{2}(m_i + m_{i-1}) &= H_i - H_{i-1} \quad i = 1, 2, \dots, n-1, \end{aligned} \quad (3.6)$$

Por outro lado, $s_{n-1}(x_n) = H_n$ pelo que

$$\begin{aligned} \frac{m_n - m_{n-1}}{2h_n}h_n^2 + m_{n-1}h_n + H_{n-1} &= H_n, \\ \iff \frac{h_n}{2}(m_n + m_{n-1}) &= H_n - H_{n-1}. \end{aligned} \quad (3.7)$$

As equações (3.6) e (3.7) constituem (3.5).

O recíproco é simples de verificar e, por isso, essa verificação é omitida neste texto. De facto, qualquer conjunto de valores (3.2) que satisfaça (3.5) faz com que $H[s] \in P_{3,x}[a, b)$, conforme (3.1) e (3.4), seja um spline quadrático. Isto conclui a demonstração. \square

Atendendo à sua forma particular, o sistema de equações (3.5) é equivalente ao sistema

$$m_i = (-1)^i m_0 + 2 \sum_{j=1}^i (-1)^{i-j} \frac{H_j - H_{j-1}}{h_j}, \quad i = 1, \dots, n, \quad (3.8)$$

pelo que aquelas variáveis m_i podem ser substituídas directamente na expressão (3.4), ficando m_0 livre de assumir qualquer valor. Condições iniciais que determinam unicidade do spline quadrático interpolador de $\{(x_i, H_i) : i = 0, 1, \dots, n\}$ incluem, por exemplo: (i) fixar m_0 a determinado valor - que fixa o valor limite do declive das

rectas tangente a H em x_0 ; (ii) fixar $(m_1 - m_0)/h_0$ a determinado valor - que fixa a concavidade de H em (x_0, x_1) ; (iii) fixar $m_n = m_0$ - que impõe igualdade no declive das rectas tangente a H em x_0 e x_n , condição mais adequada a sistemas cíclicos.

O Teorema 3.1.2 abaixo caracteriza os splines cúbicos no conjunto $P_{4,x}[a, b]$ através de um sistema de $n - 1$ equações nas $2(n + 1)$ variáveis

$$\{(H_0, M_0), (H_1, M_1), \dots, (H_n, M_n)\}, \quad (3.9)$$

onde

$$\begin{aligned} H_{i-1} &\equiv s_{i-1}(x_{i-1}), & i = 1, 2, \dots, n-1, \\ M_{i-1} &\equiv s''_{i-1}(x_{i-1}), & i = 1, 2, \dots, n-1, \end{aligned}$$

denotam os valores de H e a sua segunda derivada nos nós x_0, x_1, \dots, x_{n-1} ; e

$$H_n \equiv s_{n-1}(x_n), \quad M_n \equiv s''_{n-1}(x_n),$$

denotam os valores limite de H e sua segunda derivada em x_n .

Teorema 3.1.2. *Seja $H[s] \in P_{4,x}[a, b]$, conforme (3.1). Então, $H[s]$ é um spline cúbico se e só se, para cada $i = 1, 2, \dots, n$,*

$$s_{i-1}(t) = A_i(t - x_{i-1})^3 + B_i(t - x_{i-1})^2 + C_i(t - x_{i-1}) + D_i, \quad (3.11)$$

cujos coeficientes são definidos por

$$\begin{aligned} A_i &= \frac{M_i - M_{i-1}}{6h_i}, & B_i &= \frac{M_{i-1}}{2}, \\ C_i &= \frac{H_i - H_{i-1}}{h_i} - \frac{h_i}{3} \left(\frac{M_i}{2} + M_{i-1} \right), & D_i &= H_{i-1}, \end{aligned}$$

e com os escalares (3.9) a satisfazer

$$h_i M_{i-1} + 2(h_i + h_{i+1})M_i + h_{i+1}M_{i+1} = \frac{6}{h_i} H_{i-1} - \frac{6(h_i + h_{i+1})}{h_i h_{i+1}} H_i + \frac{6}{h_{i+1}} H_{i+1}, \quad i = 1, 2, \dots, n-1 \quad (3.12)$$

Demonstração. Como a segunda derivada de cada s_i é uma função linear então, a continuidade da segunda derivada em H no intervalo (a, b) impõe que

$$\begin{aligned} s''_{i-1}(t) &= M_{i-1} \frac{x_i - t}{h_i} + M_i \frac{t - x_{i-1}}{h_i} \\ &= \frac{M_i - M_{i-1}}{h_i} (t - x_{i-1}) + M_{i-1}, & i = 1, 2, \dots, n, \end{aligned} \quad (3.13)$$

pele que, para $i = 1, 2, \dots, n$,

$$\begin{aligned} s'_{i-1}(t) &= \frac{M_i - M_{i-1}}{2h_i} (t - x_{i-1})^2 + M_{i-1}(t - x_{i-1}) + m_{i-1}, \\ s_{i-1}(t) &= A_i(t - x_{i-1})^3 + B_i(t - x_{i-1})^2 + m_{i-1}(t - x_{i-1}) + D_i. \end{aligned} \quad (3.14)$$

A continuidade de H no intervalo (a, b) impõe que nos nós internos se verifique, para cada $i = 1, 2, \dots, n-1$,

$$\begin{aligned} & \underbrace{\frac{M_i - M_{i-1}}{6h_i}h_i^3 + \frac{M_{i-1}}{2}h_i^2 + m_{i-1}h_i + H_{i-1}}_{s_{i-1}(x_i)} = \underbrace{H_i}_{s_i(x_i)} \\ \Leftrightarrow & m_{i-1} = \frac{H_i - H_{i-1}}{h_i} - \frac{h_i}{3} \left(\frac{M_i}{2} + M_{i-1} \right). \\ \Leftrightarrow & m_{i-1} = C_i. \end{aligned} \tag{3.15}$$

Por outro lado, $s_{n-1}(x_n) = H_n$ pelo que

$$\begin{aligned} & \frac{M_n - M_{n-1}}{6h_n}h_n^3 + \frac{M_{n-1}}{2}h_n^2 + m_{n-1}h_n + H_{n-1} = H_n \\ \Leftrightarrow & m_{n-1} = \frac{H_n - H_{n-1}}{h_n} - \frac{h_n}{3} \left(\frac{M_n}{2} + M_{n-1} \right). \\ \Leftrightarrow & m_{n-1} = C_n. \end{aligned} \tag{3.16}$$

Concluimos (3.11). A continuidade da derivada de H no intervalo (a, b) impõe que nos nós internos se verifique, para $i = 1, 2, \dots, n-1$

$$\begin{aligned} & \underbrace{\frac{M_i - M_{i-1}}{2h_i}h_i^2 + M_{i-1}h_i + C_i}_{s'_{i-1}(x_i)} = \underbrace{C_{i+1}}_{s'_i(x_i)} \\ \Leftrightarrow & \frac{M_i + M_{i-1}}{2}h_i + \frac{H_i - H_{i-1}}{h_i} - \frac{h_i}{3} \left(\frac{M_i}{2} + M_{i-1} \right) = \frac{H_{i+1} - H_i}{h_{i+1}} - \frac{h_{i+1}}{3} \left(\frac{M_{i+1}}{2} + M_i \right), \end{aligned}$$

que é equivalente a (3.12).

O recíproco é simples de verificar e, por isso, essa verificação é omitida neste texto. De facto, qualquer conjunto de valores (3.9) que satisfaça (3.12) faz com que $H[s] \in P_{4,x}[a, b)$, conforme (3.1) e (3.11), seja um spline cúbico. Isto conclui a demonstração. \square

Corolário 3.1.3. *Nas condições do Teorema 3.1.2,*

$$\begin{aligned} m_{i-1} &= C_i, \quad i = 1, 2, \dots, n, \\ m_n &= \frac{H_n - H_{n-1}}{h_n} + \frac{h_n}{3} \left(\frac{M_{n-1}}{2} + M_n \right). \end{aligned}$$

Demonstração. Pela demonstração do Teorema 3.1.2 temos,

$$m_{i-1} = C_i$$

Provemos agora

$$m_n = \frac{H_n - H_{n-1}}{h_n} + \frac{h_n}{3} \left(\frac{M_{n-1}}{2} + M_n \right)$$

$$\begin{aligned}
 m_n &= s'_{n-1}(x_n) \\
 &= 3A_n(x_n - x_{n-1})^2 + 2B_n(x_n - x_{n-1}) + C_n \\
 &= 3\frac{M_n - M_{n-1}}{6h_n}h_n^2 + 2\frac{M_{n-1}}{2}h_n + \frac{H_n - H_{n-1}}{h_n} - \frac{h_n}{3}\left(\frac{M_n}{2} + M_{n-1}\right) \\
 &= \frac{M_n - M_{n-1}}{2}h_n + M_{n-1}h_n + \frac{H_n - H_{n-1}}{h_n} - \frac{h_n}{3}\left(\frac{M_n}{2} + M_{n-1}\right) \\
 &= \frac{H_n - H_{n-1}}{h_n} + h_n\left(\frac{M_n}{2} - \frac{M_{n-1}}{2} + M_{n-1} - \frac{M_n}{6} - \frac{M_{n-1}}{3}\right) \\
 &= \frac{H_n - H_{n-1}}{h_n} + h_n\left(\frac{3M_n - M_n}{6} - \frac{3M_{n-1} - 6M_{n-1} + 2M_{n-1}}{6}\right) \\
 &= \frac{H_n - H_{n-1}}{h_n} + h_n\left(\frac{M_n}{3} + \frac{M_{n-1}}{6}\right)
 \end{aligned}$$

Como queremos provar.

□

Se os valores H fossem previamente conhecidos, como acontece no caso em que o problema é o de interpolar um conjunto de pontos dados

$$\{(x_0, H_0), (x_1, H_1), \dots, (x_n, H_n)\} \quad (3.17)$$

então (3.12) define um sistema de equações lineares $(n - 1) \times (n + 1)$ nas variáveis M . A matriz do sistema é

$$B' = \begin{bmatrix} h_1 & 2(h_1 + h_2) & h_2 & 0 & \dots & 0 & 0 & 0 \\ 0 & h_2 & 2(h_2 + h_3) & h_3 & \dots & 0 & 0 & 0 \\ \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 2(h_{n-2} + h_{n-1}) & h_{n-1} & 0 \\ 0 & 0 & 0 & 0 & \dots & h_{n-1} & 2(h_{n-1} + h_n) & h_n \end{bmatrix}, \quad (3.18)$$

e o termo independente é $C'H$, para

$$C' = 6 \begin{bmatrix} \frac{1}{h_1} & -\left(\frac{1}{h_1} + \frac{1}{h_2}\right) & \frac{1}{h_2} & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{1}{h_2} & -\left(\frac{1}{h_2} + \frac{1}{h_3}\right) & \frac{1}{h_3} & \dots & 0 & 0 & 0 \\ \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & -\left(\frac{1}{h_{n-2}} + \frac{1}{h_{n-1}}\right) & \frac{1}{h_{n-1}} & 0 \\ 0 & 0 & 0 & 0 & \dots & \frac{1}{h_{n-1}} & -\left(\frac{1}{h_{n-1}} + \frac{1}{h_n}\right) & \frac{1}{h_n} \end{bmatrix} \quad (3.19)$$

A característica da matriz do sistema é $n - 1$ porque essa matriz se pode estender trivialmente a uma matriz simétrica e positiva definida (ou estritamente diagonalmente dominante). Devemos impôr duas condições iniciais para obter a unicidade do spline cúbico interpolador. Assim, obtemos o spline cúbico interpolador *natural* se impusermos

$$M_0 = 0, \quad (3.20a)$$

$$M_n = 0, \quad (3.20b)$$

o que, em termos “físicos” é adequado a um sistema que “arranca” e “termina” a uma velocidade constante; obtemos o spline cúbico interpolador *standard* se impo- sermos que um eventual $H(x_0 - h_1)$ coincida com $H(x_0 + h_1)$ bem como um eventual $H(x_n + h_n)$ coincida com $H(x_n - h_n)$. Estas condições iniciais, que são adequa- das a um sistema que já está em funcionamento no instante inicial x_0 , não termina abruptamente no instante final x_n , em torno dos quais o funcionamento é *refletido*¹, correspondem às seguintes equações, veja-se [1] por exemplo,

$$4h_1M_0 + 2h_1M_1 = \alpha_1H_0 + \beta_1H_1 + \gamma_1H_2, \quad (3.21a)$$

$$2h_nM_{n-1} + 4h_nM_n = \alpha_{n-1}H_{n-2} + \beta_{n-1}H_{n-1} + \gamma_{n-1}H_n. \quad (3.21b)$$

Finalmente, obtemos o spline cúbico interpolador *cíclico* se considerarmos que os nós extremos, x_0 e x_n , identificam o mesmo ponto². As correspondente equações são

$$-\frac{h_1}{3}M_0 - \frac{h_1}{6}M_1 - \frac{h_n}{6}M_{n-1} - \frac{h_n}{3}M_n = \frac{1}{h_1}H_0 - \frac{1}{h_1}H_1 - \frac{1}{h_n}H_{n-1} + \frac{1}{h_n}H_n \quad (3.22a)$$

$$M_0 - M_n = 0, \quad (3.22b)$$

e são adequadas a sistemas que funcionem ininterruptamente.

Agora, consideremos o problema de encontrar um spline quadrático com área previamente fixada em cada subintervalo. Mais especificamente, pretende-se um spline $H[s] \in P_{3,x}[a, b)$ tal que

$$\int_{x_{i-1}}^{x_i} s_{i-1}(t) dt = f_i, \quad \text{para } i = 1, \dots, n, \quad (3.23)$$

sendo f_1, f_2, \dots, f_n escalares reais não negativos. O próximo teorema mostra que um spline quadrático com esta restrição pode ser deduzido a partir de um adequado spline cúbico interpolador.

Teorema 3.1.4. *Se $H[S] \in P_{4,x}[a, b)$ é um spline cúbico interpolador tal que*

$$H_i = H_0 + \sum_{j=1}^i f_j, \quad i = 1, 2, \dots, n.$$

então, para $s = S'$, a função $H[s] \in P_{3,x}[a, b)$ é um spline quadrático satisfazendo (3.23).

Demonstração. Pelo Teorema 3.1.2, $H[S]$ é tal que, para cada $i = 1, 2, \dots, n$,

$$S_{i-1}(t) = A_i(t - x_{i-1})^3 + B_i(t - x_{i-1})^2 + C_i(t - x_{i-1}) + D_i,$$

¹portanto, os nós x_0 e x_n funcionam como nós internos.

²Portanto, admite-se que $H_0 = H_n$.

cujos coeficientes estão definidos em (3.9), e satisfazendo (3.12). Como, para cada $i = 1, 2, \dots, n-1$,

$$\begin{aligned} & \frac{6}{h_i} H_{i-1} - \frac{6(h_i + h_{i+1})}{h_i h_{i+1}} H_i + \frac{6}{h_{i+1}} H_{i+1} = \\ &= \frac{6}{h_i} \left(H_0 + \sum_{j=1}^{i-1} f_j \right) - \frac{6(h_i + h_{i+1})}{h_i h_{i+1}} \left(H_0 + \sum_{j=1}^i f_j \right) + \frac{6}{h_{i+1}} \left(H_0 + \sum_{j=1}^{i+1} f_j \right) \\ &= \left(\frac{6}{h_i} - \frac{6(h_i + h_{i+1})}{h_i h_{i+1}} + \frac{6}{h_{i+1}} \right) \left(H_0 + \sum_{j=1}^{i-1} f_j \right) - \frac{6(h_i + h_{i+1})}{h_i h_{i+1}} (f_i) + \frac{6}{h_{i+1}} (f_i + f_{i+1}) \\ &= \frac{6f_{i+1}}{h_{i+1}} - \frac{6f_i}{h_i} \end{aligned}$$

concluimos que, cada equação em (3.12) reduz-se a

$$h_i M_{i-1} + 2(h_i + h_{i+1}) M_i + h_{i+1} M_{i+1} = \frac{6f_{i+1}}{h_{i+1}} - \frac{6f_i}{h_i}.$$

Portanto, para cada $i = 1, 2, \dots, n$,

$$S'_{i-1}(t) = \frac{M_i - M_{i-1}}{2h_i} (t - x_{i-1})^2 + M_{i-1} (t - x_{i-1}) + C_i,$$

com

$$C_i \equiv \frac{f_i}{h_i} - \frac{h_i}{3} \left(\frac{M_i}{2} + M_{i-1} \right).$$

Ora, para $i = 1, \dots, n-1$,

$$\begin{aligned} C_{i+1} - C_i &= \\ &= \frac{f_{i+1}}{h_{i+1}} - \frac{h_{i+1}}{3} \left(\frac{M_{i+1}}{2} + M_i \right) - \frac{f_i}{h_i} + \frac{h_i}{3} \left(\frac{M_i}{2} + M_{i-1} \right) \\ &= \left(\frac{h_i}{6} M_{i-1} + \frac{h_i + h_{i+1}}{3} M_i + \frac{h_{i+1}}{6} M_{i+1} \right) - \frac{h_{i+1}}{3} \left(\frac{M_{i+1}}{2} + M_i \right) + \frac{h_i}{3} \left(\frac{M_i}{2} + M_{i-1} \right) \\ &= \frac{h_i}{2} (M_{i-1} + M_i). \end{aligned}$$

Além disso, podemos definir C_{n+1} de tal modo

$$C_{n+1} - C_n = \frac{h_n}{2} (M_{n-1} + M_n).$$

Portanto, pelo Teorema 3.1.1, concluimos que $H[S']$ é um spline quadrático. Além disso,

$$\begin{aligned} & \int_{x_{i-1}}^{x_i} S'_{i-1}(t) dt = S_{i-1}(x_i) - S_{i-1}(x_{i-1}) = \\ &= A_i h_i^3 + B_i h_i^2 + C_i h_i + D_i - D_i \\ &= \frac{h_i^2}{6} (M_i - M_{i-1}) + \frac{h_i^2}{2} M_{i-1} + (H_i - H_{i-1}) - \frac{h_i^2}{3} \left(\frac{M_i}{2} + M_{i-1} \right) \\ &= H_i - H_{i-1} \\ &= f_i, \end{aligned}$$

para cada $i = 1, 2, \dots, n$, pelo que concluimos o resultado desejado. \square

Este teorema mostra que encontrar um spline quadrático interpolador com área fixa reduz-se ao problema de encontrar um determinado spline cúbico interpolador (sem restrição de área).

Agora, consideremos o problema de encontrar um spline cúbico com área previamente fixada em cada subintervalo. Mais especificamente, pretende-se um spline $H[s] \in P_{4,x}[a, b)$ que verifique (3.23). Como

$$\begin{aligned} \int_{x_{i-1}}^{x_i} s_{i-1}(t) dt &= \\ &= \int_0^{h_i} A_i y^3 + B_i y^2 + C_i y + D_i dy = \\ &= A_i \frac{h_i^4}{4} + B_i \frac{h_i^3}{3} + C_i \frac{h_i^2}{2} + D_i h_i \\ &= \left(\frac{M_i - M_{i-1}}{6h_i} \right) \frac{h_i^4}{4} + \left(\frac{M_{i-1}}{2} \right) \frac{h_i^3}{3} + \\ &\quad + \left(\frac{H_i - H_{i-1}}{h_i} - \frac{h_i}{3} \left(\frac{M_i}{2} + M_{i-1} \right) \right) \frac{h_i^2}{2} + H_{i-1} h_i \\ &= -\frac{h_i^3}{24} (M_{i-1} + M_i) + \frac{h_i}{2} (H_{i-1} + H_i) \end{aligned}$$

então, ao sistema (3.12) teremos de acrescentar as equações,

$$-\frac{h_i^3}{24} (M_{i-1} + M_i) + \frac{h_i}{2} (H_{i-1} + H_i) = f_i, \quad i = 1, 2, \dots, n. \quad (3.24)$$

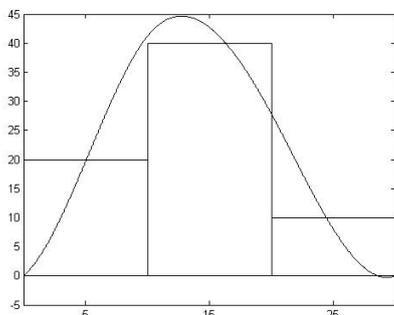
definindo um sistema com $2n - 1$ equações e $2n + 2$ variáveis. Condições iniciais para este problema incluem, por exemplo,

$$H_0 = 0 \quad (3.25a) \quad H_0 = H_n \quad (3.26a)$$

$$H_n = 0 \quad (3.25b) \quad m_0 = m_n \quad (3.26b)$$

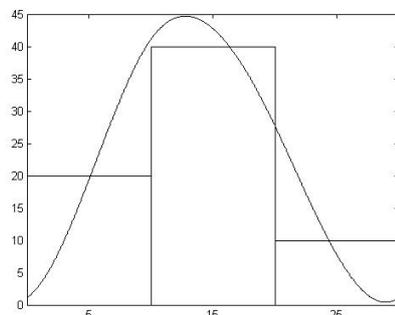
$$M_0 = M_n \quad (3.25c) \quad M_0 = M_n \quad (3.26c)$$

Nas Figuras 3.1 e 3.2 exibimos exemplos simples de aplicação destas condições iniciais. Os resultados foram obtidos utilizando os programas da secção B.6 e B.7, respetivamente, do apêndice B.



(a) condições iniciais:

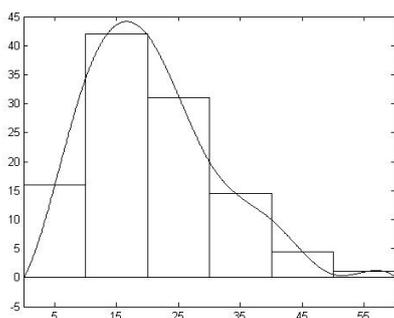
$$H_0 = 0, H_n = 0, M_0 = M_n$$



(b) condições iniciais:

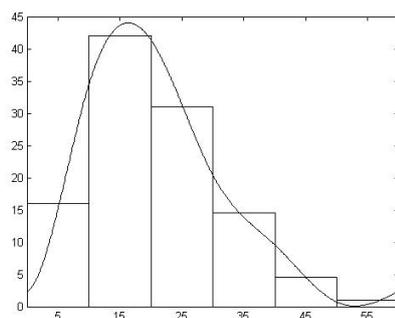
$$H_0 = H_n, m_0 = m_n, M_0 = 0 = M_n$$

Figura 3.1: Spline cúbico para os nós em $\{0, 10, 20, 30\}$ e áreas em $\{200, 400, 100\}$



(a) condições iniciais:

$$H_0 = 0, H_n = 0, M_0 = M_n$$



(b) condições iniciais:

$$H_0 = H_n, m_0 = m_n, M_0 = 0 = M_n$$

Figura 3.2: Spline cúbico para os nós $\{0, 10, 20, 30, 40, 50, 60\}$ e áreas em $\{160, 420, 310, 145, 45, 10\}$

3.2. A nossa formulação

De acordo com a nossa abordagem, procuraremos um spline cúbico $H[s]$ que, para além das restrições inerentes às condições iniciais e à restrição de área

$$\int_{x_0}^{x_n} H(t) dt = 1 \quad (3.27)$$

torne mínimo o valor de

$$\sum_{i=1}^n \sum_{j=1}^m \int_{x_{i-1}}^{x_i} \left(H(t) - \frac{f_i^j}{h_i} \right)^2 dt. \quad (3.28)$$

Seguidamente, explicamos que se trata de um problema de optimização quadrática convexa com restrições lineares de igualdade que se presta a várias simplificações.

3.2.1. A função objectivo

Primeiro, observe-se que cada uma das parcelas no duplo somatório em (3.28) é igual a

$$\begin{aligned} & \int_{x_{i-1}}^{x_i} \left(s_{i-1}(t) - \frac{f_i^j}{h_i} \right)^2 dt = \\ & = \int_{x_{i-1}}^{x_i} s_{i-1}^2(t) dt - 2 \frac{f_i^j}{h_i} \int_{x_{i-1}}^{x_i} s_{i-1}(t) dt + \left(\frac{f_i^j}{h_i} \right)^2 h_i. \end{aligned}$$

Por (3.24),

$$\int_{x_{i-1}}^{x_i} s_{i-1}(t) dt = \begin{bmatrix} -\frac{h_i^3}{24} \\ -\frac{h_i^3}{24} \\ \frac{h_i}{2} \\ \frac{h_i}{2} \end{bmatrix}^T \begin{bmatrix} M_{i-1} \\ M_i \\ H_{i-1} \\ H_i \end{bmatrix} = \begin{bmatrix} b_M^i \\ b_H^i \end{bmatrix}^T \begin{bmatrix} M \\ H \end{bmatrix}, \quad (3.29)$$

onde b^i denota uma matriz-coluna com zeros nas componentes correspondentes às variáveis distintas de $M_{i-1}, M_i, H_{i-1}, H_i$.

Lema 3.2.1. *Para cada $i = 1, 2, \dots, n$,*

$$\int_{x_{i-1}}^{x_i} s_{i-1}^2(t) dt = \begin{bmatrix} M_{i-1} \\ M_i \\ H_{i-1} \\ H_i \end{bmatrix}^T \underbrace{\begin{bmatrix} \frac{2h_i^5}{945} & \frac{31h_i^5}{15120} & -\frac{h_i^3}{45} & -\frac{7h_i^3}{360} \\ \frac{31h_i^5}{15120} & \frac{2h_i^5}{945} & -\frac{7h_i^3}{360} & -\frac{h_i^3}{45} \\ -\frac{h_i^3}{45} & -\frac{7h_i^3}{360} & \frac{h_i}{3} & \frac{h_i}{6} \\ -\frac{7h_i^3}{360} & -\frac{h_i^3}{45} & \frac{h_i}{6} & \frac{h_i}{3} \end{bmatrix}}_{S_i} \begin{bmatrix} M_{i-1} \\ M_i \\ H_{i-1} \\ H_i \end{bmatrix},$$

Demonstração. Pelo Teorema Fundamental do Cálculo,

$$\begin{aligned}
 & \int_{x_{i-1}}^{x_i} s_{i-1}^2(t) dt = \\
 &= \int_0^{h_i} (A_i y^3 + B_i y^2 + C_i y + D_i)^2 dy \\
 &= D_i^2 h_i + C_i D_i h_i^2 + \frac{(C_i^2 + 2B_i D_i)}{3} h_i^3 + \frac{(B_i C_i + A_i D_i)}{2} h_i^4 + \frac{(B_i^2 + 2A_i C_i)}{5} h_i^5 \\
 &\quad + \frac{(A_i B_i)}{3} h_i^6 + \frac{(A_i^2)}{7} h_i^7
 \end{aligned}$$

que algumas simplificações depois conduzem ao resultado desejado. \square

Pelo Lema 3.2.1,

$$\int_{x_{i-1}}^{x_i} s_{i-1}^2(t) dt = \begin{bmatrix} M \\ H \end{bmatrix}^T \begin{bmatrix} A_{MM}^i & A_{MH}^i \\ A_{HM}^i & A_{HH}^i \end{bmatrix} \begin{bmatrix} M \\ H \end{bmatrix} \quad (3.30)$$

onde A^i denota uma matriz com zeros nas linhas e colunas correspondentes às variáveis distintas de $M_{i-1}, M_i, H_{i-1}, H_i$.

Lema 3.2.2. *Para cada $i = 1, 2, \dots, n$, a matriz S_i é simétrica e positiva definida.*

Demonstração. Os valores próprios da matriz S_i , obtidos com recurso ao software Mathematica, são

$$\frac{h \left(120 + h^4 \pm \sqrt{120^2 + 160h^4 + h^8} \right)}{480}, \quad \frac{h \left(2520 + h^4 \pm \sqrt{2520^2 + 2016h^4 + h^8} \right)}{30240}.$$

(onde $h \equiv h_i$) e são todos positivos para todo $h > 0$ pois

$$(120 + h^4)^2 = 120^2 + 240h^4 + h^8 > 120^2 + 160h^4 + h^8,$$

$$(2520 + h^4)^2 = 2520^2 + 5040h^4 + h^8 > 2520^2 + 2016h^4 + h^8.$$

\square

Pelo Lema 3.2.2, toda a matriz A^i é simétrica e positiva semi-definida pois, a menos de uma permutação principal de linhas e colunas,

$$A^i = \begin{bmatrix} S_i & 0 \\ 0 & 0 \end{bmatrix}.$$

Isso implica que (3.30) é uma função estritamente convexa nas variáveis $(M_{i-1}, M_i, H_{i-1}, H_i)$ e convexa na totalidade das variáveis (M, H) .

Em conclusão, a expressão em (3.28) é uma função quadrática convexa nas variáveis (M, H) com a seguinte expressão

$$\frac{1}{2} \begin{bmatrix} M \\ H \end{bmatrix}^T \begin{bmatrix} A_{MM} & A_{MH} \\ A_{HM} & A_{HH} \end{bmatrix} \begin{bmatrix} M \\ H \end{bmatrix} + \begin{bmatrix} b_M \\ b_H \end{bmatrix}^T \begin{bmatrix} M \\ H \end{bmatrix} + c \quad (3.31)$$

onde

$$\begin{bmatrix} A_{MM} & A_{MH} \\ A_{HM} & A_{HH} \end{bmatrix} = 2m \sum_{i=1}^n \begin{bmatrix} A_{MM}^i & A_{MH}^i \\ A_{HM}^i & A_{HH}^i \end{bmatrix},$$

$$\begin{bmatrix} b_M \\ b_H \end{bmatrix} = -2 \sum_{i=1}^n \frac{\sum_{j=1}^m f_i^j}{h_i} \begin{bmatrix} b_M^i \\ b_H^i \end{bmatrix},$$

$$c = \sum_{i=1}^n \frac{\sum_{j=1}^m f_{ij}^2}{h_i}$$

Note-se que A é uma matriz simétrica e Positiva Semi-Definida porque é a soma de matrizes simétricas e positiva semi-definidas.

Ilustramos estas matrizes com um pequeno exemplo. Para $n = 2$, $m = 2$, $h \equiv h_1 = h_2 = 10$ considere

$$f = \begin{bmatrix} 0.3 & 0.8 \\ 0.7 & 0.2 \end{bmatrix} \quad (3.32)$$

que corresponde aos dois histogramas das Figuras A.1 e A.2 do apêndice A,

$$A = \begin{bmatrix} \frac{4 \times 2 \times 10^5}{945} & \frac{4 \times 31 \times 10^5}{15120} & 0 & -\frac{4 \times 10^3}{45} & -\frac{4 \times 7 \times 10^3}{360} & 0 \\ \frac{4 \times 31 \times 10^5}{15120} & \frac{4 \times 4 \times 10^5}{945} & \frac{4 \times 31 \times 10^5}{15120} & -\frac{4 \times 7 \times 10^3}{360} & -\frac{4 \times 2 \times 10^3}{45} & -\frac{4 \times 7 \times 10^3}{360} \\ 0 & \frac{4 \times 31 \times 10^5}{15120} & \frac{4 \times 2 \times 10^5}{945} & 0 & -\frac{4 \times 7 \times 10^3}{360} & -\frac{4 \times 10^3}{45} \\ -\frac{4 \times 10^3}{45} & -\frac{4 \times 7 \times 10^3}{360} & 0 & \frac{4 \times 10}{3} & \frac{4 \times 10}{6} & 0 \\ -\frac{4 \times 7 \times 10^3}{360} & -\frac{4 \times 2 \times 10^3}{45} & -\frac{4 \times 7 \times 10^3}{360} & \frac{4 \times 10}{6} & \frac{4 \times 2 \times 10}{3} & \frac{4 \times 10}{6} \\ 0 & -\frac{4 \times 7 \times 10^3}{360} & -\frac{4 \times 10^3}{45} & 0 & \frac{4 \times 10}{6} & \frac{4 \times 10}{3} \end{bmatrix}, \quad b = \begin{bmatrix} \frac{0.22 \times 10^3}{24} \\ \frac{0.4 \times 10^3}{24} \\ \frac{0.18 \times 10^3}{24} \\ -\frac{0.22 \times 10}{2} \\ -\frac{0.4 \times 10}{2} \\ -\frac{0.18 \times 10}{2} \end{bmatrix}$$

e a constante c é igual a 0.126.

3.2.2. As restrições

Antes de introduzirmos condições iniciais, as restrições que caracterizam o spline cúbico são da forma

$$B'M = C'H, \quad (3.33)$$

sendo B' definido em (3.18) e C' definido em (3.19). Para além disso, deve ser satisfeito um dos conjuntos que define as condições iniciais:

$$M_0 = 0 \quad (3.34a) \qquad H_0 = H_n \quad (3.36a)$$

$$M_n = 0 \quad (3.34b) \qquad M_0 = M_n \quad (3.36b)$$

$$M_0 = M_n \quad (3.35a) \qquad H_0 = 0 \quad (3.37a)$$

$$m_0 = m_n \quad (3.35b) \qquad H_n = 0 \quad (3.37b)$$

As condições (3.34) são referentes a um sistema natural. As (3.35) e (3.36) são para um sistema cíclico, e as (3.37) são mais apropriadas quando estamos a trabalhar com determinadas leis de probabilidade.

Após introduzirmos condições iniciais, o sistema (3.33) assume a forma

$$BM = CH. \quad (3.38)$$

No desenvolvimento que se segue assumiremos que a matriz B é invertível. Já a restrição de área (3.27) assume a forma

$$\begin{bmatrix} d_M \\ d_H \end{bmatrix}^T \begin{bmatrix} M \\ H \end{bmatrix} = 1, \quad (3.39)$$

onde

$$d = \begin{bmatrix} d_M \\ d_H \end{bmatrix} = \sum_{i=1}^n \begin{bmatrix} b_M^i \\ b_H^i \end{bmatrix}.$$

Para o exemplo (3.32) com as condições iniciais (3.34), tem-se

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 10 & 40 & 10 \\ 0 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 0 & 0 \\ \frac{6}{10} & -\left(\frac{6}{10} + \frac{6}{10}\right) & \frac{6}{10} \\ 0 & 0 & 0 \end{bmatrix}, \quad d = \begin{bmatrix} -\frac{(10)^3}{24} \\ -\frac{2 \times (10)^3}{24} \\ -\frac{(10)^3}{24} \\ \frac{10}{2} \\ \frac{2 \times 10}{2} \\ \frac{10}{2} \end{bmatrix}$$

3.2.3. O problema de optimização

O desenvolvimento nas duas subsecções anteriores permitem afirmar que para determinar o spline cúbico com área total fixa que melhor aproxima um dado conjunto de

m histogramas decorre da resolução do seguinte problema de optimização quadrática convexa

$$\left\{ \begin{array}{l} \text{minimizar} \\ \text{sujeito a} \end{array} \right. \frac{1}{2} \begin{bmatrix} M \\ H \end{bmatrix}^T \begin{bmatrix} A_{MM} & A_{MH} \\ A_{HM} & A_{MM} \end{bmatrix} \begin{bmatrix} M \\ H \end{bmatrix} + \begin{bmatrix} b_M \\ b_H \end{bmatrix}^T \begin{bmatrix} M \\ H \end{bmatrix} + c \quad (3.40)$$

$$\begin{bmatrix} B & -C \end{bmatrix} \begin{bmatrix} M \\ H \end{bmatrix} = 0$$

$$\begin{bmatrix} d_M \\ d_H \end{bmatrix}^T \begin{bmatrix} M \\ H \end{bmatrix} = 1$$

As condições de optimalidade para este problema de optimização quadrática, necessárias e suficientes por convexidade, são

$$\begin{bmatrix} A_{MM} & A_{MH} & -B^T & -d_M \\ A_{HM} & A_{HH} & C^T & -d_H \\ B & -C & 0 & 0 \\ d_M^T & d_H^T & 0 & 0 \end{bmatrix} \begin{bmatrix} M \\ H \\ \mu \\ \lambda \end{bmatrix} = \begin{bmatrix} -b_M \\ -b_H \\ 0 \\ 1 \end{bmatrix}, \quad (3.41)$$

onde $(\mu, \lambda) \in \mathbb{R}^{(n+1)+1}$ denota um vector de multiplicadores de Lagrange.

Se a matriz B é invertível então, ocorrem simplificações pois $M = (B^{-1}C)H$, pelo que o problema de optimização acima pode colocar-se apenas em termos das variáveis H ,

$$\left\{ \begin{array}{l} \text{minimizar} \\ \text{sujeito a} \end{array} \right. \begin{array}{l} \frac{1}{2} H^T \mathcal{A} H + \mathcal{B}^T H + c \\ \mathcal{D}^T H = 1, \end{array} \quad (3.42)$$

onde

$$\mathcal{A} = \begin{bmatrix} B^{-1}C \\ I \end{bmatrix}^T \begin{bmatrix} A_{MM} & A_{MH} \\ A_{HM} & A_{HH} \end{bmatrix} \begin{bmatrix} B^{-1}C \\ I \end{bmatrix},$$

$$\mathcal{B} = \begin{bmatrix} B^{-1}C \\ I \end{bmatrix}^T \begin{bmatrix} b_M \\ b_H \end{bmatrix}$$

$$\mathcal{D} = \begin{bmatrix} B^{-1}C \\ I \end{bmatrix}^T \begin{bmatrix} d_M \\ d_H \end{bmatrix}$$

As condições de optimalidade para o problema (3.42), necessárias e suficiente por convexidade, são

$$\begin{bmatrix} \mathcal{A} & -\mathcal{D} \\ \mathcal{D}^T & 0 \end{bmatrix} \begin{bmatrix} H \\ \lambda \end{bmatrix} = \begin{bmatrix} -\mathcal{B} \\ 1 \end{bmatrix}, \quad (3.43)$$

onde λ denota um multiplicador de Lagrange. Para o exemplo (3.32), tem-se

$$\begin{bmatrix} B^{-1}C \\ I \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ \frac{3}{200} & -\frac{3}{100} & \frac{3}{200} \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

$$\mathcal{A} = \begin{bmatrix} 239/21 & 39/7 & -41/21 \\ 39/7 & 272/7 & 39/7 \\ -41/21 & 39/7 & 239/21 \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} -17/20 \\ -5/2 \\ -13/20 \end{bmatrix}, \quad \mathcal{D} = \begin{bmatrix} 15/4 \\ 25/2 \\ 15/4 \end{bmatrix}$$

pelo que o sistema linear (3.43) fica

$$\begin{bmatrix} 239/21 & 39/7 & -41/21 & -15/4 \\ 39/7 & 272/7 & 39/7 & -25/2 \\ -41/21 & 39/7 & 239/21 & -15/4 \\ 15/4 & 25/2 & 15/4 & 0 \end{bmatrix} \begin{bmatrix} H_0 \\ H_1 \\ H_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} 17/20 \\ 5/2 \\ 13/20 \\ 1 \end{bmatrix}$$

cuja solução é

$$\begin{bmatrix} H_0 \\ H_1 \\ H_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} 23/400 \\ 1/20 \\ 17/400 \\ \cong 0 \end{bmatrix}$$

Portanto, e como $M = (B^{-1}C)H$ então

$$M = \begin{bmatrix} 0 \\ \cong 0 \\ 0 \end{bmatrix}$$

Vejamos agora um caso no qual a matriz B não é invertível. Considerando o mesmo exemplo mas com condição inicial (3.36) vamos ter

$$B = \begin{bmatrix} 0 & 0 & 0 \\ 10 & 40 & 10 \\ 1 & 0 & -1 \end{bmatrix}, \quad C = \begin{bmatrix} -1 & 0 & 1 \\ \frac{6}{10} & -\left(\frac{6}{10} + \frac{6}{10}\right) & \frac{6}{10} \\ 0 & 0 & 0 \end{bmatrix},$$

As restantes matrizes A, b e d mantêm-se. Como B não é invertível, teremos de

recorrer às condições de optimalidade (3.41) que possui a seguinte solução,

$$\begin{bmatrix} M_0 \\ M_1 \\ M_2 \\ H_0 \\ H_1 \\ H_2 \\ \mu_1 \\ \mu_2 \\ \mu_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} \cong 0 \\ \cong 0 \\ \cong 0 \\ 1/20 \\ 1/20 \\ 1/20 \\ -1/10 \\ \cong 0 \\ 5/6 \\ \cong 0 \end{bmatrix}$$

Concluimos que em ambos os conjuntos de condições iniciais, o spline cúbico encontrado é sensivelmente um segmento paralelo ao eixo das abcissas.

Capítulo 4

Resultados Numéricos

Neste capítulo testámos computacionalmente a nossa abordagem num conjunto de dados reais e em observações da lei exponencial geradas aleatoriamente.

Este capítulo está estruturado do seguinte modo, na secção 4.1 explicámos os dados utilizados para a nossa abordagem e na secção 4.2 descrevemos o programa implementado para os dados reais. Na secção 4.3 descrevemos os resultados preliminares obtidos que ilustram as características de cada conjunto de condições iniciais e na secção 4.4 mostramos de que modo diferentes períodos de agregação afectam a função H .

Na secção 4.5 utilizámos o estimador do núcleo para os dados reais. Esta secção está dividida do seguinte modo, na subsecção 4.5.1 descrevemos os programas implementados para as observações da lei exponencial e na subsecção 4.5.2 apresentámos os resultados obtidos para a estimação da densidade do núcleo para diferentes *janelas*.

Por último, na secção 4.6 temos os resultados obtidos para as observações da lei exponencial geradas aleatoriamente.

4.1. Dados

A implementação da abordagem descrita nesta dissertação foi testada em dois conjuntos de dados. Um primeiro conjunto de dados reais provenientes de um call center de serviços financeiros (secção B.1) e um segundo conjunto correspondente a observações de uma distribuição de probabilidade conhecida, a exponencial (secção B.8).

Os dados reais que nos foram disponibilizados correspondem ao número de chegadas (telefonemas) que chegaram a esse serviço em cada período de $1/4$ de hora ao longo de um dia completo de 24h (desde as 0:00am até às 12:00pm) em oito segundas-feiras consecutivas, consideradas normais ou típicas. Com base nestes dados construímos “histogramas” agregando os dados em períodos de $1/4$, $1/2$, 1, 2, 4 e 8 horas em cada um desses dias. A Figura 4.1 apresenta os dados que nos foram

disponibilizados em períodos de 1/4 hora, apresentados em forma de histogramas.

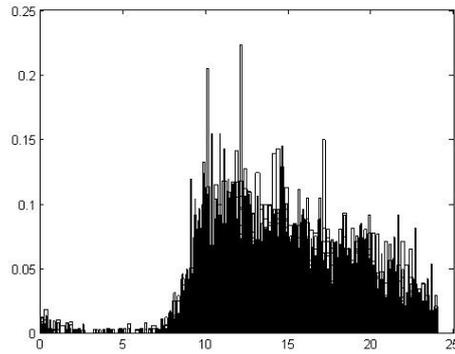


Figura 4.1: Histogramas referentes aos dados que nos foram disponibilizados em períodos de 1/4 hora.

A partir de 1000 observações geradas aleatoriamente seguindo uma lei exponencial com média 1, que se encontram na secção B.8, construímos 10 histogramas, em que cada um é constituído por 100 observações. A Figura 4.2 apresenta as observações agregadas em períodos de 0.5.

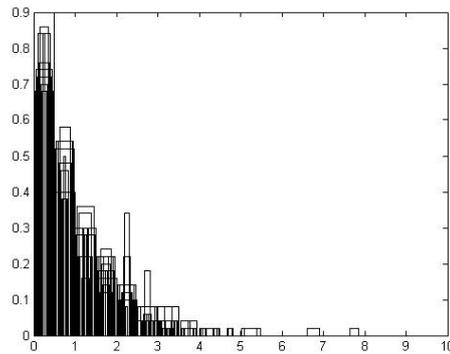


Figura 4.2: Histogramas referentes às observações (em B.8) agregadas em períodos de 0.5, com $x \in [0, 10]$.

4.2. Programa

O programa denominado `tese.m` (secção B.9) testa a abordagem por nós efectuada. Este código em MATLAB tem a seguinte sequência:

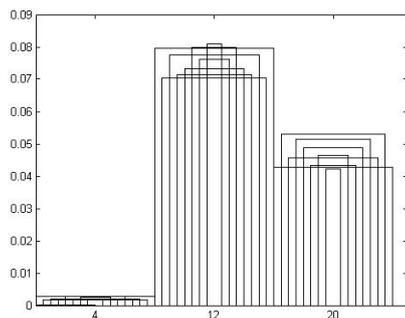
1. Leitura dos dados a satisfazerem as condições pretendidas (área fica unitária para cada segunda-feira);

2. Construção da matriz \mathcal{A} e dos vectores-coluna \mathcal{B} e \mathcal{D} , componentes referidas na secção 3.2.3;
3. Construção das matrizes B e C (secção 3.2.2),
4. Construção do sistema com as condições de optimalidade, referido em (3.43) e (3.41);
5. Construção de uma matriz com os coeficientes dos splines cúbicos correspondentes, referidos no teorema 3.1.2;
6. Construção dos histogramas referentes aos dados pretendidos.

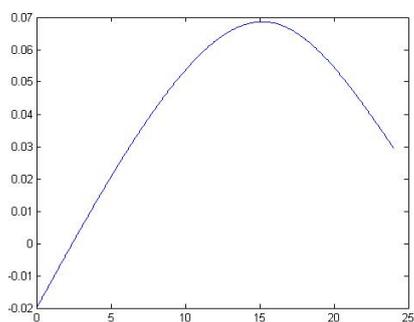
4.3. Resultados preliminares com dados reais

A Figura 4.3 apresenta o estimador spline cúbico descrito para diversas condições iniciais. São resultados preliminares que ilustram bem as características de cada condição inicial e algumas das limitações da nossa abordagem.

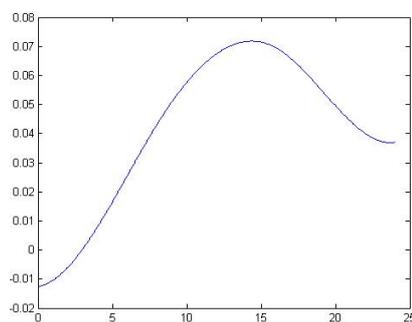
Nomeadamente, não se impõe a não negatividade da função encontrada. No entanto, nota-se que a condição inicial que nos parece mais adequada aos dados é a da Figura 4.3(d) e para esta a função encontrada é não negativa.



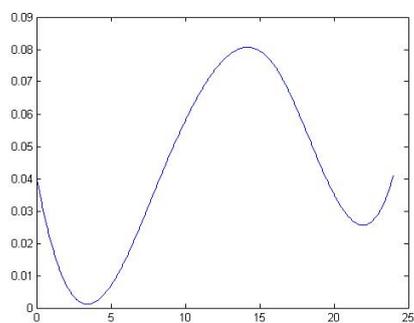
(a) Dados reais agregados em períodos de 8 horas



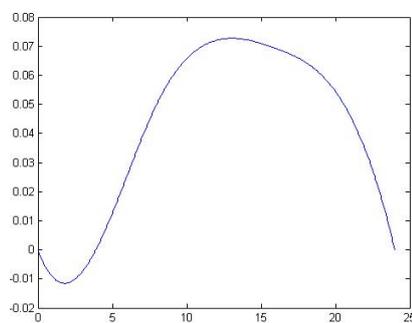
(b) $M_0 = 0 = M_n$



(c) $m_0 = m_n, M_0 = M_n$



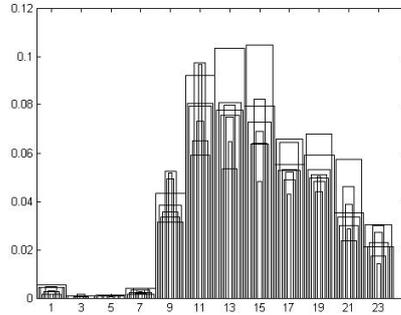
(d) $H_0 = H_n, M_0 = M_n$



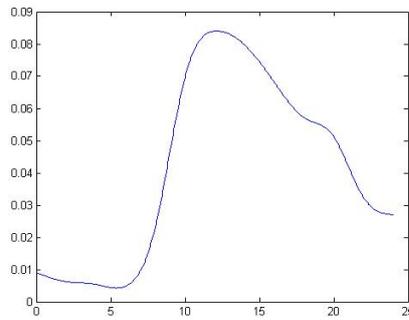
(e) $H_0 = 0, H_n = 0$

Figura 4.3: Estimador spline cúbico, para diversas condições iniciais, com os dados reais em períodos de 8 horas

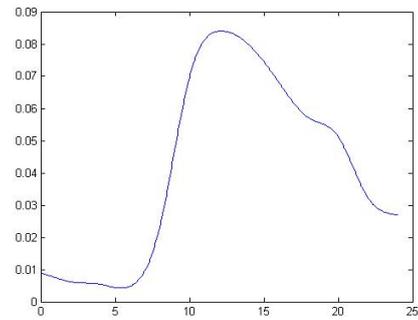
Na Figura 4.4 apresentamos os resultados obtidos a partir do mesmo conjunto de dados, em períodos de 2 horas.



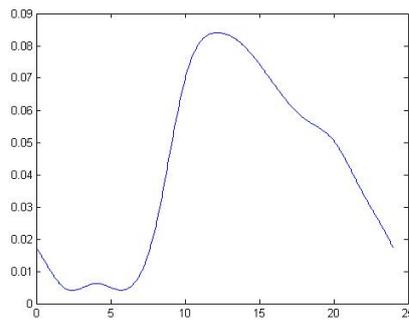
(a) Dados reais agregados em períodos de 2 horas



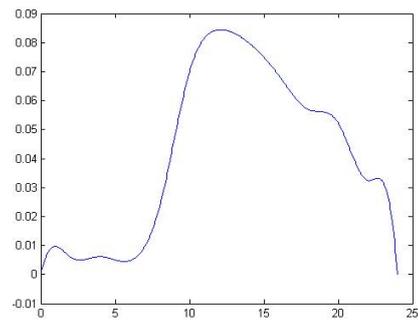
(b) $M_0 = 0 = M_n$



(c) $m_0 = m_n, M_0 = M_n$



(d) $H_0 = H_n, M_0 = M_n$



(e) $H_0 = 0, H_n = 0$

Figura 4.4: Estimador spline cúbico, para diversas condições iniciais, com os dados reais em períodos de 2 horas.

Para a restante experiência computacional, vamos considerar apenas a condição inicial $H_0 = H_n, M_0 = M_n$.

4.4. Tamanho do *bin*

Nesta secção pretendemos observar de que modo diferentes períodos de agregação afectam a função H . Assim sendo, agregamos os dados em períodos de 8, 4, 2, 1, 1/2 e 1/4 horas.

As Figuras 4.5-4.10 ilustram os resultados obtidos para esses diferentes períodos, com as condições iniciais $H_0 = H_n$; $M_0 = M_n$

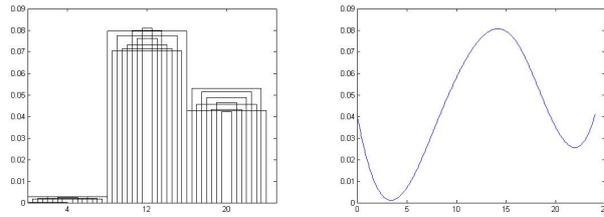


Figura 4.5: Períodos de 8 horas.

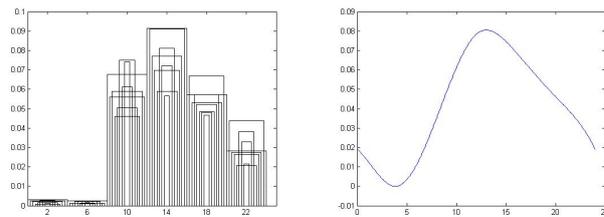


Figura 4.6: Períodos de 4 horas.

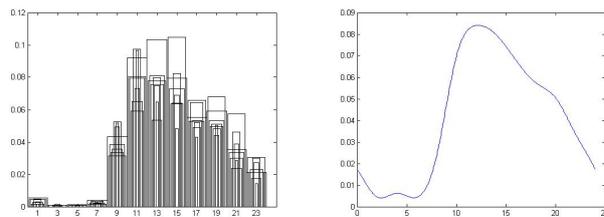


Figura 4.7: Períodos de 2 horas.

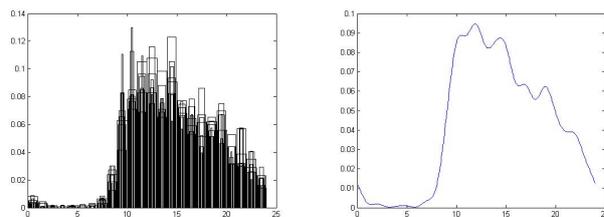


Figura 4.8: Períodos de 1 hora.

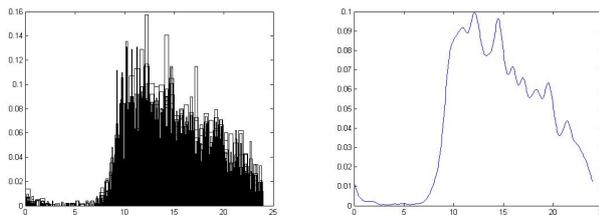


Figura 4.9: Períodos de 1/2 hora.

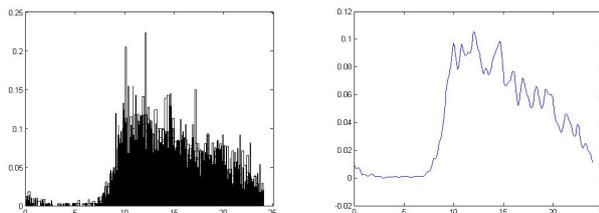


Figura 4.10: Períodos de 1/4 hora.

Com as Figuras anteriores observamos a influência de diferentes períodos (*bins*) no estimador. Concluímos então que quanto mais pequena a *bin*, mais oscilações obtemos na função H .

4.5. Método do núcleo

Utilizaremos agora o estimador do método do núcleo descrito na secção 2.2 para os mesmos dados reais.

É de notar que para este estimador não é necessário definir condições iniciais *a priori* como no estimador spline cúbico, mas é necessário definir o h (a *janela*).

4.5.1. Programa

O programa que se encontra na secção B.2 do apêndice B, procede à geração (determinística) dos instantes em que efectivamente ocorreram as chegadas dos dados reais. De seguida é utilizado o programa da secção B.3 no apêndice B, no qual obtemos os gráficos da estimação da densidade do núcleo para determinado h (definido pelo utilizador).

4.5.2. Resultados

Nas Figuras que se seguem apresentamos os resultados obtidos para a estimação da densidade do núcleo com diferentes h 's, para os dados reais com períodos de 8, 2 e 1/4 horas.

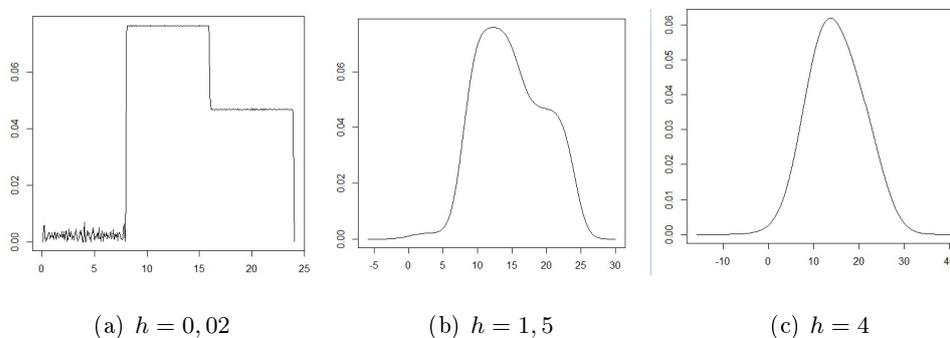


Figura 4.11: Estimador do Núcleo com os dados referentes a 8 segundas-feiras, em períodos de 8 horas com: (a) $h = 0,02$, (b) $h = 1,5$ e (c) $h = 4$.

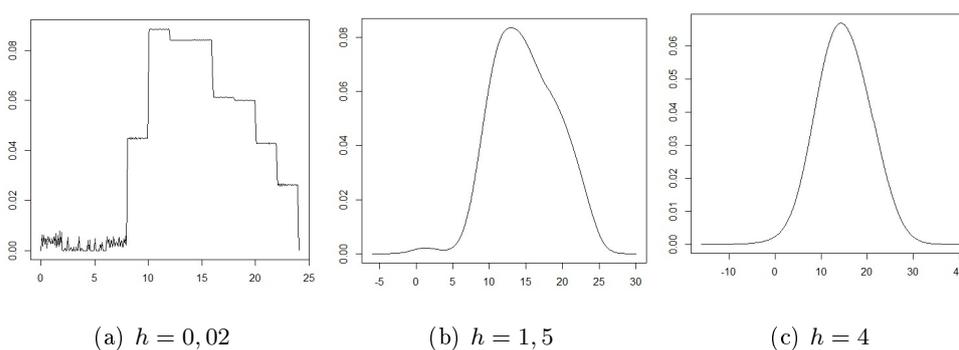


Figura 4.12: Estimador do Núcleo com os dados referentes a 8 segundas-feiras, em períodos de 2 horas com: (a) $h = 0,02$, (b) $h = 1,5$ e (c) $h = 4$.

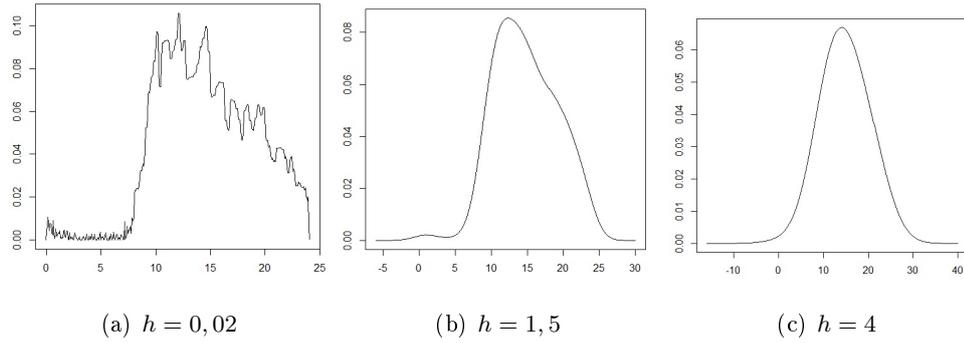


Figura 4.13: Estimador do Núcleo com os dados referentes a 8 segundas-feiras, em períodos de 1/4 horas com (a) $h = 0.02$, (b) $h = 1,5$ e (c) $h = 4$.

Verificamos que em qualquer tipo de agregação um h grande cria oscilações, já para um h pequeno obtemos uma estimação mais “real”.

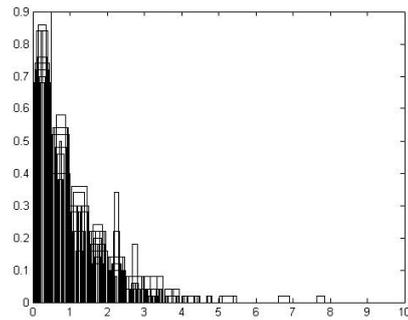
4.6. Resultados com a lei exponencial

De todas as condições iniciais referidas anteriormente a condição $H_0 = 0 = H_n$ é a que melhor se aplica quando se trata de distribuições de probabilidade. Seguidamente acrescentamos mais um conjunto de condições iniciais que melhor se ajusta ao caso da lei exponencial,

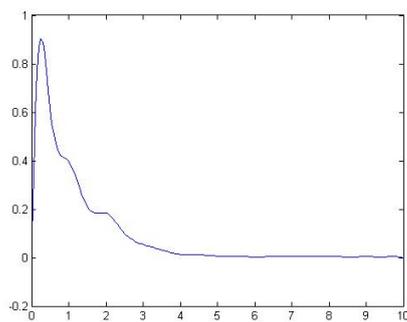
$$H_0 = 1 \quad (4.1a)$$

$$H_n = 0 \quad (4.1b)$$

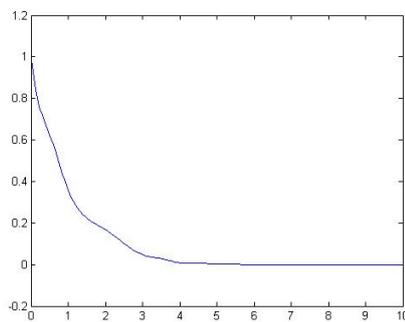
A Figura 4.14 apresenta o estimador spline cúbico para as 1000 observações da lei exponencial de média 1 geradas, com período 0.5, para diferentes condições iniciais.



(a) Observações agregadas em períodos de 0.5



(b) $H_0 = 0, H_n = 0$



(c) $H_0 = 1, H_n = 0$

Figura 4.14: Estimador spline cúbico, com período 0.5, com $x \in [0, 10]$, para diferentes condições iniciais

Sabemos que a função densidade exponencial tem a seguinte forma:

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} 1_{]0, +\infty[}(x) \quad (4.2)$$

No nosso caso como a média toma o valor 1, o nosso θ será 1.

A Figura que se segue, 4.15, foi obtida através do programa que se encontra na secção B.10. Esta figura consiste no gráfico da função densidade da lei exponencial teórica com parâmetro 1, obtida através das observações geradas anteriormente.

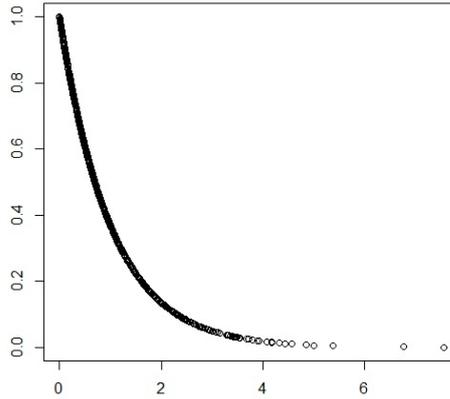


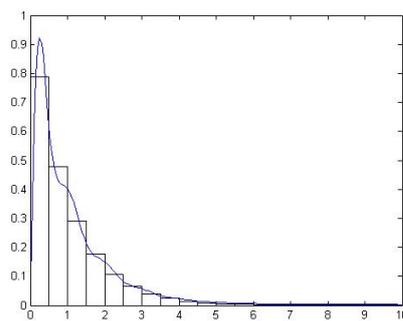
Figura 4.15: Função densidade da lei exponencial com parâmetro 1.

Comparando esta figura com as Figuras 4.14(b) e 4.14(c), observamos que a função da Figura 4.14(b) não começa como a função densidade teórica, pois impomos a condição $H_0 = 0$. Enquanto que na Figura 4.14(c) notamos que é bastante parecida com a função densidade teórica.

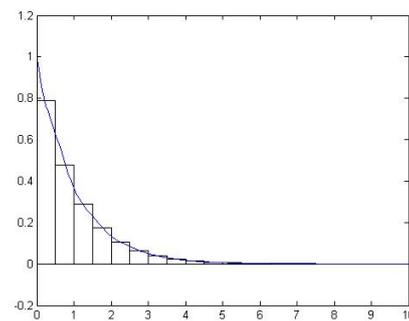
Consideramos agora um histograma no qual em cada bin a área é igual a $P((i - h)h \leq X \leq ih)$, sendo X exponencial.

$$P((i - 1)h \leq X \leq ih) = \int_{(i-1)h}^{ih} e^{-x} = -e^{-ih} + e^{-(i-1)h} = \frac{e^{-(i-1)h} - e^{-ih}}{h}$$

Utilizando este resultado no programa da secção B.9 obtemos as figuras que se seguem,



(a) $H_0 = 0, H_n = 0$



(b) $H_0 = 1, H_n = 0$

Figura 4.16: Estimador spline cúbico para função densidade teórica, com período 0.5, para diferentes condições iniciais

Observamos que obtivemos resultados idênticos aos já obtidos nas Figuras 4.14(b) e 4.16(b).

Capítulo 5

Conclusões

Nesta dissertação desenvolvemos uma metodologia baseada em splines para a estimação de uma densidade de probabilidade a partir do conhecimento de um conjunto finito de histogramas subjacentes à mesma distribuição de probabilidade. Temos em mente uma aplicação que é a estimação da função intensidade $\lambda(t)$ de um processo de Poisson não estacionário.

Comparando a nossa abordagem com o conhecido método do núcleo para estimação não paramétrica, o método apresenta algumas vantagens sobre este. Nomeadamente, permite assegurar que o suporte da distribuição de se mantenha não negativo e permite definir funções com poucas oscilações ao mesmo que identificando *tendências* facilmente identificáveis nos histogramas e curiosamente justificáveis para a aplicação em causa.

A nossa abordagem parece, de facto, obviar a questão das excessivas oscilações observadas na tese de César [1].

No caso de as observações subjacentes serem provenientes de uma distribuição exponencial, a nossa abordagem aproxima razoavelmente a densidade dessa lei para o caso da média ser unitária.

Apêndice A

Histogramas

Este histograma representa um pequeno exemplo para o vector $(3, 7)$ com período 10, ou seja, de 0 a 10 temos 3 chegadas, de 10 a 20 temos 7 chegadas.

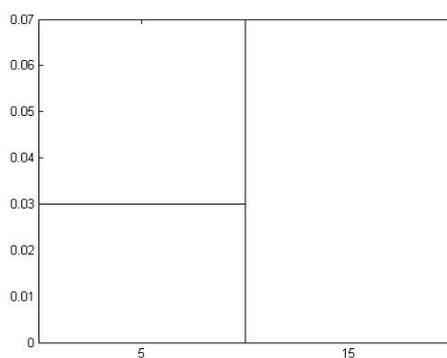


Figura A.1: Histograma para o vector $(3, 7)$

A figura seguinte representa outro pequeno exemplo com período 10 também, mas para o vector $(8, 2)$

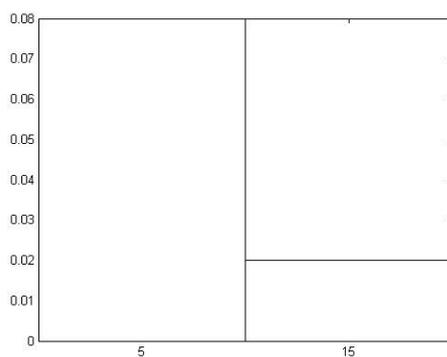


Figura A.2: Histograma para o vector $(8, 2)$

Apêndice B

Listagem de programas

B.1. dados.m

dados.m

```
%  
% Dados referentes a um Call center, registrados  
% de 1/4 em 1/4 de hora das 0:00am até 12:00pm durante 8 dias  
%  
AA = [ 3  4  1  1  1  4  0  2  
6  1  3  2  1  2  4  0  
1  1  1  1  1  2  3  1  
0  1  1  0  0  3  0  0  
1  0  3  0  2  0  0  0  
2  0  0  0  0  0  0  0  
1  1  0  0  2  1  1  0  
0  3  2  0  0  0  0  1  
1  1  0  0  0  0  1  0  
0  0  0  0  1  0  1  0  
0  1  0  0  0  0  0  2  
0  0  0  0  0  0  0  0  
0  0  0  0  0  0  0  0  
1  0  0  0  0  0  1  0  
0  0  0  0  0  1  0  0  
0  0  0  0  0  1  1  0  
0  0  0  0  1  0  0  1  
1  0  0  0  0  0  1  0  
0  0  0  0  0  0  0  0  
1  1  0  1  0  0  1  0
```

Apêndice B Listagem de programas

0 0 0 0 1 0 0 0
0 0 0 2 1 0 0 0
0 0 0 0 0 0 0 0
1 1 0 0 0 0 0 0
0 0 0 1 0 1 0 0
0 0 0 0 2 1 0 0
0 0 0 1 0 0 1 0
2 0 0 1 0 0 0 0
1 1 1 0 0 0 0 1
0 0 0 3 1 1 3 3
2 1 2 1 3 1 0 0
5 2 3 2 4 1 5 3
8 3 4 3 3 10 3 1
5 5 7 3 7 4 4 3
12 3 12 3 8 6 13 8
6 14 8 4 11 10 14 6
12 17 12 12 6 9 20 31
15 13 21 17 15 29 18 27
26 16 23 18 21 25 22 25
28 10 12 26 21 42 33 32
20 17 32 17 23 65 24 28
12 23 16 16 15 21 26 40
37 29 23 27 19 28 18 27
32 27 25 20 20 29 24 40
38 28 24 18 19 29 13 37
35 34 12 14 25 22 31 31
29 23 33 17 19 31 28 30
34 27 40 23 28 26 25 16
38 26 30 58 20 31 27 19
36 24 36 24 11 25 30 26
35 37 26 15 24 34 25 22
30 25 17 24 17 20 22 18
40 17 20 19 14 18 30 25
23 20 27 26 14 27 20 17

32 22 25 14 18 26 18 21
32 29 28 15 20 25 21 19
28 47 25 17 23 29 26 17
30 48 24 25 23 28 21 17
31 33 32 19 31 46 23 16
28 38 25 25 24 26 22 15
27 28 15 19 18 15 20 10
15 22 24 18 15 22 24 18
27 25 18 29 9 19 15 17
19 29 25 12 22 27 27 14
19 29 21 23 25 20 14 19
17 27 16 13 7 16 19 19
25 16 10 17 18 9 17 10
15 15 13 23 25 25 25 22
13 24 23 39 15 13 16 9
23 25 17 21 15 14 14 13
13 18 20 19 15 21 9 10
17 16 16 11 17 8 20 9
19 21 18 21 13 16 11 18
30 19 16 15 22 13 21 17
18 22 17 13 13 21 15 11
13 19 23 9 12 15 11 14
20 24 18 17 18 7 13 14
20 25 12 16 20 17 19 22
14 23 11 21 12 21 10 19
18 22 22 16 22 14 14 15
18 16 22 8 7 13 15 10
10 15 12 20 6 15 7 9
7 16 5 11 10 9 7 16
6 24 11 7 7 9 7 7
16 19 5 10 6 22 14 7
7 16 13 11 14 13 21 12
13 18 8 4 5 29 15 7
14 24 10 6 6 9 14 6

```
7 18 6 5 5 10 8 8
14 11 9 17 7 12 8 6
7 13 11 11 8 8 23 3
9 5 3 9 5 10 7 4
3 13 11 6 4 8 7 3
8 10 4 5 6 8 12 3
5 3 5 4 3 17 1 5
2 5 6 2 7 4 4 5 ];
```

B.2. Programa programa_abcissas_R.m

O programa programa_abcissas_R.m gera os instantes em que efectivamente ocorreram as chegadas dos dados (**dados.m**) agregados em períodos de 8, 4, 2, 1, 1/2 e 1/4 horas.

```
    programa_abcissas_R.m

%
% Programa para criar as abcissas para futura utilização no programa em R
% com dados da matriz AA
%
% Matriz AA (dados.m) contém o número de chegadas em cada quarto de hora durante
% 24 horas começando à meia-noite em oito segundas-feiras
%
% dados_Call_Center_8horas: matriz AA agregada em períodos de 8 horas
% dados_Call_Center_4horas: matriz AA agregada em períodos de 4 horas
% dados_Call_Center_2horas: matriz AA agregada em períodos de 2 horas
% dados_Call_Center_1horas: matriz AA agregada em períodos de 1 hora
% dados_Call_Center_30min: matriz AA agregada em períodos de 30 minutos
% dados_Call_Center_15min: matriz AA agregada em períodos de 15 minutos
%

p = input('Periodo do número de chegadas que pretende (em horas): ');
disp(p)

switch p
    case 8
```

```
        dados_Call_Center_8horas;
    case 4
        dados_Call_Center_4horas;
    case 2
        dados_Call_Center_2horas;
    case 1
        dados_Call_Center_1horas;
    case 0.5
        dados_Call_Center_30min;
    case 0.25
        dados_Call_Center_15min;
end;
```

```
% tamanho da matriz AA
```

```
[n,m]=size(AA)
```

```
d = input('Dia que pretende: ');
```

```
disp(d)
```

```
% j - o dia que se pretende
```

```
j = d;
```

```
x = [] ;
```

```
i=0;
```

```
while (i<n)
```

```
    ult = i*periodo ;
```

```
    i = i+1 ;
```

```
    ncheg = AA(i,j) ;
```

```
    if ncheg > 0
```

```
        h = periodo / ncheg ;
```

```
ult = ult + h/2 ;
c = 0 ;
while c < ncheg
    x = [ x ; ult ] ;
    ult = ult + h ;
    c = c + 1 ;
end;
end;
end;

x

[n,m]=size(x)
```

B.3. Programa estimadorNucleo.txt

O programa `estimadorNucleo.txt` em R que traça o gráfico da estimação da densidade do núcleo para os dados obtidos em programa `_abcissas_R.m` e em programa `_abcissas_R_exemplo_histogramas.m`

```
estimadorNucleo.m

y<-scan('primeira_segundafeira.txt')
est<-bkde(y, bandwidth=3)
plot(est,type='l')
```

B.4. Programa programa_abcissas_R_exemplo_histogramas.m

O programa `programa_abcissas_R_exemplo_histogramas.m` gera os instantes em que efectivamente ocorreram para os dados dos histogramas A.1 e A.2 que se encontram no apêndice A.

```
programa_abcissas_R_exemplo_histogramas.m

%
% Programa para criar as abcissas para futura utilização no programa em R
% Dados: f=(0.3, 0.7), ou f=(0.8,0.2)
%
```

```

AA=[3; 7]
%AA = [8; 2]

% tamanho da matriz AA
[n,m]=size(AA)

j = 1;

x = [] ;

i=0;

while (i<n)

    ult = i*10 ;
    i   = i+1   ;

    ncheg = AA(i,j) ;

    if ncheg > 0
        h = 10 / ncheg ;
        ult = ult + h/2 ;
        c = 0 ;
        while c < ncheg
            x   = [ x ; ult ] ;
            ult = ult + h ;
            c   = c + 1 ;
        end;
    end;
end;

x

[n,m]=size(x)

```

B.5. Função cubicsplinearea.m

A função **cubicsplinearea.m** é utilizada nos programas B.6 e B.7 de forma a calcular splines cúbicos com área previamente fixada em cada subintervalo para dois conjuntos de condições iniciais: $\{H_0 = 0; H_n = 0; M_0 = M_n\}$ e $\{H_0 = H_n; m_0 = m_n; M_0 = M_n\}$

cubicsplinearea.m

```
function s = cubicsplinearea (x, f, zi, type )

%
% Programa spline cúbico de área fixa
% Para matrizes completas (não singulares) e não completas (singulares)
%
% Condições iniciais:
% Se TYPE=1-> H_0=0; H_n=0; M_0=M_n
% Se TYPE=2 -> H_0=H_n; m_0=m_n; M_0=M_n
%

[n,m] = size(x);

%Transposição dos vectores

if n==1
    x=x'; f=f'; n=m;
end

h = x(2:end)-x(1:end-1) ;
h2 = h(1:end-1) ;
h1 = h(2:end)+h(1:end-1);
h3 = h(2:end) ;
h4 = h2.*h3 ;
h5=h(end);
```

```
lixo = 50;

%Matriz B
e1 = 2*[lixo; h1; lixo];
B = diag(e1)+diag(h,-1)+diag(h,1) ;
B(1,:) = 0 ;
B(end,:) = 0 ;
B

%Matriz C
e1 = 6*[lixo; -1*h1./h4; lixo] ;
e2 = 6./h;
C = diag(e1)+diag(e2,-1)+diag(e2,1) ;
C(1,:) = 0 ;
C(end,:) = 0 ;
C

A = [B -C];

rhs = [zeros(n+1,1)];

% Type 1
if type==1
    A(1,n+1) = 1 ;
    A(n,2*n) = 1 ;
    A(n+1,1) = 1 ;
    A(n+1,n) = -1 ;
end

% Type 2
if type==2
    A(1,n+1) = 1 ;
    A(1,2*n) = -1 ;
```

Apêndice B Listagem de programas

```
A(n,1) = 1 ;
A(n,n) = -1 ;

A(n+1,1) = -h(1)/3 ;
A(n+1,2) = -h(1)/6 ;
A(n+1,n-1) = -h(end)/6 ;
A(n+1,n) = -h(end)/3 ;
A(n+1,n+1) = -1/h(1) ;
A(n+1,n+2) = 1/h(1) ;
A(n+1,2*n-1) = 1/h(end) ;
A(n+1,2*n) = -1/h(end) ;
end
A

% Matriz D
e1 = [-1.*(h.*h)./24];
e2 = [-1.*(h2.*h2)./24];
D = diag(e1) + diag(e2,1);
D(n-1,n)= -1.*(h5.*h5)./24;
%D

% Matriz E
e1 = [ones(n-1,1)./2];
e2 = [ones(n-2,1)./2];
E = diag(e1) + diag(e2,1);
E(n-1,n)= 1./2;
%E

A= [A; D E];
A

% Caracteristica de A
rank(A)
```

```

% Tamanho de A
size(A)

rhs=[rhs; f./h]

[w,t]=size(A)

if w==rank(A)
    % Caracteristica completa
    y=A\rhs ;
else
    % Caracteristica nao completa
    pinv(A)
    y=pinv(A)*rhs;
end;
y

d = (y(n+2:end)-y(n+1:end-1))./h;
%d
S=zeros(n,4);
S(:,1)=y(n+1:2*n);
S(:,3)=y(1:n)/2;
%S

for m=1:n-1
    S(m,4) = (S(m+1,3)-S(m,3))/3/h(m);
    S(m,2) = d(m)-h(m)/3*(S(m+1,3)+2*S(m,3));
end

S
S = S(1:n-1, 4:-1:1);
S

pp = mkpp(x,S);

```

```
s = ppval(pp,zi);  
return
```

B.6. Programa dados_problema_dummy_areafixa.m

O programa `dados_problema_dummy_areafixa.m` traça o gráfico do spline cúbico com área previamente fixada em cada subintervalo para os pontos $\{0, 10, 20, 30\}$, num conjunto de áreas $\{200, 400, 100\}$, para os dois conjuntos de condições iniciais: $\{H_0 = 0; H_n = 0; M_0 = M_n\}$ e $\{H_0 = H_n; m_0 = m_n; M_0 = M_n\}$, utilizando a função B.5.

```
dados_problema_dummy_areafixa.m  
  
%  
% Dados para problema dummy  
% com solução explícita conhecida  
%  
% Pontos: {0,10,20,30}  
% Conjunto de áreas: {200,400,100}  
%  
  
x = [0; 10; 20; 30] ;  
f = [200; 400; 100] ;  
z = [0:0.05:30] ;  
  
w = x-5 ;  
y = f/10 ;  
bar(w(2:end), y, 1, 'w')  
hold on  
  
s=cubicsplinearea(x,f,z,2);  
plot(z,s,'k')
```

B.7. Programa dados_problema_dummy2_areafixa.m

O programa `dados_problema_dummy2_areafixa.m` traça o gráfico do spline cúbico com área previamente fixada em cada subintervalo para os pontos $\{0, 10, 20, 30, 40, 50, 60\}$,

num conjunto de áreas $\{160, 420, 310, 145, 45, 10\}$, para os dois conjuntos de condições iniciais: $\{H_0 = 0; H_n = 0; M_0 = M_n\}$ e $\{H_0 = H_n; m_0 = m_n; M_0 = M_n\}$, utilizando a função B.5.

`dados_problema_dummy2_areafixa.m`

```
%
% Dados para problema dummy
% com solução explicita conhecida
%
% Pontos: {0,10,20,30,40,50,60}
% Conjunto de áreas: {160,420,310,145,45,10}
%

x = [0; 10; 20; 30; 40; 50; 60];
f = [160; 420; 310; 145; 45; 10];
z = [0:0.1:60];

w = x-5    ;
y = f/10   ;
bar(w(2:end), y, 1, 'w')
hold on

s=cubicsplinearea(x,f,z,2);
plot(z,s,'k')
```

B.8. `dados_leiexponencial.m`

Dados que foram gerados através do comando “`exprnd(1/λ, 1000, 1)`” no MATLAB, este comando consiste em gerar 1000 observações, cujas quais seguem uma lei exponencial de média 1 ($1/\lambda = 1$), num vector-coluna.

`dados_leiexponencial.m`

B.9. Programa `tese.m`

O programa `tese.m` em MATLAB implementa a teoria desenvolvida na secção 3.2. Os gráficos resultantes das execuções encontram-se no capítulo 4.

`tese.m`

Apêndice B Listagem de programas

```
%
% Programa tese
%
type = 3 ;
dd = input('dados: 0 - dados reais; 1 - observações a seguir uma exponencial de média 1:')
if dd==0
periodo = 1/4 ;
Ind_dias = [1 2 3 4 5 6 7 8] ;
% 'f' contém as áreas de cada bin; cada coluna corresponde a um dia distinto
% periodo - tamanho da bin
% Dados do call center agregados em períodos de 8h
if periodo==8
f= [0.0225 0.0141 0.0150 0.0153 0.0219 0.0147 0.0204 0.0135;
    0.6362 0.5626 0.6193 0.5724 0.5871 0.6393 0.6090 0.6477;
    0.3413 0.4233 0.3657 0.4123 0.3910 0.3460 0.3706 0.3388];
end;
% Dados do call center agregados em períodos de 4
if periodo==4
f= [0.0124 0.0101 0.0085 0.0031 0.0062 0.0109 0.0093 0.0047
    0.0101 0.0047 0.0047 0.0093 0.0101 0.0039 0.0085 0.0062
    0.2708 0.2242 0.2358 0.1846 0.2017 0.2995 0.2452 0.2964
    0.3654 0.3646 0.3080 0.2785 0.2351 0.3243 0.2878 0.2258
    0.2281 0.2676 0.2118 0.2281 0.2087 0.1939 0.1924 0.1870
    0.1133 0.1753 0.1094 0.1055 0.0822 0.1521 0.1319 0.0861 ];
end;
% Dados do call center agregados em períodos de 2h
if periodo==2
f= [0.0109 0.0085 0.0085 0.0031 0.0054 0.0093 0.0062 0.0031
    0.0016 0.0016 0.0000 0.0000 0.0008 0.0016 0.0031 0.0016
    0.0023 0.0016 0.0000 0.0023 0.0023 0.0000 0.0016 0.0008
    0.0078 0.0031 0.0047 0.0070 0.0078 0.0039 0.0070 0.0054
    0.0869 0.0628 0.0768 0.0667 0.0714 0.1047 0.0985 0.1032
    0.1839 0.1614 0.1590 0.1179 0.1303 0.1947 0.1466 0.1932
    0.2064 0.1552 0.1621 0.1513 0.1071 0.1598 0.1497 0.1296
```

```
0.1590 0.2095 0.1458 0.1272 0.1280 0.1645 0.1381 0.0962
0.1102 0.1319 0.1055 0.1288 0.1063 0.0978 0.1040 0.0861
0.1179 0.1358 0.1063 0.0993 0.1024 0.0962 0.0884 0.1009
0.0706 0.1148 0.0667 0.0597 0.0473 0.0923 0.0776 0.0574
0.0427 0.0605 0.0427 0.0458 0.0349 0.0597 0.0543 0.0287 ];
end;
% Dados do call center agregados em períodos de 1h
if periodo==1
f = [0.0078 0.0052 0.0053 0.0038 0.0031 0.0087 0.0062 0.0029;
0.0031 0.0030 0.0044 0.0000 0.0042 0.0008 0.0009 0.0010;
0.0008 0.0015 0.0000 0.0000 0.0010 0.0000 0.0018 0.0019;
0.0008 0.0000 0.0000 0.0000 0.0000 0.0016 0.0018 0.0000;
0.0016 0.0007 0.0000 0.0010 0.0010 0.0000 0.0018 0.0010;
0.0008 0.0007 0.0000 0.0019 0.0021 0.0000 0.0000 0.0000;
0.0016 0.0000 0.0000 0.0029 0.0021 0.0016 0.0009 0.0000;
0.0062 0.0030 0.0053 0.0058 0.0083 0.0024 0.0071 0.0067;
0.0240 0.0185 0.0274 0.0125 0.0302 0.0236 0.0301 0.0173;
0.0628 0.0415 0.0601 0.0700 0.0657 0.0827 0.0824 0.1107;
0.0784 0.0712 0.0848 0.0767 0.0803 0.1127 0.0816 0.1299;
0.1055 0.0830 0.0963 0.0690 0.0949 0.0851 0.0860 0.1097;
0.1078 0.0830 0.0963 0.1160 0.0751 0.0867 0.0922 0.0818;
0.0985 0.0652 0.0883 0.0709 0.0688 0.0757 0.0789 0.0789;
0.0908 0.1231 0.0936 0.0825 0.1053 0.1017 0.0816 0.0626;
0.0683 0.0771 0.0724 0.0748 0.0667 0.0654 0.0762 0.0568;
0.0590 0.0645 0.0530 0.0729 0.0782 0.0552 0.0665 0.0674;
0.0512 0.0615 0.0671 0.0863 0.0647 0.0441 0.0523 0.0395;
0.0621 0.0600 0.0654 0.0556 0.0626 0.0512 0.0514 0.0577;
0.0559 0.0697 0.0557 0.0671 0.0751 0.0465 0.0496 0.0674;
0.0318 0.0526 0.0442 0.0441 0.0313 0.0362 0.0319 0.0404;
0.0388 0.0571 0.0318 0.0297 0.0323 0.0575 0.0567 0.0308;
0.0287 0.0348 0.0256 0.0403 0.0261 0.0315 0.0408 0.0202;
0.0140 0.0230 0.0230 0.0163 0.0209 0.0292 0.0213 0.0154];
end;
% Dados do call center agregados em períodos de 30min
```

Apêndice B Listagem de programas

```
if periodo==0.5
f = [0.0070 0.0037 0.0035 0.0029 0.0021 0.0047 0.0035 0.0019;
     0.0008 0.0015 0.0018 0.0010 0.0010 0.0039 0.0027 0.0010;
     0.0023 0.0000 0.0027 0.0000 0.0021 0.0000 0.0000 0.0000;
     0.0008 0.0030 0.0018 0.0000 0.0021 0.0008 0.0009 0.0010;
     0.0008 0.0007 0.0000 0.0000 0.0010 0.0000 0.0018 0.0000;
     0.0000 0.0007 0.0000 0.0000 0.0000 0.0000 0.0000 0.0019;
     0.0008 0.0000 0.0000 0.0000 0.0000 0.0000 0.0009 0.0000;
     0.0000 0.0000 0.0000 0.0000 0.0000 0.0016 0.0009 0.0000;
     0.0008 0.0000 0.0000 0.0000 0.0010 0.0000 0.0009 0.0010;
     0.0008 0.0007 0.0000 0.0010 0.0000 0.0000 0.0009 0.0000;
     0.0000 0.0000 0.0000 0.0019 0.0021 0.0000 0.0000 0.0000;
     0.0008 0.0007 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000;
     0.0000 0.0000 0.0000 0.0010 0.0021 0.0016 0.0000 0.0000;
     0.0016 0.0000 0.0000 0.0019 0.0000 0.0000 0.0009 0.0000;
     0.0008 0.0007 0.0009 0.0029 0.0010 0.0008 0.0027 0.0038;
     0.0054 0.0022 0.0044 0.0029 0.0073 0.0016 0.0044 0.0029;
     0.0101 0.0059 0.0097 0.0058 0.0104 0.0110 0.0062 0.0038;
     0.0140 0.0126 0.0177 0.0067 0.0198 0.0126 0.0239 0.0135;
     0.0209 0.0222 0.0292 0.0278 0.0219 0.0299 0.0337 0.0558;
     0.0419 0.0193 0.0309 0.0422 0.0438 0.0528 0.0488 0.0549;
     0.0248 0.0297 0.0424 0.0316 0.0396 0.0678 0.0443 0.0654;
     0.0535 0.0415 0.0424 0.0451 0.0407 0.0449 0.0372 0.0645;
     0.0566 0.0460 0.0318 0.0307 0.0459 0.0402 0.0390 0.0654;
     0.0489 0.0371 0.0645 0.0384 0.0490 0.0449 0.0470 0.0443;
     0.0574 0.0371 0.0583 0.0786 0.0323 0.0441 0.0505 0.0433;
     0.0504 0.0460 0.0380 0.0374 0.0428 0.0426 0.0417 0.0385;
     0.0489 0.0274 0.0415 0.0431 0.0292 0.0355 0.0443 0.0404;
     0.0497 0.0378 0.0468 0.0278 0.0396 0.0402 0.0346 0.0385;
     0.0450 0.0704 0.0433 0.0403 0.0480 0.0449 0.0417 0.0327;
     0.0458 0.0526 0.0504 0.0422 0.0574 0.0567 0.0399 0.0298;
     0.0326 0.0371 0.0345 0.0355 0.0344 0.0292 0.0390 0.0269;
     0.0357 0.0400 0.0380 0.0393 0.0323 0.0362 0.0372 0.0298;
     0.0279 0.0415 0.0327 0.0345 0.0334 0.0284 0.0293 0.0366;
```

```
0.0310 0.0230 0.0203 0.0384 0.0448 0.0268 0.0372 0.0308;
0.0279 0.0363 0.0353 0.0575 0.0313 0.0213 0.0266 0.0212;
0.0233 0.0252 0.0318 0.0288 0.0334 0.0229 0.0257 0.0183;
0.0380 0.0297 0.0300 0.0345 0.0365 0.0229 0.0284 0.0337;
0.0240 0.0304 0.0353 0.0211 0.0261 0.0284 0.0230 0.0241;
0.0310 0.0363 0.0265 0.0316 0.0396 0.0189 0.0284 0.0346;
0.0248 0.0334 0.0292 0.0355 0.0355 0.0276 0.0213 0.0327;
0.0217 0.0230 0.0300 0.0268 0.0136 0.0221 0.0195 0.0183;
0.0101 0.0297 0.0141 0.0173 0.0177 0.0142 0.0124 0.0221;
0.0178 0.0259 0.0159 0.0201 0.0209 0.0276 0.0310 0.0183;
0.0209 0.0311 0.0159 0.0096 0.0115 0.0299 0.0257 0.0125;
0.0163 0.0215 0.0133 0.0211 0.0125 0.0173 0.0142 0.0135;
0.0124 0.0133 0.0124 0.0192 0.0136 0.0142 0.0266 0.0067;
0.0085 0.0170 0.0133 0.0105 0.0104 0.0126 0.0168 0.0058;
0.0054 0.0059 0.0097 0.0058 0.0104 0.0165 0.0044 0.0096];
end;
% Dados do call center agregados em períodos de 15min
if periodo==0.25
f = [0.0023 0.0030 0.0009 0.0010 0.0010 0.0032 0.0000 0.0019;
0.0047 0.0007 0.0027 0.0019 0.0010 0.0016 0.0035 0.0000;
0.0008 0.0007 0.0009 0.0010 0.0010 0.0016 0.0027 0.0010;
0.0000 0.0007 0.0009 0.0000 0.0000 0.0024 0.0000 0.0000;
0.0008 0.0000 0.0027 0.0000 0.0021 0.0000 0.0000 0.0000;
0.0016 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000;
0.0008 0.0007 0.0000 0.0000 0.0021 0.0008 0.0009 0.0000;
0.0000 0.0022 0.0018 0.0000 0.0000 0.0000 0.0000 0.0010;
0.0008 0.0007 0.0000 0.0000 0.0000 0.0000 0.0009 0.0000;
0.0000 0.0000 0.0000 0.0000 0.0010 0.0000 0.0009 0.0000;
0.0000 0.0007 0.0000 0.0000 0.0000 0.0000 0.0000 0.0019;
0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000;
0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000;
0.0008 0.0000 0.0000 0.0000 0.0000 0.0000 0.0009 0.0000;
0.0000 0.0000 0.0000 0.0000 0.0000 0.0008 0.0000 0.0000;
0.0000 0.0000 0.0000 0.0000 0.0000 0.0008 0.0009 0.0000;
```

Apêndice B Listagem de programas

0.0000	0.0000	0.0000	0.0000	0.0010	0.0000	0.0000	0.0010;
0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0009	0.0000;
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000;
0.0008	0.0007	0.0000	0.0010	0.0000	0.0000	0.0009	0.0000;
0.0000	0.0000	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000;
0.0000	0.0000	0.0000	0.0019	0.0010	0.0000	0.0000	0.0000;
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000;
0.0008	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000;
0.0000	0.0000	0.0000	0.0010	0.0000	0.0008	0.0000	0.0000;
0.0000	0.0000	0.0000	0.0000	0.0021	0.0008	0.0000	0.0000;
0.0000	0.0000	0.0000	0.0010	0.0000	0.0000	0.0009	0.0000;
0.0016	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000	0.0000;
0.0008	0.0007	0.0009	0.0000	0.0000	0.0000	0.0000	0.0010;
0.0000	0.0000	0.0000	0.0029	0.0010	0.0008	0.0027	0.0029;
0.0016	0.0007	0.0018	0.0010	0.0031	0.0008	0.0000	0.0000;
0.0039	0.0015	0.0027	0.0019	0.0042	0.0008	0.0044	0.0029;
0.0062	0.0022	0.0035	0.0029	0.0031	0.0079	0.0027	0.0010;
0.0039	0.0037	0.0062	0.0029	0.0073	0.0032	0.0035	0.0029;
0.0093	0.0022	0.0106	0.0029	0.0083	0.0047	0.0115	0.0077;
0.0047	0.0104	0.0071	0.0038	0.0115	0.0079	0.0124	0.0058;
0.0093	0.0126	0.0106	0.0115	0.0063	0.0071	0.0177	0.0298;
0.0116	0.0096	0.0186	0.0163	0.0156	0.0229	0.0160	0.0260;
0.0202	0.0119	0.0203	0.0173	0.0219	0.0197	0.0195	0.0241;
0.0217	0.0074	0.0106	0.0249	0.0219	0.0331	0.0293	0.0308;
0.0155	0.0126	0.0283	0.0163	0.0240	0.0512	0.0213	0.0269;
0.0093	0.0170	0.0141	0.0153	0.0156	0.0165	0.0230	0.0385;
0.0287	0.0215	0.0203	0.0259	0.0198	0.0221	0.0160	0.0260;
0.0248	0.0200	0.0221	0.0192	0.0209	0.0229	0.0213	0.0385;
0.0295	0.0208	0.0212	0.0173	0.0198	0.0229	0.0115	0.0356;
0.0272	0.0252	0.0106	0.0134	0.0261	0.0173	0.0275	0.0298;
0.0225	0.0170	0.0292	0.0163	0.0198	0.0244	0.0248	0.0289;
0.0264	0.0200	0.0353	0.0221	0.0292	0.0205	0.0222	0.0154;
0.0295	0.0193	0.0265	0.0556	0.0209	0.0244	0.0239	0.0183;
0.0279	0.0178	0.0318	0.0230	0.0115	0.0197	0.0266	0.0250;

0.0272	0.0274	0.0230	0.0144	0.0250	0.0268	0.0222	0.0212;
0.0233	0.0185	0.0150	0.0230	0.0177	0.0158	0.0195	0.0173;
0.0310	0.0126	0.0177	0.0182	0.0146	0.0142	0.0266	0.0241;
0.0178	0.0148	0.0239	0.0249	0.0146	0.0213	0.0177	0.0164;
0.0248	0.0163	0.0221	0.0134	0.0188	0.0205	0.0160	0.0202;
0.0248	0.0215	0.0247	0.0144	0.0209	0.0197	0.0186	0.0183;
0.0217	0.0348	0.0221	0.0163	0.0240	0.0229	0.0230	0.0164;
0.0233	0.0356	0.0212	0.0240	0.0240	0.0221	0.0186	0.0164;
0.0240	0.0245	0.0283	0.0182	0.0323	0.0362	0.0204	0.0154;
0.0217	0.0282	0.0221	0.0240	0.0250	0.0205	0.0195	0.0144;
0.0209	0.0208	0.0133	0.0182	0.0188	0.0118	0.0177	0.0096;
0.0116	0.0163	0.0212	0.0173	0.0156	0.0173	0.0213	0.0173;
0.0209	0.0185	0.0159	0.0278	0.0094	0.0150	0.0133	0.0164;
0.0147	0.0215	0.0221	0.0115	0.0229	0.0213	0.0239	0.0135;
0.0147	0.0215	0.0186	0.0221	0.0261	0.0158	0.0124	0.0183;
0.0132	0.0200	0.0141	0.0125	0.0073	0.0126	0.0168	0.0183;
0.0194	0.0119	0.0088	0.0163	0.0188	0.0071	0.0151	0.0096;
0.0116	0.0111	0.0115	0.0221	0.0261	0.0197	0.0222	0.0212;
0.0101	0.0178	0.0203	0.0374	0.0156	0.0102	0.0142	0.0087;
0.0178	0.0185	0.0150	0.0201	0.0156	0.0110	0.0124	0.0125;
0.0101	0.0133	0.0177	0.0182	0.0156	0.0165	0.0080	0.0096;
0.0132	0.0119	0.0141	0.0105	0.0177	0.0063	0.0177	0.0087;
0.0147	0.0156	0.0159	0.0201	0.0136	0.0126	0.0098	0.0173;
0.0233	0.0141	0.0141	0.0144	0.0229	0.0102	0.0186	0.0164;
0.0140	0.0163	0.0150	0.0125	0.0136	0.0165	0.0133	0.0106;
0.0101	0.0141	0.0203	0.0086	0.0125	0.0118	0.0098	0.0135;
0.0155	0.0178	0.0159	0.0163	0.0188	0.0055	0.0115	0.0135;
0.0155	0.0185	0.0106	0.0153	0.0209	0.0134	0.0168	0.0212;
0.0109	0.0170	0.0097	0.0201	0.0125	0.0165	0.0089	0.0183;
0.0140	0.0163	0.0194	0.0153	0.0229	0.0110	0.0124	0.0144;
0.0140	0.0119	0.0194	0.0077	0.0073	0.0102	0.0133	0.0096;
0.0078	0.0111	0.0106	0.0192	0.0063	0.0118	0.0062	0.0087;
0.0054	0.0119	0.0044	0.0105	0.0104	0.0071	0.0062	0.0154;
0.0047	0.0178	0.0097	0.0067	0.0073	0.0071	0.0062	0.0067;

Apêndice B Listagem de programas

```
0.0124 0.0141 0.0044 0.0096 0.0063 0.0173 0.0124 0.0067;
0.0054 0.0119 0.0115 0.0105 0.0146 0.0102 0.0186 0.0115;
0.0101 0.0133 0.0071 0.0038 0.0052 0.0229 0.0133 0.0067;
0.0109 0.0178 0.0088 0.0058 0.0063 0.0071 0.0124 0.0058;
0.0054 0.0133 0.0053 0.0048 0.0052 0.0079 0.0071 0.0077;
0.0109 0.0082 0.0080 0.0163 0.0073 0.0095 0.0071 0.0058;
0.0054 0.0096 0.0097 0.0105 0.0083 0.0063 0.0204 0.0029;
0.0070 0.0037 0.0027 0.0086 0.0052 0.0079 0.0062 0.0038;
0.0023 0.0096 0.0097 0.0058 0.0042 0.0063 0.0062 0.0029;
0.0062 0.0074 0.0035 0.0048 0.0063 0.0063 0.0106 0.0029;
0.0039 0.0022 0.0044 0.0038 0.0031 0.0134 0.0009 0.0048;
0.0016 0.0037 0.0053 0.0019 0.0073 0.0032 0.0035 0.0048];

end;

%f = f(:,Ind_dias)
% Exemplo pequeno
%f=[0.3 0.8;
%    0.7 0.2];
% periodo=10;
% 'x' contém os nós
x = [0:periodo:24] ;
zi=[0:0.05:24];
%x = [ 0; 10; 20 ] ;
%zi=[0:0.05:20];

else
%%%%%% Programa para a Lei Exponencial %%%%%%%%%%%%%%%
%lambda = 1 ;
% Gerar 1000 numeros aleatórios a partir da distribuição exponencial com
% parametro média
% NOTA: X segue uma lei exponencial de parametro lambda, entao E(X)=1/lambda
%A = exprnd (1/lambda,1000,1 )
MatrizA
% Com os números gerados criar uma matriz de 100 linhas (observações) por
% 10 colunas (dias)
j=1;
```

```
AA=[];
    t=1;
while j<=10
    i=1;
    while i<=100
        AA(i,j)=A(t,1);
        i=i+1;
        t=t+1;
    end;
    j=j+1;
end;
AA
periodo = 0.5 ;
x = [0.25:periodo:10] ;
% Distribuir os elementos de AA(:,J) por cada 'bin' de x -> 'f'
% E modificar a matriz 'f' de modo que sum(f(:,j))=1
f=[];
j=1;
while j<=10
    f(:,j) = hist( AA(:,j), x )' ;
    f(:,j) = f(:,j)/sum(f(:,j)) ;
    j=j+1;
end;
if ( 0==1 )
    f=[];
    j=1;
    while j<=20
        f(j) = exp(-(j-1)*periodo) - exp(-j*periodo) ;
        j=j+1;
    end;
    f = f'
end
x = [0:periodo:10] ;
zi = [0:0.05:10] ;
```

Apêndice B Listagem de programas

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
end;
% 'n' é o número de bins
% 'm' é o número de dias
[n,m] = size(f);
% Construimos o vector dos hi's
hh = zeros(n,1) ;
for p=1:n
    hh(p) = x(p+1) - x(p) ;
end;
% 1. Construimos a matriz A e os vectores-coluna b e d
%
% 1.2 Construir as matriz A_i e vectores e acumula-las em A; construir tambie;m b e c.
%
% 'S' contem a matriz A
% 'Q' contem o vector-coluna b
% 'd' contem o vector-coluna d
% 'K' contem a constante c
S = zeros(2*(n+1),2*(n+1)) ;
Q = zeros(1,2*(n+1)) ;
K = 0 ;
d = zeros(1,2*(n+1)) ;
for i=2:(n+1)
    h = hh(i-1) ;
    h6 = h*h*h*h*h ;
    h7 = h*h*h ;
    T=zeros(2*(n+1),2*(n+1));
    T(i-1,i-1) = 2*h6/945 ;
    T(i ,i ) = 2*h6/945 ;
    T(i-1,i ) = 31*h6/15120 ;
    T(i ,i-1) = 31*h6/15120 ;
    T(i-1,(n+1)+(i-1)) = -h7/45 ;
    T(i-1,(n+1)+(i )) = -7*h7/360 ;
    T(i ,(n+1)+(i-1)) = -7*h7/360 ;
end;
```

```

T(i , (n+1)+(i )) = -h7/45 ;
T((n+1)+(i-1), i-1) = -h7/45 ;
T((n+1)+(i-1), i ) = -7*h7/360 ;
T((n+1)+(i ), i-1) = -7*h7/360 ;
T((n+1)+(i ), i ) = -h7/45 ;
T((n+1)+(i-1), (n+1)+(i-1)) = h/3;
T((n+1)+(i-1), (n+1)+(i )) = h/6;
T((n+1)+(i ), (n+1)+(i-1)) = h/6;
T((n+1)+(i ), (n+1)+(i )) = h/3;
d(1, i-1) = d(1, i-1) + (-h7/24) ;
d(1, i ) = d(1, i ) + (-h7/24) ;
d(1, (n+1)+(i-1)) = d(1, (n+1)+(i-1)) + (h/2) ;
d(1, (n+1)+(i )) = d(1, (n+1)+(i )) + (h/2) ;
j = 1 ;
q=zeros(1,2*(n+1));
const=0;
while j<=m
    const = const + f(i-1, j) ;
    K = K + f(i-1, j)^2/h ;
    j = j+1 ;
end;
const=const/hh(i-1) ;
q(1, i-1) = const*(-h7/24) ;
q(1, i ) = const*(-h7/24) ;
q(1, (n+1)+(i-1)) = const*(h/2);
q(1, (n+1)+(i )) = const*(h/2);
i = i+1 ;
S = S + T ;
Q = Q + q ;
end;
S = 2*m*S;
Q = -2*Q';
d = d';
%S

```

Apêndice B Listagem de programas

```
%Q
%K
%d
% 2. Construimos as matriz B e C, com as condições iniciais
% 'type' caracteriza as condições iniciais para o spline cubico
% Se TYPE=1 SPLINE NATURAL : M_0 = M_n = 0
% Se TYPE=2 SPLINE CÍCLICO 1: m_0 = m_n; M_0=M_n
% Se TYPE=3 SPLINE CICLICO 2: H_0 = H_n; M_0 = M_n
% Se TYPE=4 SPLINE PARA LEIS 1: H_0 = H_n = 0
% Se TYPE=5 SPLINE PARA LEIS 2: H_0 = 1; H_n = 0
h1 = hh(1:end-1) ;
h2 = hh(2:end) ;
h3 = h1+h2 ;
%Matriz B
lixo = 50;
e1 = 2*[lixo; h3; lixo];
B = diag(e1)+diag(hh,-1)+diag(hh,1);
%B
%Matriz C
e1 = 6*[lixo; -(1./h1+1./h2); lixo] ;
e2 = 6./hh;
C = diag(e1)+diag(e2,-1)+diag(e2,1) ;
%C
% Spline Natural: M_0=0; M_n=0
if type==1
    B(1,:) = 0;
    B(1,1) = 1 ;
    B(n+1,:) = 0 ;
    B(n+1,n+1) = 1 ;
    C(1,:) = 0 ;
    C(n+1,:) = 0 ;
end;
% Spline Ciclico 1: M_0=M_n; m_0=m_n ----> só resulta para n>2
if type==2
```

```
B(1,:) = 0;
B(1,1 ) = -hh(1)/3 ;
B(1,2 ) = -hh(1)/6 ;
B(1,n ) = -hh(n)/6 ;
B(1,n+1) = -hh(n)/3 ;
C(1,:) = 0 ;
C(1,1 ) = 1/hh(1) ;
C(1,2 ) = -1/hh(1) ;
C(1,n ) = -1/hh(n) ;
C(1,n+1) = 1/hh(n) ;
B(n+1,:) = 0 ;
B(n+1,1) = 1 ;
B(n+1,n+1) = -1 ;
C(n+1,:) = 0 ;

end;

% Spline Ciclico 2: H_0=H_n; M_0=M_n
if type==3
    B(1,:) = 0 ;
    C(1,:) = 0 ;
    C(1,1) = -1 ;
    C(1,n+1) = 1 ;
    B(n+1,:) = 0 ;
    B(n+1,1) = 1 ;
    B(n+1,n+1) = -1 ;
    C(n+1,:) = 0 ;

end;

% Spline para Leis 1: H_0=0; H_n=0
if type==4
    B(1,:) = 0 ;
    B(n+1,:) = 0 ;
    C(1,:) = 0 ;
    C(n+1,:) = 0 ;
    C(1,1) = 1 ;
    C(n+1,n+1) = 1 ;
```

Apêndice B Listagem de programas

```
end;
% Spline para Leis 2: H_0= const ; H_n=0
if type==5
    B(1,:) = 0 ;
    B(n+1,:) = 0 ;
    C(1,:) = 0 ;
    C(n+1,:) = 0 ;
    C(1,1) = 1 ;
    C(n+1,n+1) = 1 ;
end;
% Spline para Leis 3: H_n=0
if type==6
    B(1,:) = 0 ;
    B(n+1,:) = 0 ;
    C(1,:) = 0 ;
    C(n+1,:) = 0 ;
    C(1,1) = 0 ;
    C(n+1,n+1) = 1 ;
end;
%B
%C
% 3. Construímos o sistema com as condições de optimalidade, obtendo assim
% os H_i e M_i
% Se a matriz B tiver característica completa => B é não singular
if rank(B)==(n+1)
%if (0==1)
    I = eye(n+1);
    Vector = [inv(B)*C; I];
    Matriz_A = Vector'*S*Vector;
    Vector_D = Vector'*d;
    Vector_B = Vector'*Q;
    Matriz_final = [Matriz_A -Vector_D; Vector_D' 0];
    Vector_final = [-Vector_B; 1],
    Vector_Variaveis = Matriz_final\Vector_final,
```

```

    H = Vector_Variaveis(1:n+1);
    M=inv(B)*C*H;
    Variaveis = [M; H];
% Se matriz B não tiver característica completa (car(B)<n+1) => B é
% singular, logo não tem inversa
else
    Vector_BC = [-B'; C'];
    Matriz_Superior_Direita = [Vector_BC -d ];
    Matriz_final = [S Matriz_Superior_Direita; B -C zeros((n+1),(n+1)) zeros(n+1,1)
    if ( type == 5 )
        Vector_final = [-Q; -1; zeros(n,1); 1] ;
    else
        Vector_final = [-Q; zeros(n+1,1); 1] ;
    end
    Vector_Variaveis = Matriz_final\Vector_final
    Vector_Variaveis = pinv(Matriz_final)*Vector_final
    M = Vector_Variaveis(1:n+1);
    H = Vector_Variaveis(n+2: 2*(n+1));
    Variaveis = [M; H];
end;
% 4. Criamos uma matriz com os coeficientes dos splines cúbicos
% Criamos um vector com os (H_i-H_{i-1})/h_i
v=zeros(n,1)
for i=1:n
v(i,1) = (Variaveis(n+2+i,1)-Variaveis(n+2+(i-1),1))./hh(i);
i=i+1;
end
%v
Matriz=zeros(n+1,4);
Matriz(:,1)=Variaveis(n+2:2*(n+1));
Matriz(:,3)=Variaveis(1:(n+1))/2;
Matriz
for m=1:n
    Matriz(m,4) = (Matriz(m+1,3)-Matriz(m,3))/3/hh(m);

```

```
Matriz(m,2) = v(m)-hh(m)/3*(Matriz(m+1,3)+2*Matriz(m,3));
end
%Matriz
% coeficientes por ordem
Matriz = Matriz(1:n, 4:-1:1);
%Matriz
pp = mkpp(x,Matriz); % liga os polinômios nos pontos de quebra (coordenadas de x)
s = ppval(pp,zi); % valor da função spline cúbico nos elementos de zi
% Construção dos histogramas
[n,m] = size(f);
y=[];
j=1;
while j<=m
    y(:,j) = f(:,j)/periodo
    j=j+1;
end;
y
l=1;
w = x-(periodo/2)
j=0;
while j<m
    barr = bar (w(2:end), y(:,j+1), 'FaceColor', 'w', 'EdgeColor', 'k');
    set (barr, 'BarWidth', l*((m-j)/m));
    hold on;
    j=j+1;
end;
plot(zi,s,'b')
```

B.10. Programa densidade_lei_exponencial_medial.txt

O programa densidade_lei_exponencial_medial.txt, implementado em R, traça o gráfico da função densidade teórica da lei exponencial com parâmetro 1 para as observações dados_leiexponencial.m

```
densidade_lei_exponencial_medial.txt
```

```
x<-scan('matrizA.txt')
```

```
y<-dgamma(x,1,,1)
plot(x,y)
```


Bibliografia

- [1] César Carvalho. Aplicação de splines cúbicos à estimação do ritmo de chegadas a um centro de atendimento markoviano. Master's thesis, Departamento de Matemática da Universidade de Coimbra, 2009.
- [2] B. W. Silverman. *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1986.
- [3] M. P. Wand and M. C. Jones. *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London, 1995.