

UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E  
COMPUTAÇÃO CIENTÍFICA  
DEPARTAMENTO DE ESTATÍSTICA

*Estimação de Máxima Verossimilhança  
Penalizada para Funções de Regressão com  
Erros Perpendiculares*

Campinas - SP

2003

UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E  
COMPUTAÇÃO CIENTÍFICA  
DEPARTAMENTO DE ESTATÍSTICA

*Estimação de Máxima Verossimilhança  
Penalizada para Funções de Regressão com  
Erros Perpendiculares*

Clécio da Silva Ferreira

Orientadora: Prof<sup>a</sup> Dra. Nancy Lopes Garcia

Co-orientador: Prof. Dr. Ronaldo Dias

Campinas - SP  
2003

# *Estimação de Máxima Verossimilhança Penalizada para Funções de Regressão com Erros Perpendiculares*

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Clécio da Silva Ferreira e aprovada pela comissão julgadora.

Campinas, 14 de Agosto de 2003

---

Prof<sup>a</sup>. Dra.: Nancy Lopes Garcia  
Orientadora

---

Prof. Dr.: Ronaldo Dias  
Co-orientador

## Banca Examinadora

1. Nancy Lopes Garcia - Orientadora - IMECC/UNICAMP
2. Cláudia Monteiro Peixoto - IME/USP
3. Jesús Enrique Garcia - IMECC/UNICAMP
4. José Norberto Walter Dachs ( Suplente) - IMECC/UNICAMP

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de MESTRE em Estatística.

*À minha querida mãe Delizeth,  
e às memórias de meu pai José Peres,  
de meu tio Olívio e de Osmar ( Japa),  
amigo-irmão de minha família.*

# *Agradecimentos*

*À minha família, pais, irmãos(ãs), sobrinhos(as), primos(as) e a minha esposa Joelma, pelo apoio e incentivo nos momentos mais difíceis.*

*Aos meus orientadores, Prof<sup>a</sup> Dra. Nancy Lopes Garcia e Prof. Dr. Ronaldo Dias, pela paciência e confiança em meu trabalho.*

*Aos amigos, da graduação em Matemática, do mestrado em Estatística, da república em Viçosa e da minha cidade ( Ubá-MG), que, de uma forma ou de outra, contribuíram para o meu sucesso neste trabalho.*

*À Fundação CAPES, pelo apoio financeiro.*

*Aos funcionários da Secretaria de Pós-Graduação do IMECC, pela dedicação durante todo o curso, em especial nas etapas finais deste trabalho.*

*Aos professores da Universidade Federal de Viçosa, Olímpio Hiroshi Miyagaki ( Matemática) e Adair José Regazzi ( Estatística), por me incentivarem a continuar na carreira acadêmica.*

*Muito obrigado*

*Clécio da Silva Ferreira*

# *Resumo*

*Seja  $f : [a, b] \rightarrow \mathbb{R}$  uma função suave ( com quadrado da segunda derivada integrável) e considere  $N$  um processo pontual de Poisson homogêneo sobre o gráfico de  $f$ . Construa um novo processo de Poisson  $N'$  como sendo os pontos de  $N$  deslocados perpendicularmente à curva  $f$  por variáveis aleatórias iid normalmente distribuídas com média 0 e variância constante.*

*O objetivo deste trabalho é estimar  $f$  através de Estimação de Máxima Verossimilhança Penalizada (EMVP) quando se observa o processo  $N'$ . A função de verossimilhança é obtida a partir da função intensidade de processos de Poisson.*

# *Abstract*

*Let  $f : [a, b] \rightarrow \mathbb{R}$  be a smooth function ( second derivate are square integrable) and consider  $N$  to be an  $\alpha$ - homogeneous Poisson process on  $(f)$ . For each point of  $N$  dislocate it perpendicularly to  $\text{Graf}(f)$  by a normal random variable with zero mean and constant variance. The new process  $N'$  is an inhomogeneous Poisson process with intensity depending on  $f$ . We estimate  $f$  using a spline approach to penalized maximum likelihood.*

# *Lista de Figuras*

1	<i>Retas <math>g_{X_n}</math>, onde <math>(X_n, f(X_n))</math> são os pontos do processo <math>N</math>. . . . .</i>	p.11
2	<i>Função <math>h_u(x)</math> para <math>f(x) = \cos(x)</math>, <math>x \in (0, \pi)</math> para os valores <math>u = -2, 5, 2</math>. .</i>	p.14
3	<i><math>x(u, v)</math> para a reta: <math>P</math> . . . . .</i>	p.25
4	<i>Medida Média para <math>f(x) = \cos(x)</math> . . . . .</i>	p.27
5	<i>Intensidade para <math>f(x) = \cos(x)</math> . . . . .</i>	p.27
6	<i>Função <math>h_u(x)</math> para <math>f(x) = \cos(x)</math>, <math>x \in (0, \pi)</math>. . . . .</i>	p.41
7	<i>Gráficos do processo <math>N'</math> (<math>\sigma = 0.5</math>, <math>\alpha = 4</math>, <math>n = 47</math>) e das curvas estimadas, por Mínimos Quadrados, pelo Modelo Paramétrico e por Máxima Verossimilhança para a curva <math>f(x) = \tan(\pi/4)</math>, <math>x \in (0, 10)</math>. . . . .</i>	p.43
8	<i>Gráficos do processo <math>N'</math> (<math>\sigma = 0.01</math>, <math>\alpha = 12</math>, <math>n = 62</math>) e das curvas estimadas, por Mínimos Quadrados e por Máxima Verossimilhança para a curva <math>f(x) = \cos(4x)</math>, <math>x \in (1.2, 2.8)</math>. . . . .</i>	p.44
9	<i>Gráficos do processo <math>N'</math> (<math>\sigma = 0.07</math>, <math>\alpha = 17</math>, <math>n = 86</math>) e das curvas estimadas, por Mínimos Quadrados e por Máxima Verossimilhança para a curva <math>f(x) = \cos(4\pi x) \exp(-x^2/2)</math>, <math>x \in (0.66, 1.6)</math>. . . . .</i>	p.45
10	<i>Gráficos do processo <math>N'</math> (<math>\sigma = 0.04</math>, <math>\alpha = 5</math>, <math>n = 57</math>) e das curvas estimadas, por Mínimos Quadrados e por Máxima Verossimilhança para a curva <math>f(x) = 2 - 5x + 5 \exp(-100(x - 0.5)^2)</math>, <math>x \in (0, 1)</math>. . . . .</i>	p.46
11	<i>Gráficos do processo <math>N'</math> (<math>\sigma = 0.03</math>, <math>\alpha = 8</math>, <math>n = 97</math>) e das curvas estimadas, por Mínimos Quadrados e por Máxima Verossimilhança para a curva <math>f(x) = 3 * (\sin(2 * \pi * x.^3))</math>, <math>x \in (0.2, 1)</math>. . . . .</i>	p.47
12	<i>Gráficos das curvas estimadas, por Mínimos Quadrados e por Máxima Verossimilhança para o conjunto de dados de terremotos de Magnitude maior que 5 ocorridos na Califórnia, de 1980 a 1999. . . . .</i>	p.49



# Sumário

<b>1</b>	<b>Introdução</b>	p. 1
<b>2</b>	<b>Estimação de Máxima Verossimilhança para Processos de Poisson</b>	p. 4
2.1	<i>Processos de Poisson</i> . . . . .	p. 4
2.2	<i>Medida Média e Intensidade do Processo de Poisson</i> . . . . .	p. 7
2.2.1	<i>Exemplos</i> . . . . .	p. 24
<b>3</b>	<b>Estimação Não Paramétrica de Curvas</b>	p. 28
3.1	<i>Introdução</i> . . . . .	p. 28
3.2	<i>Splines</i> . . . . .	p. 29
3.3	<i>H-splines</i> . . . . .	p. 30
3.4	<i>Estimação de Máxima Verossimilhança da função <math>f</math></i> . . . . .	p. 32
3.5	<i>Regressão Não Paramétrica por Splines</i> . . . . .	p. 33
<b>4</b>	<b>Simulações</b>	p. 36
4.1	<i>Exemplos</i> . . . . .	p. 41
4.1.1	<i>Estimação de falhas geológicas</i> . . . . .	p. 48
4.2	<i>Conclusões</i> . . . . .	p. 50
	<b>Apêndice A - Splines</b>	p. 51
A.1	<i>Derivadas de B-splines</i> . . . . .	p. 53
	<b>Apêndice B - Validação cruzada</b>	p. 55
	<b>Apêndice C - Método de Monte Carlo</b>	p. 57

<b>Apêndice D – Processos Pontuais Univariados</b>	p. 59
<i>D.1 Processos Pontuais de Poisson</i> . . . . .	p. 61
<i>D.2 Processos de Poisson não Homogêneos</i> . . . . .	p. 62
<b>Apêndice E – Estimação de Máxima Verossimilhança Penalizada</b>	p. 64
<b>Apêndice F – Espaço de Funções</b>	p. 66
<b>Apêndice G – Programas</b>	p. 68
<b>Referências Bibliográficas</b>	p. 80

# 1 Introdução

Considere a seguinte situação hipotética: Um avião passa segundo uma trajetória lançando panfletos sobre uma região. Suponha que os panfletos caem perpendicularmente ao local onde foram lançados a uma distância normalmente distribuída. Através da localização dos panfletos pretende-se, neste trabalho, estimar a trajetória do avião. Este problema foi estudado por Garcia (1995) no caso de uma trajetória linear.

Neste caso, tem-se como amostra um conjunto de dados  $\{(X'_i, Y'_i), i = 1, \dots, n\}$  que forma um processo espacial. A verossimilhança de um processo pontual de Poisson é obtida a partir da função intensidade  $\mu_f$  (ver Daley e Vere-Jones (1988)),

$$L_A(\mathbf{X}', \mathbf{Y}', f) = \exp \left\{ \sum_{i=1}^{N(A)} \log \mu_f(X'_i, Y'_i) - \int_A (\mu_f(x, y) - 1) dx dy \right\}. \quad (1.1)$$

A partir da verossimilhança do processo pontual, pode-se fazer uso do método de Estimaco de Mxima Verossimilhança Penalizada (EMVP) para estimar a trajetria do avião.

Estimaco de funoes intensidade de um Processo de Poisson em  $\mathbb{R}$  no  um assunto muito novo. Considerando um Processo de Poisson no-homogneo, Kutoyants (1979) estudou a estimaco paramtrica da funao intensidade quando esta depende de um parmetro real unidimensional, isto ,  $\mu(t) = \mu(t, \theta)$ ,  $t \in [0, T]$ ,  $\theta \in \Theta \subset \mathbb{R}$ . Ramlau-Ransen (1983) propuseram uma estimativa no paramtrica da funao intensidade atravs de funoes Kernel  $K_h$  dada por

$$\hat{\mu}(x; h) := \sum_{i=1}^N K_h(x - X_i)$$

onde  $X_i$ ,  $i = 1, \dots, n$  so observacoes ordenadas em um intervalo fixo  $[0, T]$ . Outro trabalho nesse sentido  encontrado em Gomes (1991), que utiliza o mtodo de Grenander para achar a estimativa da funao intensidade. Olivares (1993) introduz o mtodo de validaco cruzada para determinar o valor timo da janela  $h$  do estimador de Kernel da funao intensidade.

Muitos estudos de intensidade de Processos de Poisson em  $[0, \infty)$  surgem na análise de ocorrências de terremotos. Ogata (1988) discute várias classes de modelos estocásticos para a origem do tempo e a magnitude de terremotos. A função de verossimilhança é dada por

$$\log L(\theta) = \sum_{i=1}^N \log \mu(t_i; \theta) - \int_0^T \mu(t; \theta) dt,$$

onde  $\mu(t; \theta)$  é a intensidade condicional parametrizada e  $\{t_i\}$  é o conjunto de tempos de ocorrência de terremotos em um determinado intervalo de tempo  $[0, T]$ . As magnitudes  $\{m_i\}$  já estão incluídas na intensidade condicional.

No caso de Processos de Poisson em  $\mathbb{R} \times [0, T]$ , Choi e Hall (1999) propõem estimativas da função intensidade de um processo de Poisson espaço-tempo dadas por

$$\hat{\mu}(x, t) = \frac{1}{h_1 h_2} \sum_i K_1 \left( \frac{X_i - x}{h_1} \right) K_1 \left( \frac{T_i - t}{h_2} \right),$$

onde  $K_1$  é um Kernel univariado e  $h_1$  é a janela para o componente espaço e  $h_2$  a janela para o componente tempo.  $X_i$  denota a projeção da  $i$ -ésima observação sobre o eixo especificado  $\mathcal{A}$  e  $T_i$  é o tempo em que a observação ocorreu.

Em dados de ecologia, Rathbun (1996) faz estimação paramétrica de intensidade de Poisson espacial usando variáveis observadas concomitantes parcialmente. O processo modulado é Poisson cuja log intensidade é uma função linear de variáveis espaciais concomitantes.

Entretanto, neste problema o objetivo não é estimar a intensidade  $\mu$  do processo, mas sim estimar a função  $f$ . Devido à complexa relação entre a intensidade do processo e  $f$ , não é possível obter-se uma estimativa de  $f$  através de uma estimativa de  $\mu_f$ .

Mais ainda, o conjunto  $(X'_i, Y'_i)$  pode ser visto como um modelo de regressão com erros em ambas as variáveis e estes são correlacionados. À primeira vista, um modelo de regressão simples poderia ser ajustado. Uma das questões a serem respondidas neste trabalho é: “quão errônea é esta abordagem?”.

Esta dissertação está organizada da seguinte forma: No Capítulo 2, são apresentadas algumas definições e resultados sobre processos de Poisson. Em seguida, determina-se a medida média e a intensidade do processo de Poisson obtido pelas posições dos panfletos. Exemplos da medida média e intensidade para o caso onde a trajetória do avião é uma reta e no caso onde a trajetória segue a curva  $y = \cos(x)$  são expostos.

O Capítulo 3 é dedicado ao método de estimação não paramétrica por Splines. Uma seção sobre Splines, incluindo Splines Híbridos ( $H$ -splines) é exibida. Tanto Estimação de Máxima Verossimilhança Penalizada quanto regressão por mínimos quadrados via  $H$ -Splines

são abordadas, apresentando uma solução analítica para o estimador da curva que representa a trajetória do avião.

O Capítulo 4 apresenta algumas simulações utilizando-se os dois métodos apresentados nos Capítulos 2 e 3. Algoritmos para geração do processo pontual e de construção dos nós que serão utilizados para construir a estimativa da trajetória, baseada em combinação linear de B-splines são mostrados. Um algoritmo alternativo ao utilizado em H-splines para mínimos quadrados é apresentado para utilização na EMVP. Gráficos das curvas estimadas pelos dois métodos são esboçados como forma de comparação gráfica. Os erros quadráticos médios dos dois métodos também são mostrados. Também é apresentado um ajuste para dados de terremotos de magnitude acima de 5 na escala Ritche ocorridos na Califórnia entre 1980 e 1999. Por último, são feitas conclusões sobre os dois métodos utilizados na estimação da trajetória do avião.

Finalmente os apêndices contem uma descrição mais detalhada sobre teoria de Splines, Validação Cruzada, método de Monte Carlo para integração de funções multivariadas, apresentação geral de processos pontuais e alguns resultados sobre processos de Poisson, Estimação de Máxima Verossimilhança Penalizada e algumas teorias sobre espaços de funções.

## 2 *Estimação de Máxima Verossimilhança para Processos de Poisson*

### 2.1 Processos de Poisson

*Processos de Poisson são os processos pontuais mais usados na prática. Não somente eles modelam diversos fenômenos na natureza como são a base para construção de processos mais complexos. Esta seção apresenta algumas definições e resultados sobre processos de Poisson. Processos univariados são vistos com mais detalhes no Apêndice D. Uma referência mais completa sobre processos pontuais em geral pode ser encontrada em Daley e Vere-Jones (1988).*

*Seja  $E$  um subconjunto de um espaço Euclidiano ( um subconjunto de  $[0, \infty)$ ,  $\mathbb{R}$  ou de  $\mathbb{R}^d$ ,  $d \geq 1$ ). Suponha que  $\{X_n, n \geq 0\}$  são elementos aleatórios de  $E$  representando pontos aleatórios no espaço de estado  $E$ .*

**Definição 2.1.** *Define-se a medida discreta*

$$\mathbb{I}_{X_n}(A) := \begin{cases} 1, & \text{se } X_n \in A, \\ 0, & \text{se } X_n \notin A, \end{cases} \quad (2.1)$$

*denominada de função indicadora do conjunto  $A$ .*

**Definição 2.2.** *Defina uma medida de contagem  $N$  por*

$$N(\cdot) := \sum_n \mathbb{I}_{X_n}(\cdot). \quad (2.2)$$

*Assim*

$$N(A) := \sum_n \mathbb{I}_{X_n}(A) \quad (2.3)$$

*é o número aleatório de pontos que pertencem ao conjunto  $A$ .  $N$  é chamado um processo*

pontual e  $\{X_n\}$  são chamados os pontos do processo  $N$ .

**Definição 2.3.** Define-se a medida média do processo  $N$  por

$$\Lambda(A) := \mathbb{E}(N(A)), \quad A \in \Phi, \quad (2.4)$$

onde  $\mathbb{E}$  denota esperança matemática e  $\Phi$  é a  $\sigma$ -álgebra de  $E$ , gerada pelos Borelianos de  $E$ . Assim,  $\Lambda(A)$  é o número esperado de pontos de  $N$  que pertencem ao conjunto  $A$ .

**Definição 2.4.** Sejam  $N$  um processo pontual com espaço de estado  $E$  e  $\Phi$  uma  $\sigma$ -álgebra de  $E$ . Então  $N$  é um processo de Poisson com medida média  $\Lambda(\cdot)$  se

1. Para  $A \in \Phi$

$$\mathbb{P}[N(A) = k] = \begin{cases} \frac{e^{-\Lambda(A)}(\Lambda(A))^k}{k!}, & \text{se } \Lambda(A) < \infty \\ 0 & \text{se } \Lambda(A) = \infty. \end{cases}$$

2. Se  $A_1, \dots, A_k$  são subconjuntos disjuntos pertencentes a  $\Phi$ , então  $N(A_1), \dots, N(A_k)$  são variáveis aleatórias (v.a's) independentes.

Assim  $N$  é processo de Poisson (PP) se o número aleatório de pontos no subconjunto  $A$  tem distribuição de Poisson com parâmetro  $\Lambda(A)$  e o número de pontos em regiões disjuntas são v.a's independentes.

Notação:  $N \sim PP(E, \Lambda)$ .

Quando  $E = \mathbb{R}$ , a Propriedade 2 é chamada de incrementos independentes, pois tem-se que para quaisquer  $t_1 < t_2 < \dots < t_k$ ,  $\{N((t_i, t_{i+1}]), i = 1, \dots, k-1\}$  são v.a's independentes. Quando a medida média é um múltiplo da medida de Lebesgue, isto é, existe um parâmetro  $\alpha > 0$  tal que para todo  $A$ , tem-se que  $N(A)$  tem distribuição de Poisson com média  $\mathbb{E}(N(A)) = \alpha|A|$ , onde  $|A|$  é a medida de Lebesgue de  $A$ ,  $N$  é chamado de Processo de Poisson homogêneo com taxa  $\alpha$ . Quando a medida média não é um múltiplo da medida de Lebesgue de  $A$ , o processo  $N$  é chamado de processo de Poisson não-homogêneo.

Uma outra caracterização de PP homogêneo é dada pelo seguinte teorema.

**Teorema 2.1.** Um processo  $N = \sum_n \mathbb{I}_{X_n}$  é PP homogêneo com taxa  $\alpha$  se, e somente se:

- i)  $N(A) \sim P(\alpha|A|)$ ;
- ii) Dado  $N(A) = n$  então os pontos  $X_1, \dots, X_n$  são iid  $U(A)$ .

Mais detalhes ver Apêndice D.

**Definição 2.5.** Define-se a intensidade do processo de Poisson  $N$  com medida média  $\Lambda$  absolutamente contínua com respeito à medida de Lebesgue como sendo uma função  $\mu$  que satisfaz a relação

$$\Lambda(A) := \iint_A \mu(x, y) dx dy. \quad (2.5)$$

Um resultado importante é que toda transformação determinística de um Processo de Poisson é também um Processo de Poisson. Formalmente, considere uma transformação

$$T : E \mapsto E',$$

onde  $E$  e  $E'$  são dois espaços Euclidianos. Dado  $A' \subset E'$ , seja a imagem inversa de  $A'$  sob  $T$

$$T^{-1}(A') = \{e \in E : T(e) \in A'\}.$$

**Proposição 2.1.** Se  $N$  é um processo de Poisson com medida média  $\Lambda$  em  $E$  com pontos  $W_n$ , então  $N' := N \circ T^{-1}$  é um Processo de Poisson em  $E'$  com pontos  $\{T(W_n)\}$  e medida média  $\Lambda' := \Lambda \circ T^{-1}$ .

**Proposição 2.2.** Suponha que  $\{X_n\}$  sejam elementos aleatórios de um espaço Euclidiano  $E_1$  tais que

$$\sum_n \mathbb{I}_{X_n}$$

é um processo de Poisson com medida média  $\Lambda$ . Suponha que  $\{J_n\}$  são elementos aleatórios de um segundo espaço Euclidiano  $E_2$  com distribuição de probabilidade  $F$  e suponha o processo de Poisson e a seqüência  $\{J_n\}$  são definidos no mesmo espaço de probabilidade e são independentes. Então o processo pontual

$$\sum_n \mathbb{I}_{(X_n, J_n)}$$

em  $E_1 \times E_2$  é um processo de Poisson com medida média  $\Lambda \times F$ , significando que se  $A_i \subset E_i$ ,  $i = 1, 2$ , então

$$(\Lambda \times F)(A_1 \times A_2) = (\Lambda \times F)(\{(e_1, e_2) : e_1 \in A_1, e_2 \in A_2\}) = \Lambda(A_1)F(A_2). \quad (2.6)$$

As provas das Proposições 2.2 e 2.1 podem ser encontradas em Resnick (1992).

O seguinte resultado é bem conhecido (ver Bustos e Orgambide (1992)) e será utilizado na geração de PP não homogêneo a partir de um PP homogêneo.

**Teorema 2.2.** (Teorema da Inversão) Seja  $F : \mathbb{R} \rightarrow [0, 1]$  uma função de distribuição acumulada;  $F^{-}(t) : [0, 1] \rightarrow \mathbb{R}$  definida por  $F^{-}(t) = \inf\{x : F(x) \geq t\}$ . Se  $U \sim U(0, 1)$  e  $X = F^{-}(U)$ , então  $F$  é a função de distribuição da variável aleatória  $X$ .



## 2.2 Medida Média e Intensidade do Processo de Poisson

O problema descrito na Introdução pode ser modelado através de um processo de Poisson não homogêneo. Considere a trajetória do avião no plano como o gráfico de uma função suave  $f$  e a localização dos panfletos em  $\mathbb{R}^2$  como um processo pontual  $N' = \{(X'_i, Y'_i); i = 1, 2, \dots\}$ . Suponha que o processo  $N'$  é observado somente em uma faixa  $\mathbf{R} = \{(x, y); a \leq x \leq b\}$ .

O processo  $N'$  pode ser construído através de um processo pontual de Poisson homogêneo com intensidade  $\alpha$  no conjunto

$$\text{Graf}(f) = \{(x, y) \in \mathbf{R}; y = f(x)\}.$$

Chame este processo de  $N$ .

**Teorema 2.3.** O processo  $N$

$$N([a, b] \times [c, d]) = \sum_{i=1}^L \mathbb{I}_{(X_n, f(X_n))}([a, b] \times [c, d]), \quad (2.7)$$

é Processo de Poisson homogêneo em  $\text{Graf}(f)$  com taxa  $\alpha$  se, e somente se,  $L \sim P(\alpha C)$  e dado  $L = n$ ,  $X_1, X_2, \dots, X_n$  são v.a's iid com função de distribuição acumulada (fda) dada por

$$F(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{1}{C} \int_a^x \sqrt{1 + f'(t)^2} dt & \text{se } a \leq x < b \\ 1 & \text{se } x \geq b \end{cases} \quad (2.8)$$

onde  $C = \int_a^b \sqrt{1 + f'(t)^2} dt$  é o comprimento da curva  $f$  no intervalo  $[a, b]$ .

**Prova:**

Dado que  $N$  é um  $PP(\text{Graf}(f), \alpha t)$ , pelo Teorema 2.1,  $L$  tem distribuição de Poisson com média  $\alpha C$ . Dado  $L = n$ , sejam  $X_1, \dots, X_n$  tais que  $(X_1, f(X_1)), \dots, (X_n, f(X_n))$  são os pontos de ocorrência de  $N$  em  $\text{Graf}(f)$ , isto é, dado  $L = n$ ,  $(X_1, f(X_1)), \dots, (X_n, f(X_n))$  são iid  $U(\text{Graf}(f))$ . Defina

$$H(x) = \int_a^x \sqrt{1 + f'(t)^2} dt. \quad (2.9)$$

Portanto  $H(X_1), \dots, H(X_n)$  são iid  $U(0, C)$ , pelo Teorema 2.1.

Fazendo  $U_i = \frac{H(X_i)}{C} \sim U(0, 1)$ ,  $i = 1, \dots, n$ , tem-se que  $U_1, \dots, U_n$  são iid  $U(0, 1)$ .

Sendo  $F(b) = 1$ , tem-se que  $0 \leq F(x) \leq 1$ ,  $F(x)$  é contínua. Logo,  $F(x)$  é uma função distribuição acumulada (fda) e  $(\sqrt{1 + f'(t)^2}/C)\mathbb{I}_t([a, b])$  é uma função densidade de probabilidade (fdp). Assim

$$\begin{aligned} X_i &= \inf \left\{ x : \int_a^x \sqrt{1 + f'(t)^2} dt \geq U_i C \right\} \\ &= \inf \left\{ x : \frac{1}{C} \int_a^x \sqrt{1 + f'(t)^2} dt \geq U_i \right\} \\ &= F^{-}(U_i). \end{aligned} \tag{2.10}$$

Portanto, pelo Teorema 2.2,  $X_i \sim F$ .

Reciprocamente, sendo  $L \sim P(\alpha C)$  e dado que  $L = n$ ,  $X_1, X_2, \dots, X_n \sim F$ , sejam as v.a.'s

$$U_i = H(X_i) = \int_a^{X_i} \sqrt{1 + f'(t)^2} dt, \quad a \leq X_i \leq b, \quad i = 1, \dots, n.$$

Assim, para  $i = 1, \dots, n$ , tem-se que

$$\begin{aligned} F_{U_i}(x) &= P(U_i \leq x) \\ &= P(H(X_i) \leq x) \\ &= P(X_i \leq H^{-}(x)) \\ &= F_{X_i}(H^{-}(x)) \\ &= F_{X_i}(F^{-}(x/C)) \end{aligned}$$

pois

$$H^{-}(x) = \inf \left\{ y : \int_a^y \sqrt{1 + f'(t)^2} dt \geq x \right\} = F^{-}(x/C).$$

Portanto,  $U_i = H(X_i)$  são independentes, pois  $X_i$  são independentes e  $U_i \sim U(0, C)$ . ■

Seja  $\{Z_n, n \geq 1\}$  uma coleção de v.a.'s iid normalmente distribuídas com média 0 e variância  $\sigma^2$  conhecida. Considere a seguinte transformação

$$(X'_n, Y'_n) = (X_n, f(X_n)) + (-f'(X_n), 1) \frac{Z_n}{\sqrt{1 + f'(X_n)^2}}, \quad (2.11)$$

onde  $\{X_n, f(X_n)\}$  formam o processo  $N$ .

Seja  $N'$  o processo pontual

$$N'(\cdot) := \sum_n \mathbb{I}_{(X'_n, Y'_n)}(\cdot). \quad (2.12)$$

**Teorema 2.4.** *O processo pontual  $N'$  definido por (2.12) é um processo de Poisson não-homogêneo em  $\mathbb{R}^2$  com medida média  $\Lambda_f$  dada por*

$$\begin{aligned} \Lambda_f(u, v) &:= \Lambda_f((-\infty, u] \times (-\infty, v]) \\ &= \frac{\alpha}{\sqrt{2\pi}\sigma} \int_a^b \sqrt{1 + f'(x)^2} \left[ \int_{-\infty}^{\min(Z_1(u, x), Z_2(v, x))} e^{-z^2/2\sigma^2} dz \mathbb{I}_{\{f'(x) < 0\}} + \right. \\ &\quad \left. + \int_{Z_1(u, x)}^{Z_2(v, x)} e^{-z^2/2\sigma^2} dz \mathbb{I}_{\{f'(x) > 0\}} \mathbb{I}_{\{Z_1(u, x) < Z_2(v, x)\}} \right] dx, \quad u, v \in \mathbb{R} \end{aligned} \quad (2.13)$$

onde

$$Z_1(u, x) = \frac{x - u}{f'(x)} \sqrt{1 + f'(x)^2} \quad (2.14)$$

e

$$Z_2(v, x) = (v - f(x)) \sqrt{1 + f'(x)^2}. \quad (2.15)$$

**Prova:** Sendo  $Z_i \sim N(0, \sigma^2)$  e  $f_{X_i}(x) = \frac{\sqrt{1+f'(x)^2}}{C} \mathbb{I}_x[a, b]$ , pela Proposição 2.2,

$$F_Z(A_2) = \int_{A_2} \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2/2\sigma^2} dz$$

e

$$N_1(\cdot) = \sum_n \mathbb{I}_{(X_n, Z_n)}(\cdot)$$

é um processo de Poisson com medida média

$$\Lambda_1(A_1, A_2) = \alpha \int_{A_1} \sqrt{1 + f'(t)^2} dt \int_{A_2} \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2/2\sigma^2} dz, \quad A_1, A_2 \in \Phi,$$

onde  $\Phi$  é a  $\sigma$ -álgebra gerada pelos Borelianos de  $\mathbb{R}$ . Considere a seguinte transformação

$$T(X_n, Z_n) = (X_n, f(X_n)) + \frac{(-f'(X_n), 1)}{\sqrt{1 + f'(X_n)^2}} Z_n. \quad (2.16)$$

Note que

$$N'(A_1 \times A_2) = \sum_n \mathbb{I}_{T(X_n, Z_n)}(A_1 \times A_2), \quad (2.17)$$

e portanto,  $N'$  é um processo pontual de Poisson.

Para determinar a medida média de  $N'$ , considere a família de retângulos

$$\Phi = \{A(u, v) = (-\infty, u] \times (-\infty, v], u, v \in \mathbb{R}\},$$

a qual forma um anel em  $\mathbb{R}^2$ , que gera a  $\sigma$ -álgebra dos Borelianos em  $\mathbb{R}^2$ .

A imagem inversa de  $A(u, v)$  sob  $T$  é:

$$T^{-1}(A(u, v)) = \{x \in [a, b], z \in \mathbb{R} : T(x, z) \in A(u, v)\}.$$

Dado  $x \in [a, b]$ , considere a reta

$$g_x(t) = f(x) - \frac{1}{f'(x)}(t - x), \quad (2.18)$$

ou seja, a reta normal à curva  $f$  no ponto  $(x, f(x))$ . Assim, se  $X_n$  é um ponto do processo pontual  $N$ , então o possível ponto  $(X'_n, Y'_n)$  correspondente do processo  $N'$  estará sobre a reta  $g_{X_n}(t)$  (veja Figura 1).

De acordo com (2.11),

$$Z_n^2 = (Y'_n - f(X_n))^2 + (X'_n - X_n)^2,$$

ou seja,  $Z_n$  é a distância de  $(X'_n, Y'_n)$  à  $(X_n, f(X_n))$ , na direção normal à curva  $f$  no ponto  $(x, f(x))$ .

O ponto de interseção da reta  $g_x(t)$  com o eixo vertical  $t = u$  é  $(u, f(x) - \frac{u-x}{f'(x)})$  e o ponto de interseção da reta  $g_x(t)$  com o eixo horizontal  $y = v$  é  $(x + f'(x)(f(x) - v), v)$ . Assim,  $Z_1(u, x)$ , dado em (2.14), é a distância do ponto  $(x, f(x))$  ao ponto  $(u, f(x) - \frac{u-x}{f'(x)})$  e  $Z_2(v, x)$ , dado em (2.15), é a distância do ponto  $(x, f(x))$  ao ponto  $(x + f'(x)(f(x) - v), v)$ .

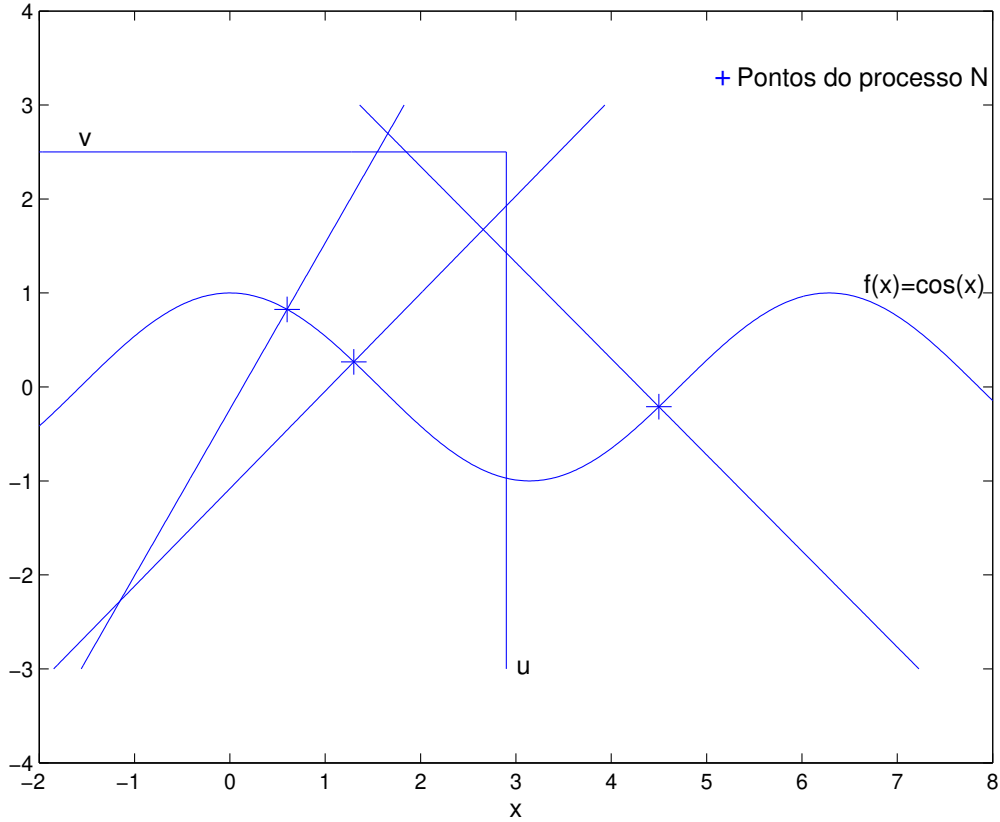


Figura 1: Retas  $g_{X_n}$ , onde  $(X_n, f(X_n))$  são os pontos do processo  $N$ .

Note que as “distâncias” aqui são positivas no sentido positivo da reta normal à curva  $f$  no ponto  $(x, f(x))$  tendo este ponto como a origem e negativas caso contrário.  $Z_1(u, x)$  e  $Z_2(v, x)$  foram obtidos levando isto em consideração. Isto pode ser feito devido ao fato que  $Z$  tem distribuição normal, simétrica em torno de zero.

Assim, tem-se que

$$\begin{aligned}
 T^{-1}(A(u, v)) &= \{x \in [a, b], z \in \mathbb{R} : T(x, z) \in A(u, v)\} \\
 &= \{x \in [a, b], z \in (-\infty, \min\{Z_1(u, x), Z_2(v, x)\}]\mathbb{I}_{\{f'(x) < 0\}} \cup \\
 &\quad [Z_1(u, x), Z_2(v, x)]\mathbb{I}_{\{Z_1(u, x) < Z_2(v, x)\}}\mathbb{I}_{\{f'(x) > 0\}}\}.
 \end{aligned}$$

Portanto, a medida média do processo  $N'$  é:

$$\begin{aligned} \Lambda_f(u, v) &= \frac{\alpha}{\sqrt{2\pi}\sigma} \int_a^b \sqrt{1 + f'(x)^2} \left[ \int_{-\infty}^{\min(Z_1(u,x), Z_2(v,x))} e^{-z^2/2\sigma^2} dz \mathbb{I}_{\{f'(x) < 0\}} + \right. \\ &\quad \left. + \int_{Z_1(u,x)}^{Z_2(v,x)} e^{-z^2/2\sigma^2} dz \mathbb{I}_{\{f'(x) > 0\}} \mathbb{I}_{\{Z_1(u,x) < Z_2(v,x)\}} \right] dx. \end{aligned} \quad (2.19)$$

■

Somente para ilustração, é exibida outra forma de determinar a medida média do processo  $N'$ , através de esperança condicional.

Seja  $(X_i, f(X_i))$ ,  $i = 1, 2, \dots, L$  uma realização do processo  $N$ . Por hipótese, cada  $(X'_i, Y'_i)$  está na direção normal à curva  $f$  no ponto  $(X_i, f(X_i))$  a uma distância  $Z_i$  deste último ponto. Além disso,  $(X_i, f(X_i))$  é um processo de Poisson homogêneo sobre  $\text{Graf}(f)$ . Sendo  $C$  o comprimento da curva  $f$  no intervalo  $[a, b]$  e  $L$  o número de pontos do processo sobre a curva  $f$ , de acordo com o Teorema 2.3, tem-se que  $L$  tem distribuição de Poisson com média  $\alpha C$ .

Assim,

$$N'(A(u, v)) = \sum_{j=1}^L \mathbb{I}_{(X'_j, Y'_j)}(A(u, v))$$

e

$$\mathbb{E}[N'(A(u, v)) | N] = \sum_{i=1}^L \mathbb{P}((X'_i, Y'_i) \in A(u, v) | N).$$

Dada a observação  $X_i$ , de acordo com a equação (2.18), tem-se que  $g_{X_i}$  é a reta normal à curva  $f$  no ponto  $(X_i, f(X_i))$ .

Seja  $A_i$  a interseção da reta  $g_{X_i}$  com  $A(u, v)$ . Assim,

$$\mathbb{P}((X'_i, Y'_i) \in A(u, v) | N) = \mathbb{P}(Z_i \in A_i | X_i) = \int_{A_i} \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2/2\sigma^2} dz.$$

De acordo com (2.11),  $Z_i^2 = (Y'_i - f(X_i))^2 + (X'_i - X_i)^2$ , ou seja,  $Z_i$  é a distância de  $(X'_i, Y'_i)$  à  $(X_i, f(X_i))$ , na direção normal à curva  $f$  no ponto  $(X_i, f(X_i))$ .

Assim,

$$\begin{aligned} \mathbb{P}((X'_i, Y'_i) \in A(u, v) | N) &= \frac{1}{\sqrt{2\pi}\sigma} \left[ \int_{-\infty}^{\min(Z_1(u, X_i), Z_2(v, X_i))} e^{-z^2/2\sigma^2} dz \mathbb{I}_{\{f'(X_i) < 0\}} + \right. \\ &\quad \left. \int_{Z_1(u, X_i)}^{Z_2(v, X_i)} e^{-z^2/2\sigma^2} dz \mathbb{I}_{\{f'(X_i) > 0\}} \mathbb{I}_{\{Z_1(u, X_i) < Z_2(v, X_i)\}} \right]. \end{aligned} \quad (2.20)$$

Denote por  $p(X_i)$  o lado direito da equação (2.20).

Assim,

$$\mathbb{E}[N'(A(u, v)) | N] = \sum_{i=1}^L \mathbb{P}((X'_i, Y'_i) \in A(u, v) | N) = \sum_{i=1}^L p(X_i).$$

$$\mathbb{E}[N'(A(u, v)) | L] = \mathbb{E} \left[ \sum_{i=1}^L p(X_i) | L \right] = \sum_{i=1}^L \mathbb{E}[p(X_i) | L] = L \int_a^b p(x) \frac{\sqrt{1 + f'(x)^2}}{C} dx,$$

pois, dado  $L$ , as variáveis  $X_i$  são iid.

$$\begin{aligned} \mathbb{E}[N'(A(u, v))] &= \mathbb{E}[L] \int_a^b p(x) \frac{\sqrt{1 + f'(x)^2}}{C} dx \\ &= \alpha C \int_a^b p(x) \frac{\sqrt{1 + f'(x)^2}}{C} dx \\ &= \alpha \int_a^b p(x) \sqrt{1 + f'(x)^2} dx, \end{aligned} \quad (2.21)$$

que é da forma (2.13). ■

As seguintes definições e resultados serão necessários para determinar-se a intensidade  $\mu_f$  do processo a partir da medida média  $\Lambda_f$ .

Sejam  $a_j \in [a, b]$ ,  $j = 1, \dots, J$  os pontos tais que  $f'(a_j) = 0$ . Defina  $A_j = (a_j, a_{j+1}]$ ,  $j = 0, \dots, J$ , onde  $A_0 = [a, a_1]$ ,  $A_J = (a_J, b]$ , os sub-intervalos de  $[a, b]$  onde  $f$  é monótona. Considere

$$h_u(x) := f(x) - \frac{(u - x)}{f'(x)}, \quad (2.22)$$

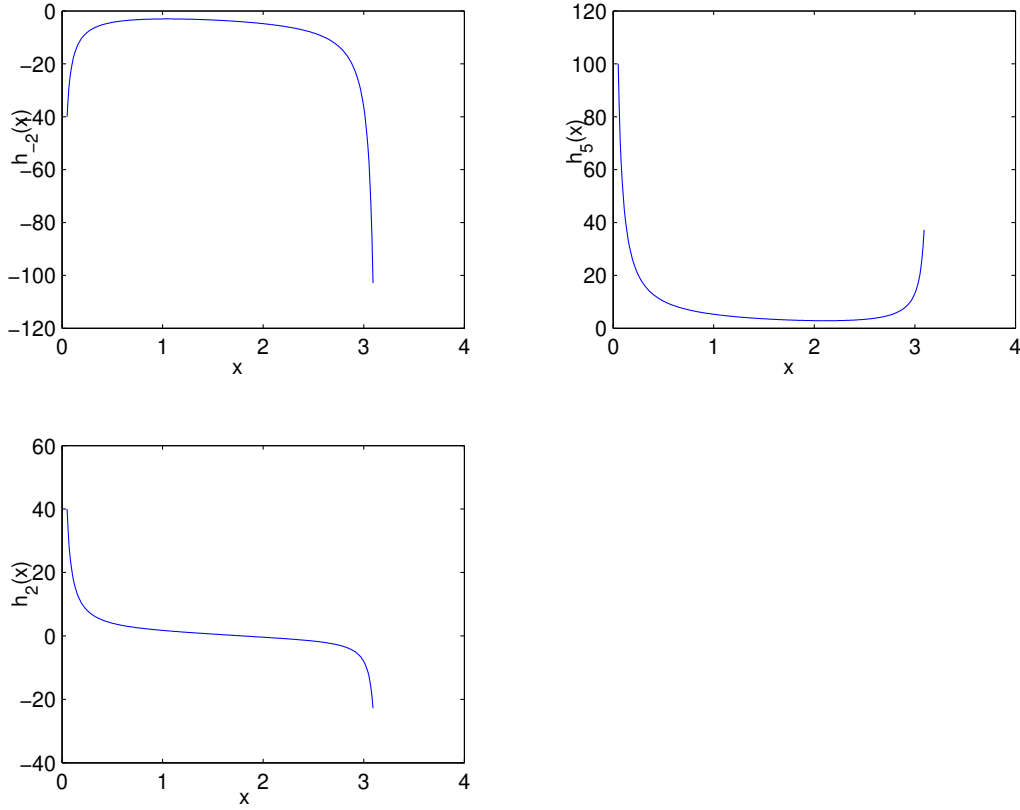


Figura 2: Função  $h_u(x)$  para  $f(x) = \cos(x)$ ,  $x \in (0, \pi)$  para os valores  $u = -2, 5, 2$ .

isto é, a função altura da curva normal à reta no ponto  $(x, f(x))$  relativa ao eixo  $x = u$ . A Figura 2 mostra a função  $h_u$  no subintervalo  $A_j = (0, \pi]$  para valores de  $u$  à esquerda, à direita e no interior do subintervalo  $A_j$ .

Denote por  $x_{j_k}(u, v)$ ,  $k = 1, \dots, K_j$  os valores, pertencentes ao subintervalo  $A_j$ , que satisfazem à equação

$$h_u(x_{j_k}(u, v)) = v. \quad (2.23)$$

Dado o sub-intervalo  $A_j$ , seja  $A_{j_k}(u, v) = (x_{j_k}(u, v), x_{j_{k+1}}(u, v)]$ ,  $k = 1, \dots, K_j - 1$ ,  $A_{j_0}(u, v) = (a_j, x_{j_1}(u, v)]$ ,  $A_{j_{K_j}}(u, v) = (x_{j_{K_j}}(u, v), a_{j+1}]$ , os sub-intervalos onde a função  $h$  é sempre maior que ou sempre menor que  $v$ . Note que o conjunto  $x_{j_k}(u, v)$  pode ser vazio.

**Definição 2.6.** Defina a seguinte função indicadora:

$$\mathbb{I}_{\{h_u(A) < v\}} := \begin{cases} 1, & \text{se } h_u(x) < v, \text{ para todo } x \in A \\ 0, & \text{caso contrário.} \end{cases} \quad (2.24)$$

A função indicadora  $\mathbb{I}_{\{h_u(A) > v\}}$  é definida de forma análoga à Definição 2.6. Também se



utiliza aqui neste trabalho a notação  $\mathbb{I}_{\{x < y\}}$  indicando que esse valor vale 1 se a desigualdade é satisfeita e 0 caso contrário.

Segue abaixo um resultado sobre derivadas de integrais. Mais detalhes podem ser encontrados em *Dudewicz (1988)*.

**Teorema 2.5.** (Teorema de Leibniz) *Se  $f(x, \theta)$  é uma função integrável com respeito à  $x$ , e  $\frac{\partial f(x, \theta)}{\partial \theta}$  é uma função contínua de ambas variáveis  $x$  e  $\theta$ , então*

$$\frac{d}{d\theta} \int_{a(\theta)}^{b(\theta)} f(x, \theta) dx = f(b(\theta), \theta) \frac{d(b(\theta))}{d\theta} - f(a(\theta), \theta) \frac{d(a(\theta))}{d\theta} + \int_{a(\theta)}^{b(\theta)} \frac{\partial f(x, \theta)}{\partial \theta} dx. \quad (2.25)$$

**Teorema 2.6.** *O processo  $N'$  tem intensidade  $\mu_f$  dada por*

$$\begin{aligned} \mu_f(u, v) = & \frac{\alpha}{\sqrt{2\pi}\sigma} \left\{ \sum_{j=1}^{J-1} \left[ \sum_{k=1}^{K_j-1} [q(x_{j_{k+1}}(u, v)) - q(x_{j_k}(u, v))] \times \right. \right. \\ & \left. \left[ \mathbb{I}_{\{h_u(A_{j_k}(u, v)) < v\}} \mathbb{I}_{\{f'(A_j) \geq 0\}} + \mathbb{I}_{\{h_u(A_{j_k}(u, v)) > v\}} \mathbb{I}_{\{f'(A_j) < 0\}} \right] \mathbb{I}_{\{a_j > u\}} + \right. \\ & \left. \left[ q(x_{j_1}(u, v)) - q(x_{j_{K_j}}(u, v)) + \sum_{k=1}^{K_j-1} [q(x_{j_{k+1}}(u, v)) - q(x_{j_k}(u, v))] \times \right. \right. \\ & \left. \left[ \mathbb{I}_{\{h_u(A_{j_k}(u, v)) < v\}} \mathbb{I}_{\{f'(A_j) \geq 0\}} + \mathbb{I}_{\{h_u(A_{j_k}(u, v)) > v\}} \mathbb{I}_{\{f'(A_j) < 0\}} \right] \mathbb{I}_{\{a_{j+1} < u\}} + \right. \\ & \left. \left[ q(x_{j_1}(u, v)) + \sum_{k=1}^{K_j-1} [q(x_{j_{k+1}}(u, v)) - q(x_{j_k}(u, v))] \times \right. \right. \\ & \left. \left. \left. \left[ \mathbb{I}_{\{h_u(A_{j_k}(u, v)) < v\}} \mathbb{I}_{\{f'(A_j) \geq 0\}} + \mathbb{I}_{\{h_u(A_{j_k}(u, v)) > v\}} \mathbb{I}_{\{f'(A_j) < 0\}} \right] \mathbb{I}_{\{a_j < u < a_{j+1}\}} \right] \right\}. \end{aligned} \quad (2.26)$$

onde

$$q(x(u, v)) = \frac{(1 + f'(x(u, v))^2) e^{-(v-f(x(u, v)))^2 (1+f'(x(u, v))^2)/2\sigma^2}}{1 + f'(x(u, v))^2 - f''(x(u, v))(v - f(x(u, v)))}, \quad (2.27)$$

$h_u$  é dada como em (2.22) e  $x_{j_k}(u, v)$  são soluções de (2.23).

**Prova:** De acordo com (2.5), a intensidade do processo de Poisson  $N'$  pode ser obtida como

$$\mu_f(u, v) = \frac{\partial^2 \Lambda_f(u, v)}{\partial v \partial u},$$

onde  $\Lambda_f$  é dada por (2.19).

Seja  $\Lambda_f(u, v)_j$  a parte de  $\Lambda_f(u, v)$  no subintervalo  $A_j$ . Logo

$$\begin{aligned} \Lambda_f(u, v)_j = & \frac{\alpha}{\sqrt{2\pi}\sigma} \left[ \left[ \sum_{k=1}^{K_j-1} \int_{A_{j_k}(u,v)} \sqrt{1+f'(x)^2} \int_{Z_1(u,x)}^{Z_2(v,x)} e^{-z^2/2\sigma^2} dz dx \mathbb{I}_{\{h_u(A_{j_k}(u,v)) < v\}} \mathbb{I}_{\{f'(A_j) \geq 0\}} \right. \right. \\ & + \left[ \int_{A_{j_0}(u,v)} \sqrt{1+f'(x)^2} \int_{-\infty}^{Z_1(u,x)} e^{-z^2/2\sigma^2} dz dx \right. \\ & + \left. \sum_{k=1}^{K_j-1} \left[ \int_{A_{j_k}(u,v)} \sqrt{1+f'(x)^2} \int_{-\infty}^{Z_1(u,x)} e^{-z^2/2\sigma^2} dz dx \mathbb{I}_{\{h_u(A_{j_k}(u,v)) < v\}} \right. \right. \\ & + \left. \left. \int_{A_{j_k}(u,v)} \sqrt{1+f'(x)^2} \int_{-\infty}^{Z_2(v,x)} e^{-z^2/2\sigma^2} dz dx \mathbb{I}_{\{h_u(A_{j_k}(u,v)) > v\}} \right] \right. \\ & + \left. \int_{A_{j_{K_j}}(u,v)} \sqrt{1+f'(x)^2} \int_{-\infty}^{Z_1(u,x)} e^{-z^2/2\sigma^2} dz dx \right] \mathbb{I}_{\{f'(A_j) \leq 0\}} \mathbb{I}_{\{a_j > u\}} \\ & + \left[ \sum_{k=0}^{K_j} \int_{A_{j_k}(u,v)} \sqrt{1+f'(x)^2} \int_{Z_1(u,x)}^{Z_2(v,x)} e^{-z^2/2\sigma^2} dz dx \mathbb{I}_{\{h_u(A_{j_k}(u,v)) < v\}} \mathbb{I}_{\{f'(A_j) > 0\}} \right. \\ & + \left[ \int_{A_{j_0}(u,v)} \sqrt{1+f'(x)^2} \int_{-\infty}^{Z_2(v,x)} e^{-z^2/2\sigma^2} dz dx \right. \\ & + \left. \sum_{k=1}^{K_j-1} \left[ \int_{A_{j_k}(u,v)} \sqrt{1+f'(x)^2} \int_{-\infty}^{Z_1(u,x)} e^{-z^2/2\sigma^2} dz dx \mathbb{I}_{\{h_u(A_{j_k}(u,v)) < v\}} \right. \right. \\ & + \left. \left. \int_{A_{j_k}(u,v)} \sqrt{1+f'(x)^2} \int_{-\infty}^{Z_2(v,x)} e^{-z^2/2\sigma^2} dz dx \mathbb{I}_{\{h_u(A_{j_k}(u,v)) > v\}} \right] \right. \\ & + \left. \int_{A_{j_{K_j}}(u,v)} \sqrt{1+f'(x)^2} \int_{-\infty}^{Z_2(v,x)} e^{-z^2/2\sigma^2} dz dx \right] \mathbb{I}_{\{f'(A_j) \leq 0\}} \mathbb{I}_{\{a_{j+1} < u\}} \end{aligned}$$

$$\begin{aligned}
& + \left[ \left[ \int_{A_{j_0}(u,v)} \sqrt{1+f'(x)^2} \int_{Z_1(u,x)}^{Z_2(v,x)} e^{-z^2/2\sigma^2} dz dx \right. \right. \\
& + \left. \sum_{k=1}^{K_j-1} \int_{A_{j_k}(u,v)} \sqrt{1+f'(x)^2} \int_{Z_1(u,x)}^{Z_2(v,x)} e^{-z^2/2\sigma^2} dz dx \mathbb{I}_{\{h_u(A_{j_k}(u,v)) < v\}} \right] \mathbb{I}_{\{f'(A_j) > 0\}} \\
& + \left[ \int_{A_{j_0}(u,v)} \sqrt{1+f'(x)^2} \int_{-\infty}^{Z_2(v,x)} e^{-z^2/2\sigma^2} dz dx \right. \\
& + \sum_{k=1}^{K_j-1} \left[ \int_{A_{j_k}(u,v)} \sqrt{1+f'(x)^2} \int_{-\infty}^{Z_1(u,x)} e^{-z^2/2\sigma^2} dz dx \mathbb{I}_{\{h_u(A_{j_k}(u,v)) < v\}} \right. \\
& + \left. \left. \int_{A_{j_k}(u,v)} \sqrt{1+f'(x)^2} \int_{-\infty}^{Z_2(v,x)} e^{-z^2/2\sigma^2} dz dx \mathbb{I}_{\{h(A_{j_k}(u,v)) > v\}} \right] \right] \\
& + \left. \int_{A_{j_{K_j}}(u,v)} \sqrt{1+f'(x)^2} \int_{-\infty}^{Z_1(u,x)} e^{-z^2/2\sigma^2} dz dx \right] \mathbb{I}_{\{f'(A_j) \leq 0\}} \mathbb{I}_{\{a_j < u < a_{j+1}\}}.
\end{aligned} \tag{2.28}$$

Para diferenciar  $\Lambda_f(u, v)_j$  com relação à  $v$ , será necessário estudar cada caso do subintervalo  $A_j$ , quando  $f'(A_j) > 0$  ou  $f'(A_j) < 0$  e quando  $a_j > u$ ,  $a_{j+1} < u$  ou  $a_j < u < a_{j+1}$  pois nestes casos alguns termos se cancelam.

Inicialmente, note que  $Z_1(u, x_{j_k}(u, v)) = Z_2(v, x_{j_k}(u, v))$ , pois  $h_u(x_{j_k}(u, v)) = v$ , isto é,

$$f(x_{j_k}(u, v)) - \frac{(u - x_{j_k}(u, v))}{f'(x_{j_k}(u, v))} = v.$$

Ou seja,

$$v - f(x_{j_k}(u, v)) = \frac{(x_{j_k}(u, v)) - u}{f'(x_{j_k}(u, v))}. \tag{2.29}$$

Logo

$$\begin{aligned}
Z_1(u, x_{j_k}(u, v)) &= \frac{x_{j_k}(u, v) - u}{f'(x_{j_k}(u, v))} \sqrt{1 + f'(x_{j_k}(u, v))^2} \\
&= (v - f(x_{j_k}(u, v))) \sqrt{1 + f'(x_{j_k}(u, v))^2} \\
&= Z_2(v, x_{j_k}(u, v)).
\end{aligned}$$

Utilizando a expressão (2.25), tem-se que:

**Primeiro caso:**  $a_j > u$

a) Se  $f'(A_j) > 0$ :

$$\frac{\partial}{\partial v} (\Lambda_f(u, v)_j) = \frac{\alpha}{\sqrt{2\pi}\sigma} \sum_{k=1}^{K_j-1} \int_{A_{j_k}(u, v)} (1 + f'(x)^2) e^{-(Z_v(x))^2/2\sigma^2} dx \mathbb{I}_{\{h_u(A_{j_k}(u, v)) < v\}}. \quad (2.30)$$

b) Se  $f'(A_j) < 0$ :

$$\begin{aligned}
\frac{\partial}{\partial v} (\Lambda_f(u, v)_j) &= \frac{\alpha}{\sqrt{2\pi}\sigma} \left[ \frac{\partial}{\partial v} (x_{j_1}(u, v)) \sqrt{1 + f'(x_{j_1}(u, v))^2} \int_{-\infty}^{Z_1(u, x_{j_1}(u, v))} e^{-z^2/2\sigma^2} dz \right. \\
&+ \sum_{k=1}^{K_j-1} \left[ \frac{\partial}{\partial v} (x_{j_{k+1}}(u, v)) \sqrt{1 + f'(x_{j_{k+1}}(u, v))^2} \int_{-\infty}^{Z_1(u, x_{j_{k+1}}(u, v))} e^{-z^2/2\sigma^2} dz \right. \\
&- \left. \frac{\partial}{\partial v} (x_{j_k}(u, v)) \sqrt{1 + f'(x_{j_k}(u, v))^2} \int_{-\infty}^{Z_1(u, x_{j_k}(u, v))} e^{-z^2/2\sigma^2} dz \right] \\
&- \frac{\partial}{\partial v} (x_{j_{K_j}}(u, v)) \sqrt{1 + f'(x_{j_{K_j}}(u, v))^2} \int_{-\infty}^{Z_1(u, x_{j_{K_j}}(u, v))} e^{-z^2/2\sigma^2} dz \\
&+ \left. \sum_{k=1}^{K_j-1} \int_{A_{j_k}(u, v)} (1 + f'(x)^2) e^{-(Z_2(v, x))^2/2\sigma^2} dx \mathbb{I}_{\{h_u(A_{j_k}(u, v)) > v\}} \right]. \\
&= \frac{\alpha}{\sqrt{2\pi}\sigma} \sum_{j=1}^{K_j-1} \int_{A_{j_k}(u, v)} (1 + f'(x)^2) e^{-(Z_2(v, x))^2/2\sigma^2} dx \mathbb{I}_{\{h_u(A_{j_k}(u, v)) > v\}}. \quad (2.31)
\end{aligned}$$

**Segundo caso:**  $a_{j+1} < u$

a) Se  $f'(A_j) \geq 0$ :

$$\begin{aligned}
\frac{\partial}{\partial v}(\Lambda_f(u, v)_j) &= \frac{\alpha}{\sqrt{2\pi}\sigma} \left[ \int_{A_{j_0}(u, v)} (1 + f'(x)^2) e^{-(Z_2(v, x))^2/2\sigma^2} dx \right. \\
&+ \sum_{k=1}^{K_j-1} \int_{A_{j_k}(u, v)} (1 + f'(x)^2) e^{-(Z_2(v, x))^2/2\sigma^2} dx \mathbb{I}_{\{h_u(A_{j_k}(u, v)) < v\}} \\
&\left. + \int_{A_{j_{K_j}}(u, v)} (1 + f'(x)^2) e^{-(Z_2(v, x))^2/2\sigma^2} dx \right]. \tag{2.32}
\end{aligned}$$

b) Se  $f'(A_j) < 0$

$$\begin{aligned}
\frac{\partial}{\partial v}(\Lambda_f(u, v)_j) &= \\
&\frac{\alpha}{\sqrt{2\pi}\sigma} \left[ \frac{\partial}{\partial v}(x_{j_1}(u, v)) \sqrt{1 + f'(x_{j_1}(u, v))^2} \int_{-\infty}^{Z_2(v, x_{j_1}(u, v))} e^{-z^2/2\sigma^2} dz \right. \\
&+ \int_{A_{j_0}(u, v)} (1 + f'(x)^2) e^{-(Z_2(v, x))^2/2\sigma^2} dx \\
&+ \sum_{k=1}^{K_j-1} \left[ \frac{\partial}{\partial v}(x_{j_{k+1}}(u, v)) \sqrt{1 + f'(x_{j_{k+1}}(u, v))^2} \int_{-\infty}^{Z_2(v, x_{j_{k+1}}(u, v))} e^{-z^2/2\sigma^2} dz \right. \\
&- \frac{\partial}{\partial v}(x_{j_k}(u, v)) \sqrt{1 + f'(x_{j_k}(u, v))^2} \int_{-\infty}^{Z_2(v, x_{j_k}(u, v))} e^{-z^2/2\sigma^2} dz \\
&\left. + \int_{A_{j_k}(u, v)} (1 + f'(x)^2) e^{-(Z_2(v, x))^2/2\sigma^2} dx \mathbb{I}_{\{h(A_{j_k}(u, v)) > v\}} \right] \\
&- \frac{\partial}{\partial v}(x_{j_{K_j}}(u, v)) \sqrt{1 + f'(x_{j_{K_j}}(u, v))^2} \int_{-\infty}^{Z_2(v, x_{j_{K_j}}(u, v))} e^{-z^2/2\sigma^2} dz \\
&\left. + \int_{A_{j_{K_j}}(u, v)} (1 + f'(x)^2) e^{-(Z_2(v, x))^2/2\sigma^2} dx \right]. \tag{2.33}
\end{aligned}$$

$$\begin{aligned}
&= \frac{\alpha}{\sqrt{2\pi}\sigma} \left[ \int_{A_{j_0}(u,v)} (1 + f'(x)^2) e^{-(Z_2(v,x))^2/2\sigma^2} dx \right. \\
&+ \sum_{k=1}^{K_j-1} \int_{A_{j_k}(u,v)} (1 + f'(x)^2) e^{-(Z_2(v,x))^2/2\sigma^2} dx \mathbb{I}_{\{h(A_{j_k}(u,v)) > v\}} \\
&\left. + \int_{A_{j_{K_j}}(u,v)} (1 + f'(x)^2) e^{-(Z_2(v,x))^2/2\sigma^2} dx \right]. \tag{2.34}
\end{aligned}$$

**Terceiro caso:**  $a_j < u < a_{j+1}$

a) Se  $f'(A_j) \geq 0$

$$\begin{aligned}
\frac{\partial}{\partial v} (\Lambda_f(u, v)_j) &= \frac{\alpha}{\sqrt{2\pi}\sigma} \left[ \int_{A_{j_0}(u,v)} (1 + f'(x)^2) e^{-(Z_2(v,x))^2/2\sigma^2} dx \right. \\
&+ \left. \sum_{k=1}^{K_j-1} \int_{A_{j_k}(u,v)} (1 + f'(x)^2) e^{-(Z_2(v,x))^2/2\sigma^2} dx \mathbb{I}_{\{h_u(A_{j_k}(u,v)) < v\}} \right]. \tag{2.35}
\end{aligned}$$

b) Se  $f'(A_j) < 0$

$$\begin{aligned}
\frac{\partial}{\partial v}(\Lambda_f(u, v)_j) &= \\
&\frac{\alpha}{\sqrt{2\pi}\sigma} \left[ \frac{\partial}{\partial v}(x_{j_1}(u, v)) \sqrt{1 + f'(x_{j_1}(u, v))^2} \int_{-\infty}^{Z_2(v, x_{j_1}(u, v))} e^{-z^2/2\sigma^2} dz \right. \\
&+ \int_{A_{j_0}(u, v)} (1 + f'(x)^2) e^{-(Z_2(v, x))^2/2\sigma^2} dx \\
&+ \sum_{k=1}^{K_j-1} \left[ \frac{\partial}{\partial v}(x_{j_{k+1}}(u, v)) \sqrt{1 + f'(x_{j_{k+1}}(u, v))^2} \int_{-\infty}^{Z_2(v, x_{j_{k+1}}(u, v))} e^{-z^2/2\sigma^2} dz \right. \\
&- \frac{\partial}{\partial v}(x_{j_k}(u, v)) \sqrt{1 + f'(x_{j_k}(u, v))^2} \int_{-\infty}^{Z_2(v, x_{j_k}(u, v))} e^{-z^2/2\sigma^2} dz \\
&+ \left. \int_{A_{j_k}(u, v)} (1 + f'(x)^2) e^{-(Z_2(v, x))^2/2\sigma^2} dx \mathbb{I}_{\{h_u(A_{j_k}(u, v)) > v\}} \right] \\
&- \left. \frac{\partial}{\partial v}(x_{j_{K_j}}(u, v)) \sqrt{1 + f'(x_{j_{K_j}}(u, v))^2} \int_{-\infty}^{Z_1(u, x_{j_{K_j}}(u, v))} e^{-z^2/2\sigma^2} dz \right]. \\
&= \frac{\alpha}{\sqrt{2\pi}\sigma} \left[ \int_{A_{j_0}(u, v)} (1 + f'(x)^2) e^{-(Z_2(v, x))^2/2\sigma^2} dx \right. \\
&+ \left. \sum_{k=1}^{K_j-1} \int_{A_{j_k}(u, v)} (1 + f'(x)^2) e^{-(Z_2(v, x))^2/2\sigma^2} dx \mathbb{I}_{\{h_u(A_{j_k}(u, v)) > v\}} \right].
\end{aligned} \tag{2.36}$$

A derivação  $\frac{\partial^2}{\partial u \partial v}(\Lambda_f(u, v))$ , será feita de forma análoga à derivada primeira, ou seja, será estudado cada caso do subintervalo  $A_j$ , quando  $f'(A_j) > 0$  ou  $f'(A_j) < 0$  e quando  $a_j > u$ ,  $a_{j+1} < u$  ou  $a_j < u < a_{j+1}$ .

Antes, é necessário determinar  $\frac{\partial}{\partial u}(x(u, v))$ . Da equação (2.29), tem-se que

$$x(u, v) - u = f'(x(u, v))(v - f(x(u, v))). \tag{2.37}$$

Assim, derivando implicitamente a equação (2.37), segue que

$$\begin{aligned}
\frac{\partial}{\partial u}(x(u, v)) - 1 &= f''(x(u, v)) \frac{\partial}{\partial u}(x(u, v))(v - f(x(u, v))) \\
&- f'(x(u, v)) f'(x(u, v)) \frac{\partial}{\partial u}(x(u, v)). \\
&= \frac{\partial}{\partial u}(x(u, v)) [f''(x(u, v))(v - f(x(u, v))) - f'(x(u, v))^2].
\end{aligned}$$

Logo, obtem-se

$$\frac{\partial}{\partial u}(x(u, v)) = \frac{1}{1 + f'(x(u, v))^2 - f''(x(u, v))(v - f(x(u, v)))}. \quad (2.38)$$

Note que

$$q(x(u, v)) = \frac{\partial}{\partial u}(x(u, v)) (1 + f'(x(u, v))^2) e^{-(Z_v(x(u, v)))^2 / 2\sigma^2}.$$

**Primeiro Caso:**  $a_j > u$

a) Se  $f'(A_j) \geq 0$ :

$$\frac{\partial^2}{\partial u \partial v}(\Lambda_f(u, v)_j) = \frac{\alpha}{\sqrt{2\pi}\sigma} \sum_{k=1}^{K_j-1} [q(x_{j_{k+1}}(u, v)) - q(x_{j_k}(u, v))] \mathbb{I}_{\{h_u(A_{j_k}(u, v)) < v\}}.$$

b) Se  $f'(A_j) < 0$ :

$$\frac{\partial^2}{\partial u \partial v}(\Lambda_f(u, v)_j) = \frac{\alpha}{\sqrt{2\pi}\sigma} \sum_{k=1}^{K_j-1} [q(x_{j_{k+1}}(u, v)) - q(x_{j_k}(u, v))] \mathbb{I}_{\{h_u(A_{j_k}(u, v)) > v\}}.$$

**Segundo Caso:**  $a_{j+1} < u$

a) Se  $f'(A_j) \geq 0$ :

$$\begin{aligned}
\frac{\partial^2}{\partial u \partial v}(\Lambda_f(u, v)_j) &= \frac{\alpha}{\sqrt{2\pi}\sigma} [q(x_{j_1}(u, v)) - q(x_{j_{K_j}}(u, v))] \\
&+ \sum_{k=1}^{K_j-1} [q(x_{j_{k+1}}(u, v)) - q(x_{j_k}(u, v))] \mathbb{I}_{\{h_u(A_{j_k}(u, v)) < v\}}.
\end{aligned}$$



b) Se  $f'(A_j) < 0$ :

$$\begin{aligned} \frac{\partial^2}{\partial u \partial v}(\Lambda_f(u, v)_j) &= \frac{\alpha}{\sqrt{2\pi\sigma}} [q(x_{j_1}(u, v)) - q(x_{j_{K_j}}(u, v))] \\ &\quad + \sum_{k=1}^{K_j-1} [q(x_{j_{k+1}}(u, v)) - q(x_{j_k}(u, v))] \mathbb{I}_{\{h_u(A_{j_k}(u, v)) > v\}}. \end{aligned}$$

**Terceiro Caso:**  $a_j < u < a_{j+1}$

a) Se  $f'(A_j) \geq 0$ :

$$\frac{\partial^2}{\partial u \partial v}(\Lambda_f(u, v)_j) = \frac{\alpha}{\sqrt{2\pi\sigma}} [q(x_{j_1}(u, v)) + \sum_{k=1}^{K_j-1} [q(x_{j_{k+1}}(u, v)) - q(x_{j_k}(u, v))] \mathbb{I}_{\{h_u(A_{j_k}(u, v)) < v\}}].$$

b) Se  $f'(A_j) < 0$ :

$$\frac{\partial^2}{\partial u \partial v}(\Lambda_f(u, v)_j) = \frac{\alpha}{\sqrt{2\pi\sigma}} [q(x_{j_1}(u, v)) + \sum_{k=1}^{K_j-1} [q(x_{j_{k+1}}(u, v)) - q(x_{j_k}(u, v))] \mathbb{I}_{\{h_u(A_{j_k}(u, v)) > v\}}].$$

Portanto, tem-se que a função intensidade do processo  $N'$  é:

$$\begin{aligned}
\mu_f(u, v) = & \frac{\alpha}{\sqrt{2\pi}\sigma} \left\{ \sum_{j=1}^{J-1} \left[ \sum_{k=1}^{K_j-1} [q(x_{j_{k+1}}(u, v)) - q(x_{j_k}(u, v))] \times \right. \right. \\
& \left. \left[ \mathbb{I}_{\{h_u(A_{j_k}(u, v)) < v\}} \mathbb{I}_{\{f'(A_j) \geq 0\}} + \mathbb{I}_{\{h_u(A_{j_k}(u, v)) > v\}} \mathbb{I}_{\{f'(A_j) < 0\}} \right] \mathbb{I}_{\{a_j > u\}} + \right. \\
& \left. \left[ q(x_{j_1}(u, v)) - q(x_{j_{K_j}}(u, v)) + \sum_{k=1}^{K_j-1} [q(x_{j_{k+1}}(u, v)) - q(x_{j_k}(u, v))] \times \right. \right. \\
& \left. \left[ \mathbb{I}_{\{h_u(A_{j_k}(u, v)) < v\}} \mathbb{I}_{\{f'(A_j) \geq 0\}} + \mathbb{I}_{\{h_u(A_{j_k}(u, v)) > v\}} \mathbb{I}_{\{f'(A_j) < 0\}} \right] \mathbb{I}_{\{a_{j+1} < u\}} + \right. \\
& \left. \left[ q(x_{j_1}(u, v)) + \sum_{k=1}^{K_j-1} [q(x_{j_{k+1}}(u, v)) - q(x_{j_k}(u, v))] \times \right. \right. \\
& \left. \left. \left. \left[ \mathbb{I}_{\{h_u(A_{j_k}(u, v)) < v\}} \mathbb{I}_{\{f'(A_j) \geq 0\}} + \mathbb{I}_{\{h_u(A_{j_k}(u, v)) > v\}} \mathbb{I}_{\{f'(A_j) < 0\}} \right] \mathbb{I}_{\{a_j < u < a_{j+1}\}} \right\} \right. \\
& \left. \right. \left. \right\}. \tag{2.39}
\end{aligned}$$

■

## 2.2.1 Exemplos

Nesta seção serão apresentados exemplos ilustrativos da medida média e da intensidade do processo de Poisson  $N'$  para duas funções específicas.

1)  $f(x) = x \tan(\theta)$ , onde  $\theta > 0$

Primeiro, determina-se a solução de (2.23) analiticamente pois, no caso da reta,  $x(u, v)$  é o único ponto da reta cuja normal passa pelo ponto  $(u, v)$ , representado por  $P$  na Figura 3.

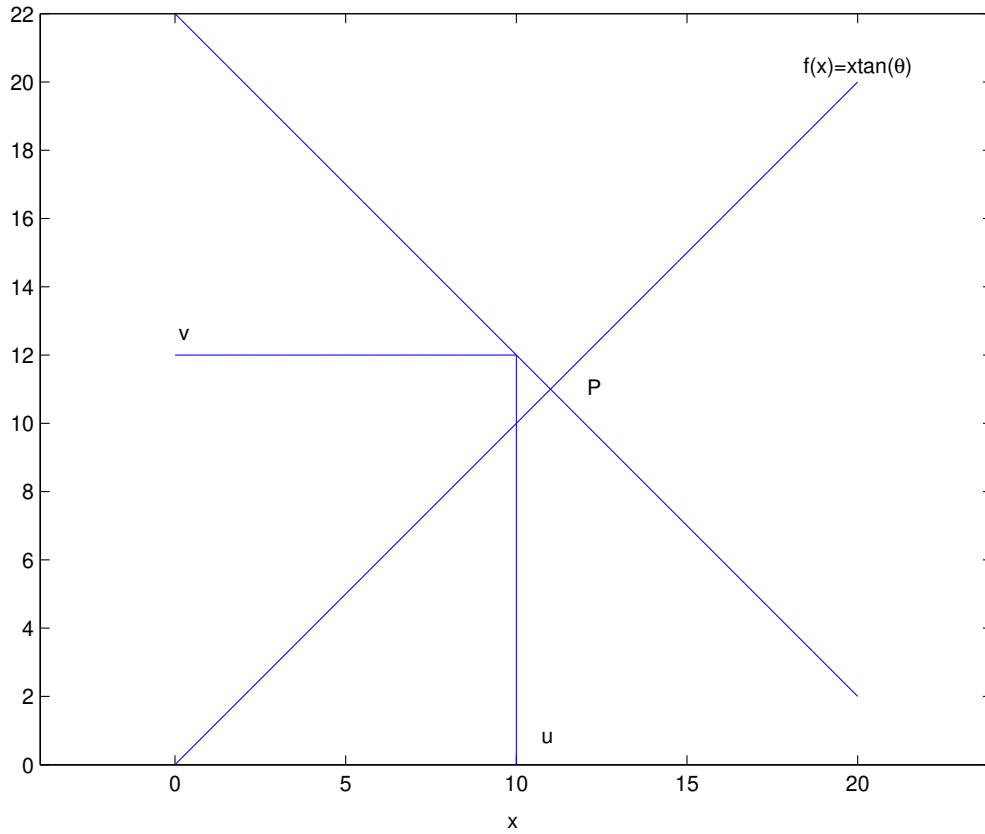


Figura 3:  $x(u,v)$  para a reta: P

Assim,

$$h_u(x(u,v)) = f(x(u,v)) - \frac{(u-x(u,v))}{f'(x(u,v))} \Rightarrow x(u,v) \tan \theta - \frac{u-x(u,v)}{\tan \theta} = v.$$

Portanto

$$x(u,v) = \frac{u + v \tan \theta}{\sec^2 \theta}.$$

Agora, determina-se  $Z_1(u,x)$  e  $Z_2(v,x)$  como em (2.14) e (2.15), respectivamente.

$$Z_1(u,x) = (x-u) \frac{\sqrt{1+f'(x)^2}}{f'(x)} = (x-u) \frac{\sec \theta}{\tan \theta},$$

$$Z_2(v,x) = (v-f(x)) \sqrt{1+f'(x)^2} = (v-x \tan \theta) \sec \theta.$$

De acordo com (2.38), tem-se que

$$\frac{\partial}{\partial u}(x(u, v)) = \frac{1}{1 + \tan^2 \theta} = \frac{1}{\sec^2 \theta}.$$

De acordo com (2.13) ou com (2.28), tem-se que a medida média do processo  $N'$  é:

$$\begin{aligned} \Lambda_f(u, v) &= \frac{\alpha}{\sqrt{2\pi\sigma}} \int_a^{x(u,v)} \sqrt{1 + f'(x)^2} \int_{Z_1(u,x)}^{Z_2(v,x)} e^{-z^2/2\sigma^2} dz dx \\ &= \frac{\alpha}{\sqrt{2\pi\sigma}} \int_a^{\frac{u+v \tan \theta}{\sec^2 \theta}} \sec \theta \int_{(x-u) \frac{\sec \theta}{\tan \theta}}^{(v-x \tan \theta) \sec \theta} e^{-z^2/2\sigma^2} dz dx. \end{aligned}$$

A função intensidade do processo  $N'$  é, de acordo com (2.39):

$$\begin{aligned} \mu_f(u, v) &= \frac{\alpha}{\sqrt{2\pi\sigma}} q(x(u, v)) \\ &= \frac{\alpha}{\sqrt{2\pi\sigma}} \frac{(1 + f'(x(u, v))^2) e^{-(v-f(x(u, v)))^2 (1+f'(x(u, v))^2)/2\sigma^2}}{1 + f'(x(u, v))^2 - f''(x(u, v))(v - f(x(u, v)))} \\ &= \frac{\alpha}{\sqrt{2\pi\sigma}} \frac{1}{\sec^2 \theta} (1 + \tan^2 \theta) e^{-(v - \frac{u+v \tan \theta}{\sec^2 \theta} \tan \theta)^2 (1+\tan^2 \theta)/2\sigma^2} \\ &= \frac{\alpha}{\sqrt{2\pi\sigma}} e^{-(v - (u+v \tan \theta) \sin \theta \cos \theta)^2 \sec^2 \theta / 2\sigma^2} \\ &= \frac{\alpha}{\sqrt{2\pi\sigma}} e^{-(v - u \sin \theta \cos \theta - v \sin^2 \theta)^2 \sec^2 \theta / 2\sigma^2} \\ &= \frac{\alpha}{\sqrt{2\pi\sigma}} e^{-(v(1 - \sin^2 \theta) - u \sin \theta \cos \theta)^2 \sec^2 \theta / 2\sigma^2} \\ &= \frac{\alpha}{\sqrt{2\pi\sigma}} e^{-(v \cos^2 \theta - u \sin \theta \cos \theta)^2 \sec^2 \theta / 2\sigma^2} \\ &= \frac{\alpha}{\sqrt{2\pi\sigma}} e^{-[\cos \theta (v \cos \theta - u \sin \theta)]^2 \sec^2 \theta / 2\sigma^2} \\ &= \frac{\alpha}{\sqrt{2\pi\sigma}} e^{-(v \cos \theta - u \sin \theta)^2 / 2\sigma^2}. \end{aligned} \tag{2.40}$$

A solução (2.40) está de acordo com a apresentada em Garcia (1995).

2)  $f(x) = \cos(x)$

Para este exemplo, não existe solução analítica para (2.23). Assim, serão exibidos somente os gráficos da medida média e da intensidade do processo ( Figuras 4 e 5).

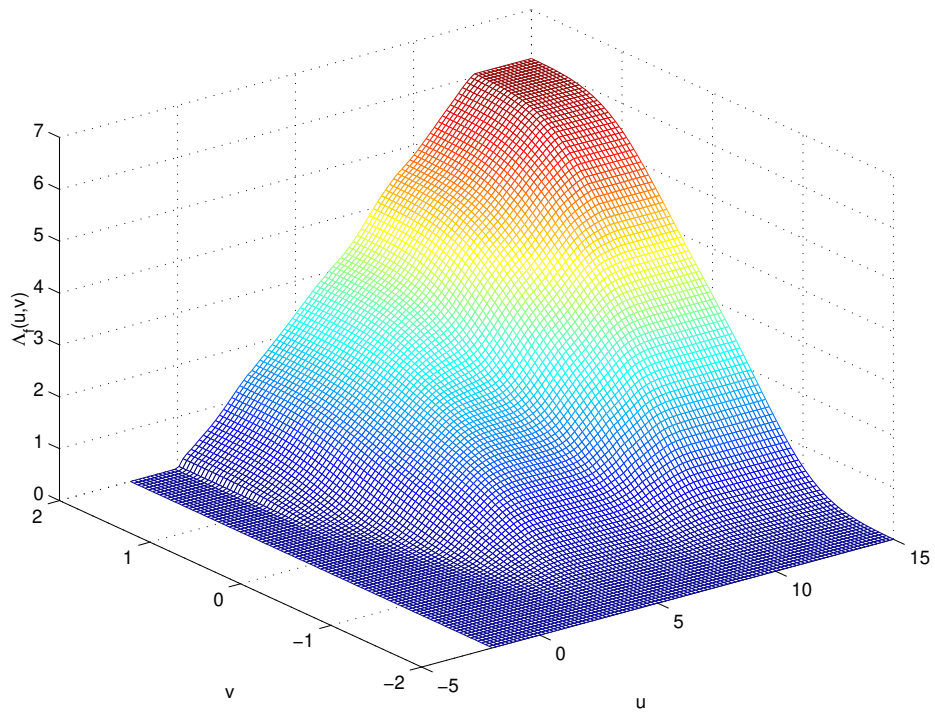


Figura 4: Medida Média para  $f(x) = \cos(x)$

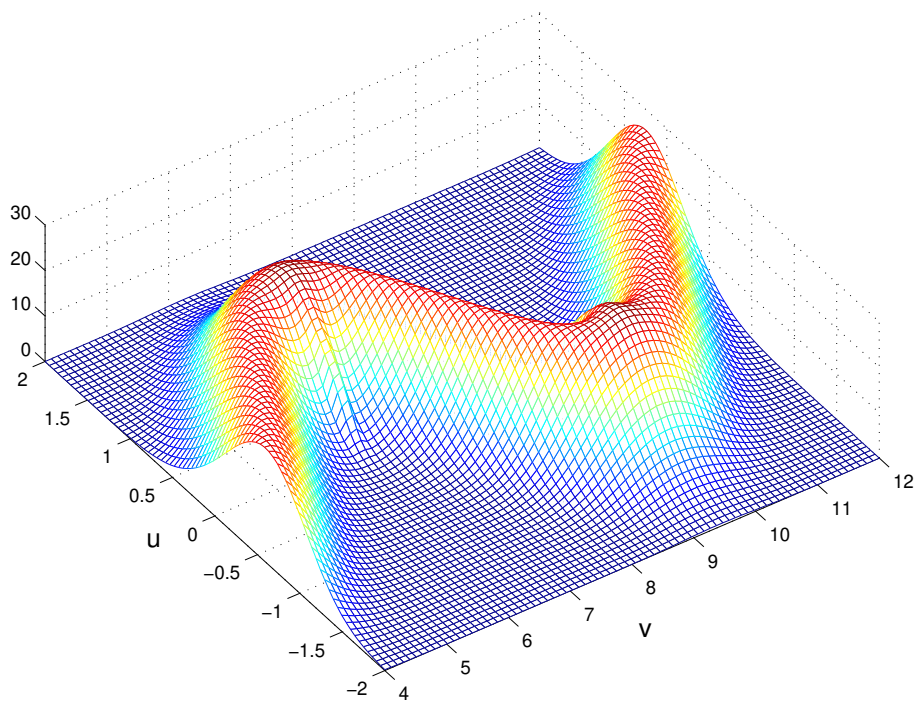


Figura 5: Intensidade para  $f(x) = \cos(x)$

# 3 Estimação Não Paramétrica de Curvas

## 3.1 Introdução

Suponha que se tenha um conjunto de observações  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , onde  $Y$  depende de  $X$  através de  $f$ . Uma forma de estimar a curva  $f$  é minimizar o escore  $\mathcal{L}(f) + \lambda J(f)$ , onde  $\mathcal{L}(f)$  é uma medida de bondade de ajuste,  $J(f)$  é um função de penalização da “rugosidade” da curva e  $\lambda$  é o parâmetro de suavização que faz o balanço entre fidelidade aos dados e suavização da curva.

A “suavidade” de uma curva  $f$  pode ser medida pelo funcional  $J(f)$ , que dará um valor grande se  $f$  tiver uma grande curvatura local e um valor próximo de zero será dado caso  $f$  seja tenha uma curvatura local pequena.

Uma outra justificativa para se usar o funcional  $J(f)$  é que, em alguns casos, a função perda  $\mathcal{L}(f)$  pode divergir para  $+\infty$ . Isso acontece por exemplo quando o objetivo é estimação de densidade e a função perda  $\mathcal{L}(f)$  neste caso é a função de verossimilhança. Mais detalhes desse exemplo podem ser vistos em Dias (1994).

Neste capítulo são apresentadas duas medidas de bondade de ajuste. A primeira é a log-verossimilhança obtida do processo de Poisson  $N'$  utilizada para EMVP. A outra medida é baseada no método de mínimos quadrados. Nos dois processos de otimização, log-verossimilhança e mínimos quadrados, supõe-se que a curva  $f$  pertença a um espaço de Sobolev definido a seguir.

**Definição 3.1.** Defina o espaço de Sobolev  $W_2^k$  de  $k$ -ésima ordem

$$W_2^k[a, b] = \left\{ \begin{array}{l} f : f^{(j)} \text{ é absolutamente contínua,} \\ j = 0, \dots, k - 1 \text{ e } f^{(k)} \in L_2[a, b] \end{array} \right\}$$

que é um subconjunto de  $L_2[a, b]$ , espaço das funções definidas em  $[a, b]$  tal que  $\int f^2(t)dt < \infty$ , para toda  $f \in L_2[a, b]$ .

Neste trabalho, o funcional  $J(f)$  será  $\int (f'')^2$ . Para isto, supõe-se que  $f \in W_2^2[a, b]$ . Schoenberg (1964) mostrou que a solução do problema variacional  $\mathbb{L}(f) + \lambda J(f)$ , quando  $\mathbb{L}(f)$  e  $J(f)$  são quadráticas, é necessariamente um Spline com nós sendo os próprios dados  $X_i$ .

## 3.2 Splines

Dada uma seqüência de pontos  $\xi := (\xi_i)$  não-decrescente e uma função  $f$ , seja  $[\xi_i, \dots, \xi_{i+k} : \xi] = \{\xi_i, \dots, \xi_{i+k}\}$ . Sendo

$$[\xi_i : \xi]f(\xi) = f(\xi_i)$$

e

$$[\xi_i, \xi_{i+1} : \xi]f(\xi) = \frac{f(\xi_{i+1}) - f(\xi_i)}{\xi_{i+1} - \xi_i}, \text{ se } \xi_i \neq \xi_{i+1},$$

segue a seguinte definição.

**Definição 3.2.** Dada uma seqüência de pontos  $\xi := (\xi_i)$  não-decrescente e uma função  $f$ , denomina-se a ***k*-ésima diferença dividida** de  $f$  sobre os pontos  $\xi_i, \dots, \xi_{i+k}$  como

$$[\xi_i, \dots, \xi_{i+k} : \xi]f(\xi) = \begin{cases} \frac{[\xi_{i+1}, \dots, \xi_{i+k} : \xi]f(\xi) - [\xi_i, \dots, \xi_{i+k-1} : \xi]f(\xi)}{\xi_{i+k} - \xi_i} & \text{se } \xi_i \neq \xi_{i+k} \\ f^{(k)}(\xi_i)/k! & \text{se } \xi_i = \xi_{i+k} \end{cases} \quad (3.1)$$

onde  $f^{(k)}$  é a *k*-ésima derivada de  $f$ .

**Definição 3.3.** Seja  $\xi := (\xi_i)$  uma seqüência não-decrescente (que pode ser finita ou infinita). O *i*-ésimo **B-spline normalizado de ordem *k*** para a seqüência  $\xi$  é :

$$B_{i,k,\xi}(x) := (\xi_{i+k} - \xi_i)[\xi_i, \dots, \xi_{i+k}](\xi - x)_+^{k-1}, \forall x \in \mathbb{R}. \quad (3.2)$$

Note que  $B_{i,k,\xi}$  tem suporte compacto, isto é,

$$B_{i,k,\xi}(x) = 0 \text{ para } x \notin [\xi_i, \xi_{i+k}].$$

Uma fórmula de recorrência para B-splines é dada por

$$B_{j,1,\xi}(x) = \begin{cases} 1 & \text{se } \xi_j \leq x < \xi_{j+1} \\ 0 & \text{caso contrário} \end{cases}$$

e  $B_{j,k-1,\xi}(x) > 0$  para  $\xi_j < x < \xi_{j+k-1}$ , para todo  $j$ , tem-se que

$$B_{i,k,\xi}(x) = \frac{x - \xi_i}{\xi_{i+k-1} - \xi_i} B_{i,k-1,\xi}(x) + \frac{\xi_{i+k} - x}{\xi_{i+k} - \xi_{i+1}} B_{i+1,k-1,\xi}(x). \quad (3.3)$$

**Definição 3.4.** Uma **função spline de ordem  $k$  com seqüência de nós  $\xi := (\xi_i)$**  é qualquer combinação linear de  $B$ -splines de ordem  $k$  para a seqüência de nós  $\xi$

$$S_{k,\xi} = \left\{ \sum_i \alpha_i B_{i,k,\xi} : \alpha_i \in \mathbb{R}, \text{ para todo } i \right\}.$$

Seja  $S^k(\xi_1, \dots, \xi_l)$  o conjunto formado por todos os splines de ordem  $k$  com nós  $\xi_1, \dots, \xi_l$ . Pode-se afirmar que este conjunto é um espaço vetorial de dimensão  $k + l$ , pois as funções  $1, x, x^2, \dots, x^{k-1}, (\xi_1 - x)_+^{k-1}, \dots, (\xi_l - x)_+^{k-1}$  são LI (linearmente independentes).

Maiores detalhes, veja Apêndice A e Eubank (1988).

Este trabalho propõe estimar a curva  $f$  através de estimação não paramétrica de curvas, supondo que  $f$  seja uma combinação linear de  $B$ -splines. No Capítulo 4, simulações serão feitas usando  $B$ -splines cúbicos (de ordem 4) para estimação de curvas por mínimos quadrados e por máxima verossimilhança de processos de Poisson.

### 3.3 H-splines

Considere o modelo (3.8), com  $a \leq x_1, \dots, x_n \leq b$ , suponha que a função  $f$  pertença ao espaço  $W_2^2[a, b]$  e que a função de aproximação  $\hat{f} \in S^4(\xi_1, \dots, \xi_l)$ , com bases  $B_{1,4,\xi}, \dots, B_{m,4,\xi}$ .

Suavização por splines é feita quando se considera o número de bases  $m = n$ , onde  $\lambda$  é o parâmetro que controla bondade de ajuste e suavidade. Regressão por splines é feita quando o número de bases é fixo, porém menor que o número de pontos, isto é,  $m \leq n$ . Neste caso, o número de bases atua como parâmetro de suavização.

O método de regressão híbrida por Splines ( $H$ -splines), desenvolvido por Dias (1994), é um método híbrido de estimação de curva que combina idéias de regressão por splines e suavização por splines, escolhendo o número de bases e o parâmetro de suavização seqüencialmente e iterativamente. O número de bases é aumentado por 1 (um) até que se satisfaça um critério de parada que é baseado na pseudo-distância de Hellinger entre duas estimativas consecutivas.

Experimentos de Dias (1999) mostram através de gráficos um relacionamento entre o número de bases  $m$  e o parâmetro de suavização  $\lambda$ . Quando  $m$  aumenta,  $\lambda$  decresce para um ponto e então aumenta com  $m$ . Todos os resultados desta seção podem ser encontrados em



Dias (1999).

**Definição 3.5.** Dadas duas funções  $f, g \in L_2[a, b]$ , define-se a **Distância de Hellinger** por

$$\delta(f, g) = \left[ \int_a^b (f^{1/2} - g^{1/2})^2 \right]^{1/2}. \quad (3.4)$$

Dada qualquer função  $f \in L_2[a, b]$ , tome

$$t_f = \frac{f^2}{\int_a^b f^2}.$$

Assim,  $t_f \geq 0$  e  $\int_a^b t_f = 1$ .

**Definição 3.6.** Dadas duas funções  $f, g \in L_2[a, b]$ , define-se a **afinidade entre  $f$  e  $g$**  por

$$\rho(f, g) = \int_a^b \sqrt{t_f t_g} = \int_a^b \sqrt{\frac{f^2 g^2}{\int_a^b f^2 \int_a^b g^2}} = \frac{\int_a^b |fg|}{\sqrt{\int_a^b f^2 \int_a^b g^2}}. \quad (3.5)$$

Da definição de afinidade, vê-se que  $0 \leq \rho(f, g) \leq 1$ .

Note que

$$\begin{aligned} \delta^2(t_f, t_g) &= \int_a^b (t_f^{1/2} - t_g^{1/2})^2 \\ &= \int_a^b (t_f - 2t_f^{1/2}t_g^{1/2} + t_g) \\ &= \int_a^b t_f - 2 \int_a^b (t_f t_g^{1/2}) + \int_a^b t_g \\ &= 2(1 - \rho(f, g)). \end{aligned}$$

Assim, a pseudo distância  $d(f, g) = \delta^2(t_f, t_g)$  é mínima quando  $\rho(f, g) = 1$ , isto é,  $(\int_a^b f^2 g^2)^{1/2} = \int_a^b |fg|$ .

Seja  $f_m$  uma aproximação finita de  $f$  tal que  $f_m(x) = \sum_{i=1}^n c_j B_{j,4,\xi}(x)$  são os *B-splines*. Assim, aumentando o número de funções bases  $m$  por 1 (um), o procedimento termina quando  $f_{m,\hat{\lambda}} \approx f_{m+1,\hat{\lambda}}$  no sentido da afinidade parcial

$$\rho(f_{m,\hat{\lambda}}, f_{m+1,\hat{\lambda}}) = \frac{\int_a^b |f_{m,\hat{\lambda}}, f_{m+1,\hat{\lambda}}|}{\sqrt{\int_a^b f_{m,\hat{\lambda}}^2 \int_a^b f_{m+1,\hat{\lambda}}^2}} \approx 1.$$

O método *H-splines* será usado neste trabalho para estimação de curvas por regressão não paramétrica, considerando o modelo (3.8) e estimação por Máxima Verossimilhança Penalizada de processos de Poisson, considerando o modelo (2.12).

### 3.4 Estimação de Máxima Verossimilhança da função $f$

Sejam  $(X'_i, Y'_i), i = 1, 2, \dots, N'(A)$  os pontos do processo  $N'$  no conjunto  $A$  e  $\mu_f$  a intensidade desse processo, dada em (2.39). Denote por  $\mathbf{X}' = \{X'_i, i = 1, \dots, n\}$  e  $\mathbf{Y}' = \{Y'_i, i = 1, \dots, n\}$  os vetores dos pontos obtidos. A função de verossimilhança de  $N'$  é dada por (Daley e Vere-Jones, 1988)

$$L_A(\mathbf{X}', \mathbf{Y}', f) = \exp \left\{ \sum_{i=1}^{N(A)} \log \mu_f(X'_i, Y'_i) - \int_A (\mu_f(x, y) - 1) dx dy \right\}. \quad (3.6)$$

Neste caso, a medida de bondade de ajuste é:

$$\mathcal{L}(f) = -\log(L_A(\mathbf{X}', \mathbf{Y}', f)),$$

o que resulta em maximizar a função

$$\begin{aligned} L_A^*(\mathbf{X}', \mathbf{Y}', f; \lambda) &= \log(L_A(\mathbf{X}', \mathbf{Y}', f)) - \lambda \int_{A_X} (f''(x))^2 dx \\ &= \sum_{i=1}^{N(A)} \log \mu_f(X'_i, Y'_i) - \int_A (\mu_f(x, y) - 1) dx dy - \lambda \int_A (f''(x))^2 dx. \end{aligned} \quad (3.7)$$

Assim o problema de estimação de uma curva é composta de dois objetivos: maximizar a adequação aos dados medido pela verossimilhança e evitar curvas “pouco suaves” ou que mostrem uma variabilidade muito rápida.

Maximizar  $L_A^*(\mathbf{X}', \mathbf{Y}', f)$  representa então um critério de compromisso entre os dois objetivos. Assim o possível problema de obter um valor grande de  $L_A(\mathbf{X}', \mathbf{Y}', f)$  e também valores excessivos de  $\int_A (f''(x))^2 dx$  é resolvido subtraindo da verossimilhança um múltiplo de  $\int_A (f''(x))^2 dx$ .

### 3.5 Regressão Não Paramétrica por Splines

O problema dos panfletos descrito na Introdução pode ser visto como um conjunto de observações da forma  $(X_i, Y_i)$ ,  $i = 1, \dots, n$  e um modelo de regressão considerando erro somente na ordenada  $Y$  poderia ser erroneamente proposto. Esta seção apresenta este modelo e uma estimação da curva que ajuste esses dados, através da regressão por mínimos quadrados penalizados.

**Definição 3.7.** Suponha que observações são tomadas em uma variável aleatória contínua  $Y$  para  $n$  valores predeterminados de uma variável contínua independente  $x$ . Sejam  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , os valores resultantes dessa amostra e assumamos que os  $x_i$  e  $y_i$  são relacionados pelo seguinte modelo de regressão:

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (3.8)$$

onde as variáveis aleatórias  $\epsilon_i$  tem média zero, são não correlacionadas, com variância comum  $\sigma^2$  e  $f(x_i)$  são os valores obtidos de alguma função  $f$  desconhecida, calculada nos pontos  $x_1, \dots, x_n$ . A função  $f$  é chamada de **função de regressão** ou **curva de regressão**.

Considere o modelo (3.8), com  $a \leq x_1, \dots, x_n \leq b$ , suponha que a função  $f$  pertença ao espaço  $W_2^2[a, b]$ .

Neste caso, a medida de bondade de ajuste é

$$\mathcal{L}(f) = \sum_{j=1}^n (y_j - f(x_j))^2.$$

Obtém-se o critério conhecido como soma de quadrados penalizada

$$SQP_{\lambda^*}(f) = \sum_{j=1}^n (y_j - f(x_j))^2 + \lambda^* \int_a^b (f''(x))^2 dx, \quad \lambda^* > 0. \quad (3.9)$$

Fazendo  $\lambda' = \lambda^*/(1 + \lambda^*)$ , obtém-se um funcional equivalente

$$SP_{\lambda'}(f) = (1 - \lambda') \sum_{j=1}^n (y_j - f(x_j))^2 + \lambda' \int_a^b (f''(x))^2 dx, \quad (3.10)$$

para  $\lambda' \in (0, 1)$ .

Schoenberg (1964) mostra que o estimador  $f_{\lambda'}$  da função de regressão que minimiza (3.9) é o mesmo minimizador de (3.10) e é chamado de **estimador de suavização por splines**.

A constante  $\lambda'$  controla o balanço entre a suavidade e a bondade de ajuste, pois para valores grandes de  $\lambda'$ , o estimador é mais suave, enquanto que para valores pequenos de  $\lambda'$ , a bondade de ajuste tem maior peso no critério (3.9).

Sendo  $B_{1,k,\xi}, \dots, B_{n,k,\xi}$  uma base para  $NS^k(\xi_1, \dots, \xi_n)$ , com  $k \leq n$ , e  $k \in \mathbb{N}$ ,  $k$  um número par, (onde  $NS^k(\xi_1, \dots, \xi_n)$  é chamado de Natural Splines, ver De Boor (1979)), então existe para  $0 < \lambda' < \infty$  fixo, um único estimador  $f_{\lambda'}$  sob o critério (3.9). Além disso, o estimador  $f_{\lambda'}$  pertence ao espaço  $NS^k(\xi_1, \dots, \xi_n)$  e conseqüentemente este estimador pode ser escrito como combinação linear dessas bases. Ou seja,

$$f_{n,\lambda'}(x) = \sum_{j=1}^n \beta_{\lambda'j} B_{j,k,\xi}(x) \quad (3.11)$$

tal que o vetor  $\beta_{\lambda'} = (\beta_{\lambda'1}, \dots, \beta_{\lambda'n})^t$  satisfaz a equação

$$(X'X + n\lambda'\Psi)\beta_{\lambda'} = X'Y,$$

onde a matriz

$$\Psi_{i,j} = \int_a^b B''_{i,k,\xi}(x) B''_{j,k,\xi}(x) dx, \text{ para } i, j = 1, \dots, n \quad (3.12)$$

e a matriz  $X_{i,j} = B_{i,k,\xi}(x_j)$ . O problema tem solução única

$$\beta_{\lambda'} = (X'X + \lambda'\Psi)^{-1} X'Y.$$

No caso de  $H$ -splines, utiliza-se neste trabalho  $f_{m,\lambda'}(x) = \sum_{j=1}^m \beta_{\lambda'j} B_{j,k,\xi}(x)$  com  $m \ll n$ .

#### ★ Algoritmo GCV em Regressão Não Paramétrica

1. Seja  $m_0$  o número inicial de funções bases e  $\lambda_0$  fixo.
2. Calcule  $\beta_{\lambda_0}$  por resolver  $(X^T X + \lambda_0 \Psi) \beta_{\lambda_0} = X^T y$ .
3. Encontre  $\hat{\lambda}'$  que minimiza a equação

$$GCV(\lambda') = \frac{n^{-1} \sum_{i=1}^n (y_i - f_{m_0,\lambda'}(x_i))^2}{1 - n^{-1} Tr(H(\lambda'))},$$

onde  $H(\lambda') = X(X^T X + \lambda' \Psi)^{-1}$ .

4. Calcule  $f_{m_0,\hat{\lambda}'} = H(\hat{\lambda}')y$ .
5. Incremente o número de funções bases por um e repita os passos (2) a (4) para calcular  $f_{m_0+1,\hat{\lambda}'}$ .

6. Para um número real  $\delta > 0$ , se a distância  $d(f_{m_0, \hat{\chi}}, f_{m_0+1, \hat{\chi}}) < \delta$ , pare o procedimento. O número  $\delta$  pode ser determinado empiricamente de acordo com uma distância particular  $d(., .)$ .

*Experimentos numéricos realizados em Dias (1999), para o caso de regressão não paramétrica, mostram que a afinidade é uma função côncava do número de funções bases e a afinidade parcial aproxima-se muito bem da afinidade. Além disso, o máximo da afinidade e a estabilização da afinidade parcial coincidem. Ou seja, aumentar  $m$  arbitrariamente não somente aumenta o custo computacional, mas também não melhora o ajuste da curva.*

## 4 Simulações

Este estudo tem como objetivo a estimação por máxima verossimilhança Penalizada (EMVP) de processos de Poisson vista no Capítulo 3, considerando o modelo (2.11), utilizando a teoria de H-splines. Somente para efeito de comparação será também utilizada a regressão por mínimos quadrados via H-splines, vista no Capítulo 3, considerando o modelo (3.8).

Simulações serão feitas para algumas funções e uma forma de se comparar os métodos é comparando-se o Erro Médio Quadrático dado por

$$EMQ = \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 / n. \quad (4.1)$$

Para tanto, supõe-se que

$$\hat{f}_m(x) = \sum_{i=1}^m \hat{\beta}_i B_{i,4,\xi}(x), \quad (4.2)$$

onde  $B_{i,4,\xi}(x)$  são splines normalizados como em (3.2),  $m \leq n$ , sendo  $n$  o número de pontos gerados.

Assim, a função de verossimilhança  $L_A^*(f; \lambda)$  dada em (3.7) é uma função dos parâmetros  $\beta = [\beta_1, \dots, \beta_m]$  e  $\lambda$ , isto é,  $L_A^*(f; \lambda) = L_A^*(\beta; \lambda)$ .

A estimativa  $\hat{\beta}$  é obtida da maximização da verossimilhança dada na equação (3.7). Os parâmetros  $\alpha$  e  $\sigma$  nas simulações são considerados constantes.

Para o cálculo do funcional  $J(f) = \int (f''(x))^2 dx$ , observe que

$$(f''(x))^2 = \left( \sum_{i=1}^m \beta_i B_{i,k,\xi}''(x) \right)^2 = \sum_{i=1}^m \beta_i^2 B_{i,k,\xi}''(x)^2 + 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^m \beta_i \beta_j B_{i,k,\xi}''(x) B_{j,k,\xi}''(x).$$

Logo

$$\begin{aligned}
\int (f''(x))^2 dx &= \sum_{i=1}^m \beta_i^2 \int B_{i,k,\xi}''(x)^2 dx + 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^m \beta_i \beta_j \int B_{i,k,\xi}''(x) B_{j,k,\xi}''(x) dx \\
&= \beta^T \Phi \beta,
\end{aligned} \tag{4.3}$$

pois as integrais em (4.3) são simplesmente os elementos da matriz  $\Psi$  definida em (3.12).

**Teorema 4.1.** *Um processo de Poisson homogêneo de taxa  $\alpha$  em um intervalo finito de comprimento  $C$  pode ser realizado em dois estágios: 1. gerar o número de ocorrências,  $n$ , em um intervalo de acordo com a distribuição de Poisson com parâmetro  $\alpha C$ ; 2. escolher  $n$  pontos no intervalo independentemente de acordo com a distribuição uniforme nesse intervalo.*

**★ Algoritmo para geração do processo de Poisson  $N'$**

1. Gere um processo de Poisson homogêneo de intensidade  $\alpha$  (conhecido) em uma curva  $f$  contínua no intervalo  $[a, b]$  de comprimento  $C$ . Sejam  $U_1, \dots, U_n$  os pontos gerados. Este processo é realizado segundo o Teorema 4.1.
2. Aplique a transformação  $X_i = \inf \left\{ x : \frac{1}{C} \int_a^x \sqrt{1 + f'(t)^2} dt \geq U_i \right\}$ ,  $i = 1, \dots, n$  dada em (2.10) para obter os pontos do processo  $N$ .
3. Aplique aos pontos do processo gerado  $N$  uma variável aleatória Normal univariada como no modelo (2.11)

$$(X'_n, Y'_n) = (X_n, f(X_n)) + (-f'(X_n), 1) \frac{Z_n}{\sqrt{1 + f'(X_n)^2}}, \tag{4.4}$$

obtendo o processo de Poisson não-homogêneo  $N'$ .

Dados os pontos gerados  $(X'_i, Y'_i)$ ,  $i = 1, \dots, n$ , faz-se estimação via  $H$ -splines, mostrada na Seção 3.3, com o valor inicial  $m_0$  dependendo da função que está sendo analisada.

**★ Construção do vetor de nós**

Considerando  $m$  o número de bases,  $k = 4$  a ordem dos  $B$ -splines (cúbicos), os elementos do vetor  $\xi = \{\xi_1, \dots, \xi_{m+4}\}$  de nós são construídos de tal forma que eles tenham multiplicidade 1 no interior do intervalo  $[X'(1), X'(n)]$  e replica-se os extremos  $X'(1)$  e  $X'(n)$  3 vezes. Assim,

$$\xi_1 = \dots = \xi_4 = X'(1) \text{ e } \xi_{m+1} = \dots = \xi_{m+4} = X'(n). \quad (4.5)$$

Os  $m - 4$  valores do interior dos nós serão os  $(m - 4)$ -percentis do vetor  $X' = \{X'(1), X'(2), \dots, X'(n)\}$ . Isto é feito da seguinte forma:

1. Seja  $c=1/(m-3)$ .
2. Construa o vetor  $k_1$  de elementos igualmente espaçados no intervalo  $[c, 1-c]$  com  $m-4$  elementos.
3. Faça  $k_2$  o vetor de  $m - 4$  elementos, onde a  $i$ -ésima posição  $k_2(i)$  vale o percentil de  $k_1(i)\%$  de  $X'$ .
4. Faça  $\xi(5, \dots, m) = k_2$ .

Utilizando essas condições, de Boor (1978) garante que a seqüência  $B_{1,4,\xi}, \dots, B_{m,4,\xi}$  de B-splines de ordem 4 é uma base para  $\mathbf{P}_{4,t,\nu}$ , o espaço de polinômios por partes de ordem  $k$  com seqüência de nós  $t$  e vetor de condições de continuidade  $\nu$ . (Mais detalhes, veja Apêndice A).

Utilizando o vetor  $\xi$  de nós, De Boor (1978) exibe teoremas que garantem boas taxas de convergência para aproximação de funções com algumas condições de suavidade.

As integrais e a maximização em (3.7) não podem ser obtidas analiticamente, portanto será utilizado o método de Monte Carlo para obter a integral (ver Apêndice C).

No problema de mínimos quadrados, o algoritmo  $GCV(\lambda')$  da Seção 3.3 é calculado com um custo computacional relativamente baixo pois, dado  $\lambda'$ , a solução  $\hat{f}_{\lambda'} = H(\lambda')Y = X(X'X + \lambda'\Psi)^{-1}X'Y$  é calculada matricialmente, restando achar o melhor  $\lambda'$  que minimize o GCV.

Na EMV do processo de Poisson  $N'$ , o mesmo não acontece pois, dado  $\lambda$ , na solução

$$\hat{f}_{m,\lambda}(x) = \sum_{i=1}^m \hat{\beta}_{i\lambda} B_{i,4,\xi}(x) \quad (4.6)$$

a estimativa  $\hat{\beta}_{i\lambda}$  é calculada numericamente, minimizando a equação (3.7), com  $\lambda$  constante. Este trabalho demanda um custo computacional alto.

Outro problema que surge na tentativa de se produzir um algoritmo H-splines na EMV do processo  $N'$  é a inexistência direta da equação (B.3) (ver Apêndice B), a fórmula do GCV utilizada no problema de mínimos quadrados. Uma alternativa seria fazer o procedimento original de Validação Cruzada (veja Apêndice B.1), usando o método "leave-one-out".



Porém esta solução teria um custo  $n$  vezes maior, com  $n$  o número de pontos do processo  $N'$ , pois teria que fazer  $n$  estimativas  $\hat{f}_\lambda^{(-i)}$ , soluções da função de verossimilhança (3.7) para cada  $\lambda$ .

Com esse problema em mente, decidiu-se neste trabalho manter a idéia do método H-splines, porém ao invés de estimar um  $\hat{\lambda}$  minimizando o CV, para cada passo ( número de funções bases), obtem-se esse parâmetro de tal forma que o peso da função de suavidade seja o mesmo tanto no critério de Mínimos Quadrados dado em (3.10) quanto no critério de Máxima Verossimilhança do processo de poisson  $N'$  dado em (3.7). O seguinte procedimento explica como se chega ao algoritmo de H-splines para EMVP.

Considere o critério de mínimos quadrados

$$SP_f(\lambda') = (1 - \lambda') \sum_{j=1}^n (y_j - f(x_j))^2 + \lambda' \int_a^b (f''(x))^2 dx \quad (4.7)$$

$$= (1 - \lambda')MQO + \lambda'J(f) \quad (4.8)$$

onde  $MQO = \sum_{j=1}^n (y_j - f(x_j))^2$ ,  $J(f) = \int_a^b (f''(x))^2 dx$  e o problema de maximizar a função de verossimilhança do processo  $N'$  dada por

$$\begin{aligned} L_A^*(\mathbf{X}', \mathbf{Y}', f; \lambda) &= (1 - \lambda) \left[ \sum_{i=1}^{N(A)} \log \mu_f(X'_i, Y'_i) - \int_A [\mu_f(x, y) - 1] dx dy \right] - \lambda \int_A (f''(x))^2 dx \\ &= (1 - \lambda)L(\mathbf{X}', \mathbf{Y}', f) - \lambda J(f), \end{aligned} \quad (4.9)$$

onde  $L(\mathbf{X}', \mathbf{Y}', f)$  é a função de verossimilhança do processo  $N'$ . O peso da Penalty  $J(f)$  no problema de Mínimos Quadrados é

$$PMINQ(\lambda') = \frac{\lambda' J(f)}{SP_f(\lambda')} \quad (4.10)$$

enquanto que o peso da Penalty  $J(f)$  na EMV de  $N'$  é

$$PEMV(\lambda) = \frac{\lambda J(f)}{(1 - \lambda)L(\mathbf{X}', \mathbf{Y}', f)}. \quad (4.11)$$

Igualando as equações (4.10) e (4.11),

$$\frac{\lambda J(f)}{(1 - \lambda)L(\mathbf{X}', \mathbf{Y}', f)} = \frac{\lambda' J(f)}{SP_f(\lambda')}.$$

Considerando  $\lambda$  como uma função de  $\lambda'$ , obtem-se que

$$\lambda = \frac{\lambda' L(f)}{SP_f(\lambda') + \lambda' L(\mathbf{X}', \mathbf{Y}', f)}. \quad (4.12)$$

★ **Algoritmo H-splines para EMVP**

1. Seja  $m_0$  o número inicial de funções bases.
2. Encontre  $\hat{\lambda}'$  que minimiza a equação

$$GCV(\lambda') = \frac{n^{-1} \sum_{i=1}^n (y_i - f_{m_0, \lambda'}(x_i))^2}{1 - n^{-1} Tr(H(\lambda'))},$$

onde  $H(\lambda') = X(X^T X + \lambda' \Psi)^{-1}$ .

3. Encontre  $\hat{\lambda}$  na equação (4.12).
4. Calcule  $\hat{f}_{m_0, \hat{\lambda}}$  como em (4.6).
5. Aumente o número de bases por um e repita os passos 2 a 3 para calcular  $\hat{f}_{m_0+1, \hat{\lambda}}$ .
6. Calcule  $d(\hat{f}_{m_0, \hat{\lambda}}, \hat{f}_{m_0+1, \hat{\lambda}})$ .
7. Para um número real fixado  $\delta > 0$ , se a distância  $d(\hat{f}_{m_0, \hat{\lambda}}, \hat{f}_{m_0+1, \hat{\lambda}}) < \delta$ , pare o procedimento. Senão repita os passos 4 a 6.

A distância  $d(\cdot, \cdot)$  utilizada neste trabalho é a distância de Hellinger dada em (3.4).

Para determinar as soluções da equação (2.23), isto é, os valores  $x_{j_k}(u, v) \in A_j$  tais que

$$h_u(x_{j_k}(u, v)) = f(x_{j_k}(u, v)) - \frac{(u - x_{j_k}(u, v))}{f'(x_{j_k}(u, v))} = v, \quad k = 1, \dots, K_j \quad (4.13)$$

será necessário transformar a função  $h_u$  em um spline para utilizar uma função do Matlab que procura os zeros (raízes) de uma função na forma de um spline. Isto é necessário, pois para uma função na forma “natural”, o Matlab não dispõe de nenhuma função para calcular mais de um zero, se existir, de uma função em um dado intervalo. Essa transformação é feita usando interpolação dos pontos  $(x_i, h_u(x_i))$  para um determinado número de valores de  $x$  no subintervalo  $A_j$  onde  $f$  é monótona. Para mostrar que o erro de aproximação é muito pequeno, quase zero, os gráficos da Figura 2 serão rerepresentados, porém juntos com os gráficos da função  $h_u$  transformada em um spline.

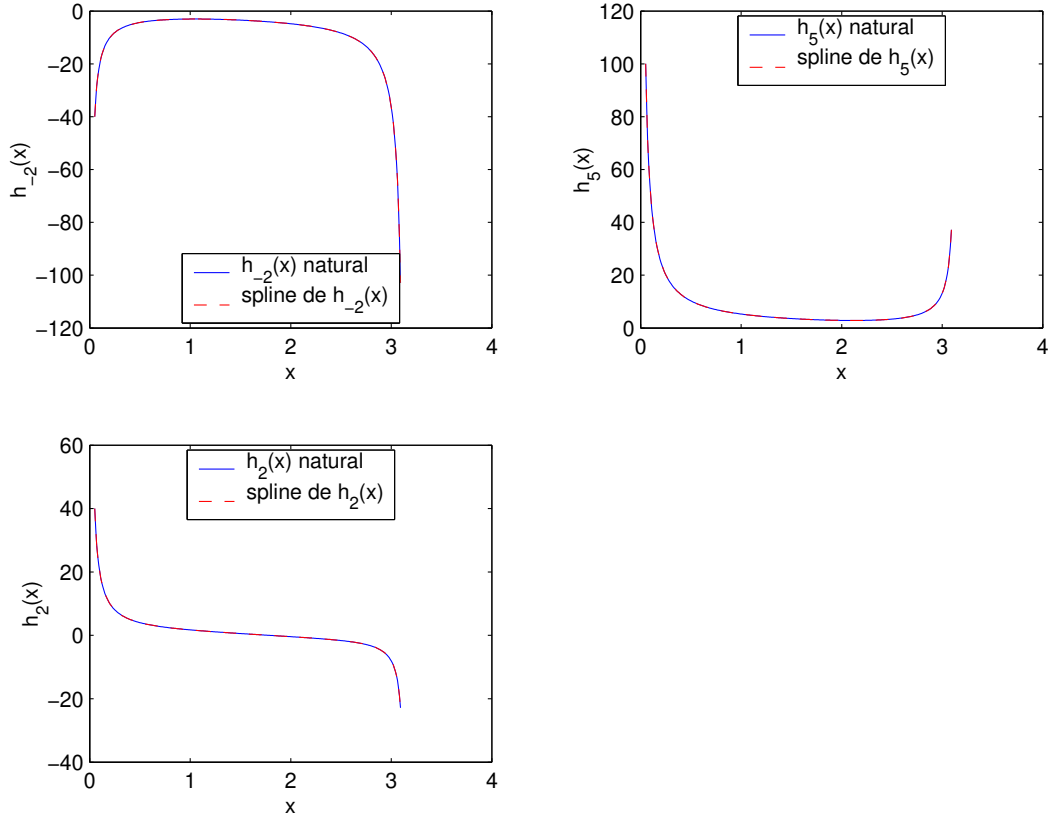


Figura 6: Função  $h_u(x)$  para  $f(x) = \cos(x)$ ,  $x \in (0, \pi)$ .

## 4.1 Exemplos

Nesta seção serão exibidos alguns exemplos de simulações feitas para algumas funções e um exemplo com dados reais de terremotos ocorridos na Califórnia, EUA.

Para os exemplos de dados simulados, cada figura contém os pontos do processo  $N'$  dado em (4.4), a curva verdadeira e as curvas estimadas, por Mínimos Quadrados e por Máxima Verossimilhança do Processo de Poisson. Para cada figura, serão exibidos quadros comparativos entre os dois métodos, relatando a Distância de Hellinger atingida, o coeficiente  $\lambda$  de penalização, o número de nós e o Erro Médio Quadrático atingido.

Para o exemplo em que  $f(x) = x \tan(\theta)$ , uma reta centrada na origem (sem perda de generalidade), o estimador de máxima verossimilhança é encontrado de forma analítica, dado por

$$\tilde{\theta} = \frac{1}{2} \arctan \frac{2 \sum_{\Gamma_t} X'_i Y'_i}{\sum_{\Gamma_t} (X_i'^2 - Y_i'^2)} + \frac{\pi}{2} \operatorname{sgn} \left( \sum_{\Gamma_t} X'_i Y'_i \right) \mathbb{I} \left( \sum_{\Gamma_t} (X_i'^2 - Y_i'^2) < 0 \right),$$

onde  $\Gamma_t = \{(x, y) : x^2 + y^2 \leq t\}$  e  $(X'_i, Y'_i)$ ,  $i = 1, \dots, n$  são os pontos do processo  $N'$ . Assim  $\hat{f}(x) = x \tan(\hat{\theta})$ . Mais detalhes, veja Garcia (1995).

No trabalho de estimação dos dados, foi utilizado um microcomputador com 20 Gb de memória, velocidade de processamento de 950 Mhz e 128 Mb de memória RAM. O software usado foi o Matlab, versão 6.1 Release 12, no sistema operacional Linux.

No processo de otimização (maximização da verossimilhança penalizada dada em (3.7), foram adotados os seguintes critérios:

- Número de iterações:  $300 * m$ , sendo  $m$  o número de parâmetros a estimar;
- Taxa de convergência: Tolerância de  $10^{-4}$  para os parâmetros e de  $10^{-4}$  para a função de verossimilhança penalizada.

Como neste trabalho se utiliza o método de H-splines no processo de estimação da curva  $f$ , o critério de parada é baseado na distância de Hellinger dada em (3.4). Assim, o processo iterativo de aumentar o número de bases (B-splines) necessárias para a estimação de  $f$  para quando a distância  $d$  atingir o valor de 0,995. Não houve replicações no processo de estimação neste trabalho.

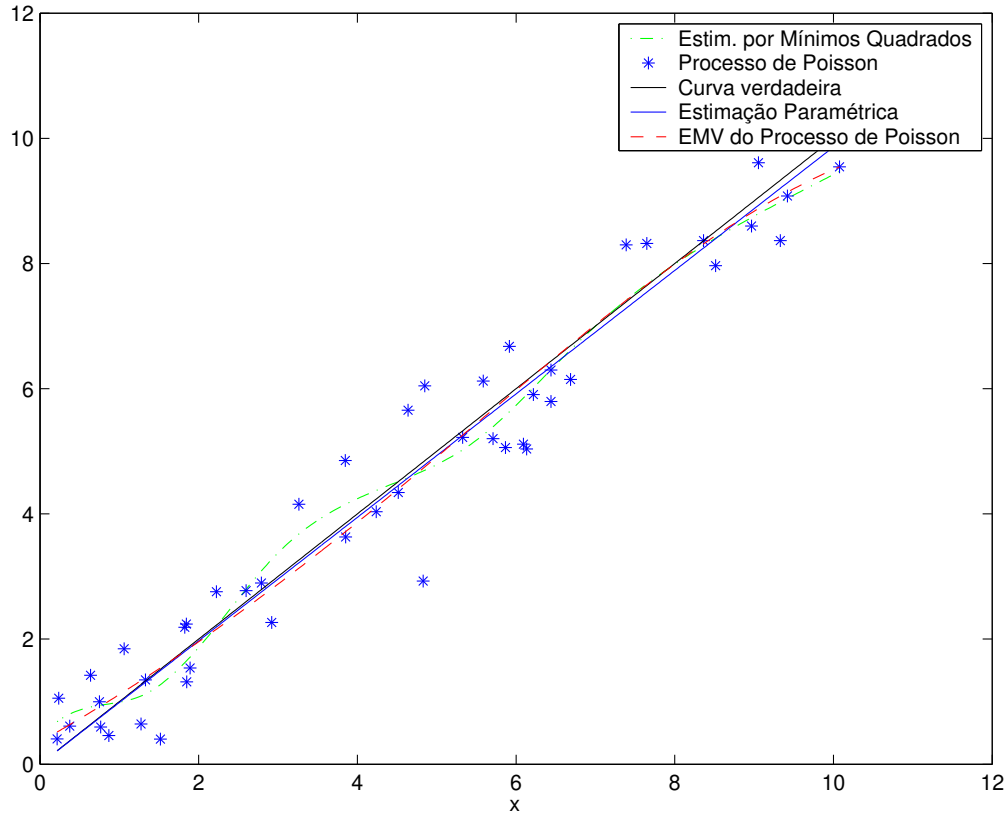


Figura 7: Gráficos do processo  $N'$  ( $\sigma = 0.5$ ,  $\alpha = 4$ ,  $n = 47$ ) e das curvas estimadas, por Mínimos Quadrados, pelo Modelo Paramétrico e por Máxima Verossimilhança para a curva  $f(x) = \tan(\pi/4)$ ,  $x \in (0, 10)$ .

Quadro Comparativo relativo à Figura 7

Parâmetros	EMV	Mín. Quadrados	Modelo Paramétrico
Dist. Hellinger	-	0,9998	-
$\lambda$	0,001	0,0123	-
Número de nós	9	12	-
EMQ	0,0226	0,0666	0,0066

No exemplo em que a curva  $f$  é uma reta, pode-se notar que a EMVP do processo de Poisson atinge uma EMQ pequeno, embora maior que o EQM da estimativa paramétrica. A estimativa de Mínimos Quadrados tem um EQM maior que as outras duas. As estimativas pelos três processos são satisfatórias para o caso da reta.

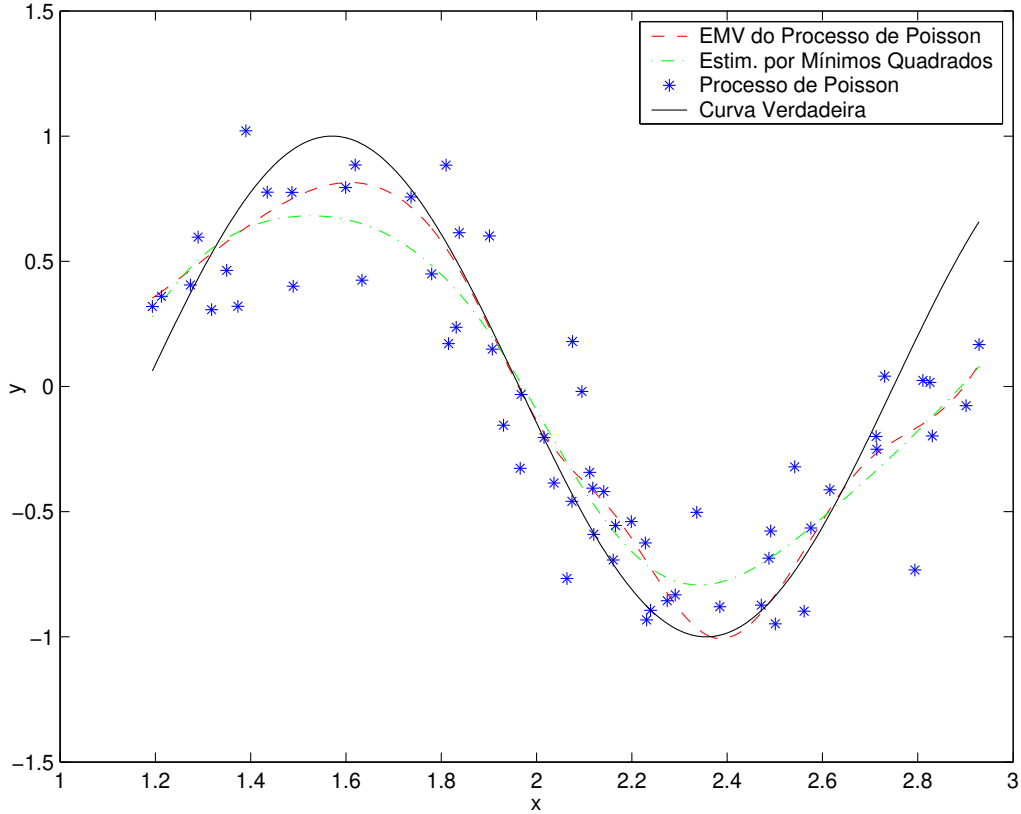


Figura 8: Gráficos do processo  $N'$  ( $\sigma = 0.01$ ,  $\alpha = 12$ ,  $n = 62$ ) e das curvas estimadas, por Mínimos Quadrados e por Máxima Verossimilhança para a curva  $f(x) = \cos(4x)$ ,  $x \in (1.2, 2.8)$ .

Quadro Comparativo relativo à Figura 8

Parâmetros	EMV	Mín. Quadrados
Dist. Hellinger	0,9954	0,9998
$\lambda$	0,0272	0,0039
Número de nós	14	13
EMQ	0,0341	0,0583

No exemplo da Figura 8, em que  $f(x) = \cos(4x)$ , existem mais pontos na parte interior das curvaturas locais (pela própria configuração do processo  $N'$ ), fazendo com que as estimativas passem sempre pelo interior dessas curvaturas. Comparando as duas estimativas, nota-se que a EMV sofre menos dessa peculiaridade do processo. No extremo direito da Figura 8 nota-se que as estimativas ficam ruins, onde se vê que a configuração dos pontos neste local se distancia da curva verdadeira.

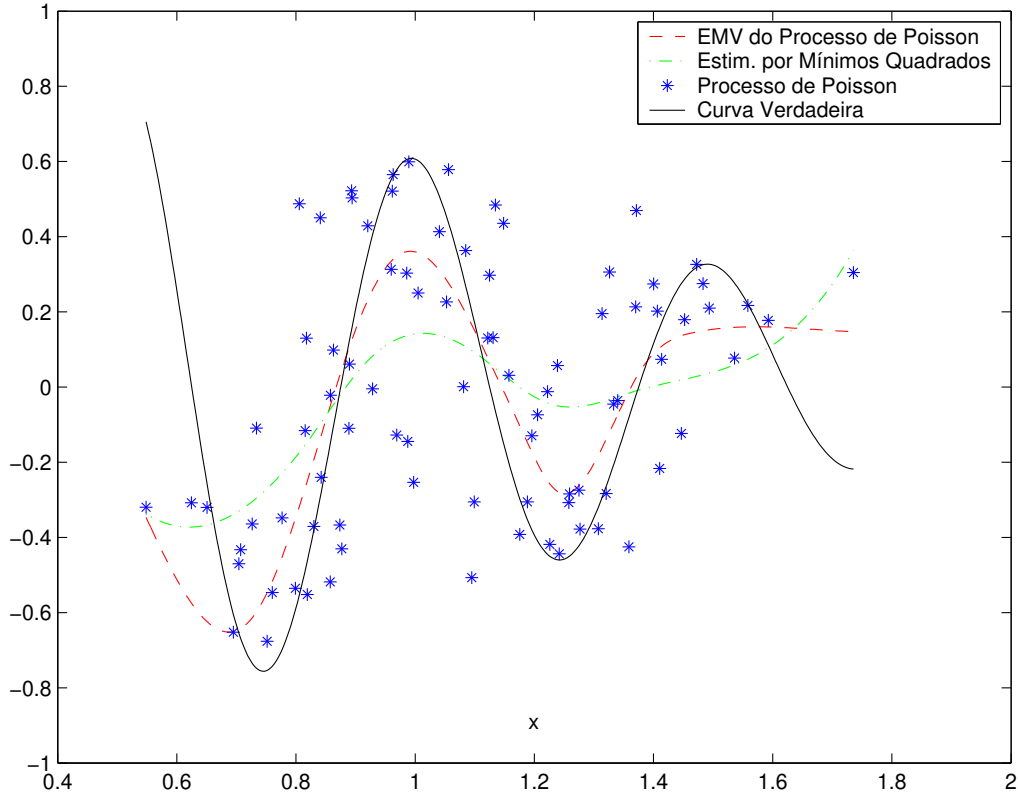


Figura 9: Gráficos do processo  $N'$  ( $\sigma = 0.07$ ,  $\alpha = 17$ ,  $n = 86$ ) e das curvas estimadas, por Mínimos Quadrados e por Máxima Verossimilhança para a curva  $f(x) = \cos(4\pi x) \exp(-x^2/2)$ ,  $x \in (0.66, 1.6)$ .

Quadro Comparativo relativo à Figura 9

Parâmetros	EMV	Mín. Quadrados
Dist. Hellinger	0,9963	0,9966
$\lambda$	0,0167	0,0123
Número de nós	20	17
EMQ	0,0827	0,1260

Na Figura 9, nota-se que a curva verdadeira apresenta uma curvaturas locais fortes, onde se vêem pontos no interior das curvaturas distantes de seu vértice. É nesta configuração de pontos que a estimação de mínimos quadrados falha, pois para este método, esses pontos estariam vindo de pontos próximos ao vértice, o que é equivocado, pois pelo processo  $N'$ , esses pontos vêm de uma direção normal à curva. Isto faz com que a curva estimada por mínimos quadrados passe muito distante dos vértices das curvaturas. Note novamente o problema nos extremos, onde a estimação fica insatisfatória. O problema de fronteira é natural em estimação não paramétrica, não desqualificando portanto a estimativa de máxima verossimilhança.

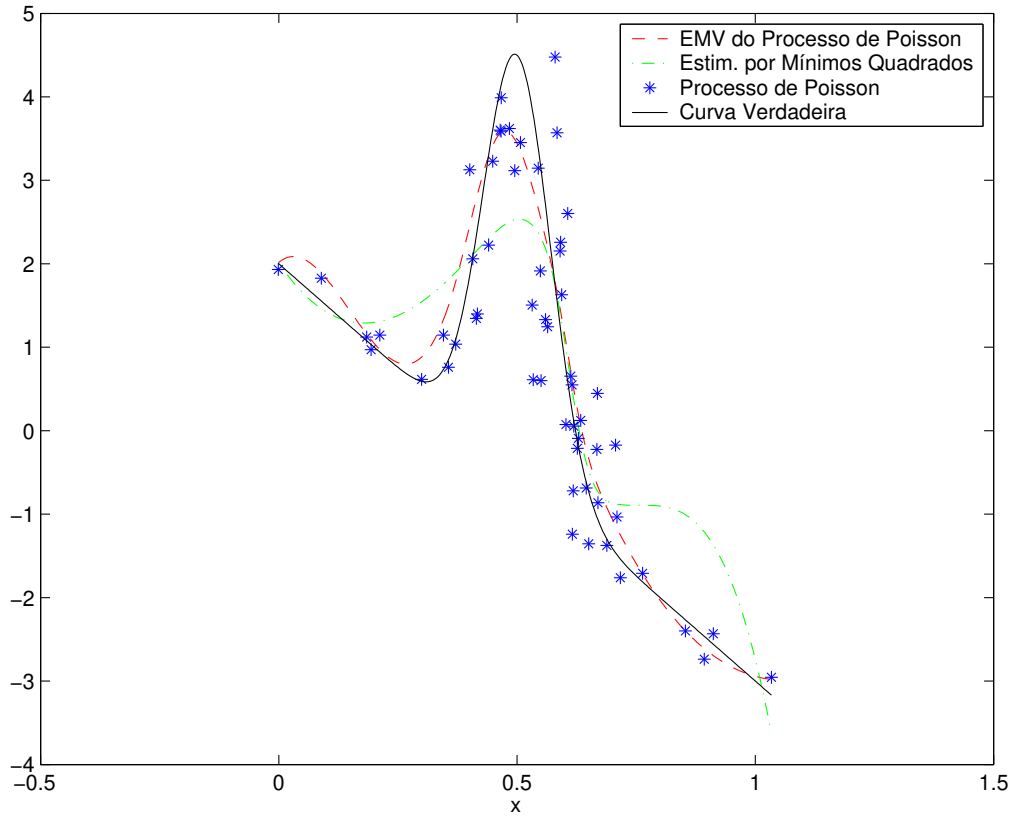


Figura 10: Gráficos do processo  $N'$  ( $\sigma = 0.04$ ,  $\alpha = 5$ ,  $n = 57$ ) e das curvas estimadas, por Mínimos Quadrados e por Máxima Verossimilhança para a curva  $f(x) = 2 - 5x + 5 \exp(-100(x - 0.5)^2)$ ,  $x \in (0, 1)$ .

Quadro Comparativo relativo à Figura 10

Parâmetros	EMV	Mín. Quadrados
Dist. Hellinger	0,9951	0,9996
$\lambda$	0,0012	0,0039
Número de nós	14	15
EMQ	0,1464	0,7270

Para o exemplo da curva  $f$  mostrada na Figura 10 novamente percebe-se a dificuldade do método de mínimos quadrados de estimar a curvatura da função. Por outro lado, nota-se um desempenho razoável da EMV em estimar a curva  $f$ .



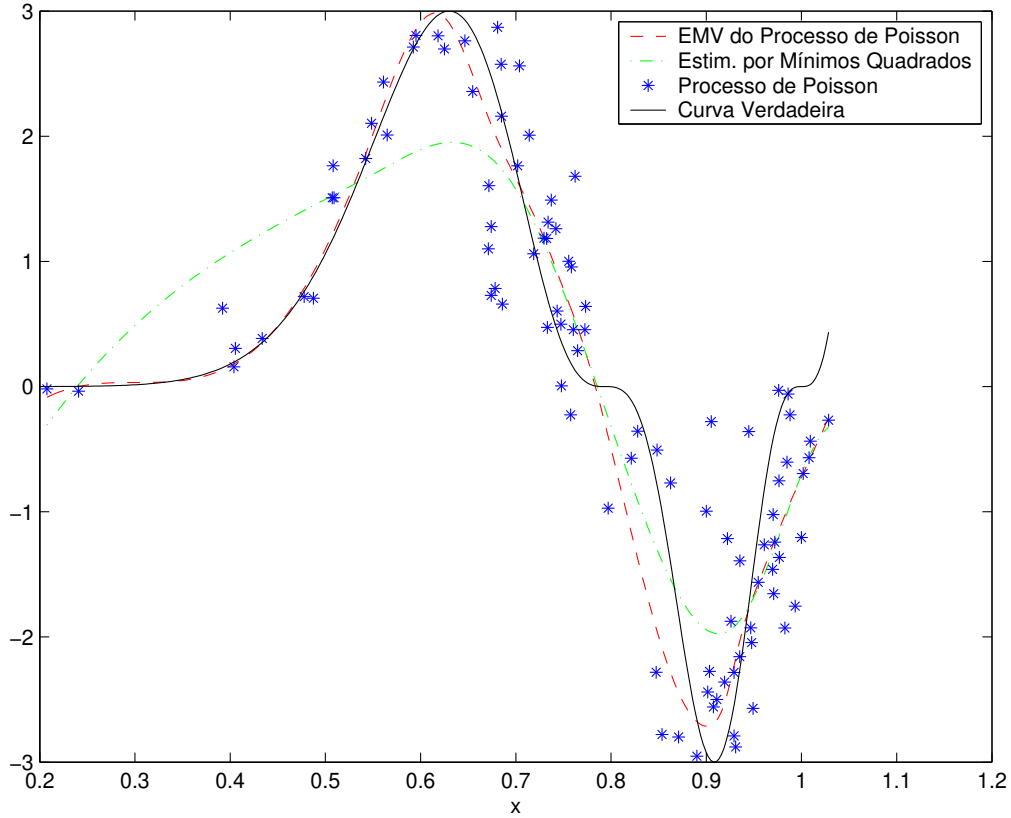


Figura 11: Gráficos do processo  $N'$  ( $\sigma = 0.03$ ,  $\alpha = 8$ ,  $n = 97$ ) e das curvas estimadas, por Mínimos Quadrados e por Máxima Verossimilhança para a curva  $f(x) = 3 * (\sin(2 * \pi * x.^3))$ ,  $x \in (0.2, 1)$ .

Quadro Comparativo relativo à Figura 11

Parâmetros	EMV	Mín. Quadrados
Dist. Hellinger	0,9963	0,9995
$\lambda$	0,0015	0,0015
Número de nós	15	14
EMQ	0,1671	0,3980

Na Figura 11 novamente tem-se que o método de estimação de curvas por mínimos quadrados falha no processo de estimar a curva  $f$ , com os dados vindos do processo  $N'$ . Note que, em todos o exemplos aqui mostrados, a Distância de Hellinger alcançada pelo método de mínimos quadrados é sempre maior que a de EMV. Isto reflete que, mesmo aumentando o número de bases, a estimativa da curva por mínimos quadrados para esse número de bases não muda muito em relação ao passo anterior (uma base a menos).

### 4.1.1 Estimação de falhas geológicas

*Nesta subseção aplica-se o procedimento de EMV relativo ao processo de Poisson  $N^l$ , dado no Capítulo 2 para um conjunto de dados reais provenientes de terremotos ocorridos na Califórnia, EUA. Foram selecionados 95 abalos ocorridos nos anos de 1980 a 1999, de magnitude maior que 5, mais precisamente terremotos próximos à falha de San Andreas, na parte norte. Os dados podem ser obtidos pela Internet no endereço: < <http://www.scecdc.scec.org/sanandre.html> >.*

*Sabendo que epicentros de terremotos não se repetem em um mesmo lugar e que a ocorrência de epicentros em lugares distintos são independentes, tem-se que o processo de ocorrência desses abalos em uma região segue uma distribuição de Poisson. Sabe-se também que terremotos ocorrem ao redor de falhas geológicas existentes no planeta.*

*Com essas hipóteses pretende-se modelar, a partir dos dados apresentados acima, uma curva que se ajuste aos dados, que no caso seria a localização da falha de San Andreas, de onde surgem os terremotos ocorridos nessa região.*

*Na estimação da curva, o procedimento será o mesmo adotado nas simulações, com exceção que os parâmetros  $\alpha$  e  $\sigma$  agora terão de ser estimados, além de  $\beta$  e  $\lambda$ . Ou seja, a função de verossimilhança  $L_A^*(f; \lambda) = L_A^*(\beta, \alpha, \sigma; \lambda)$ .*

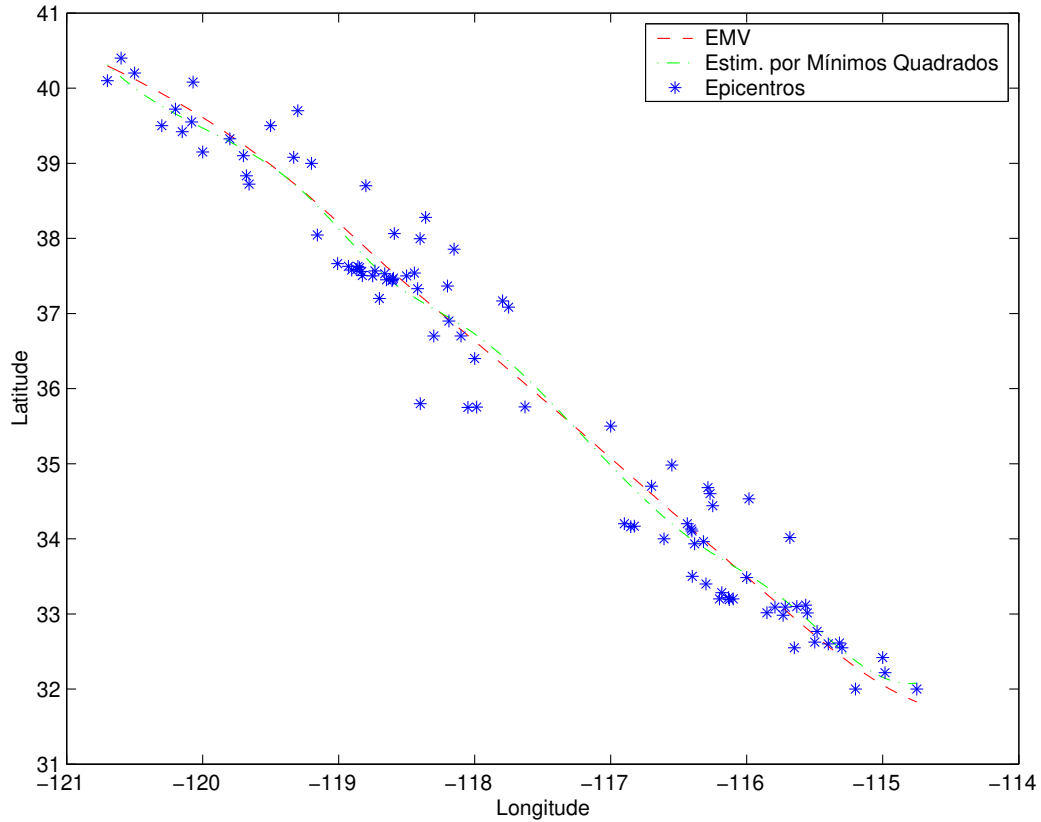


Figura 12: Gráficos das curvas estimadas, por Mínimos Quadrados e por Máxima Verossimilhança para o conjunto de dados de terremotos de Magnitude maior que 5 ocorridos na Califórnia, de 1980 a 1999.

Quadro Comparativo relativo à Figura 12

Parâmetros	EMV	Mín. Quadrados
Dist. Hellinger	0,9989	0,9999
$\lambda$	0,081	0,011
Número de nós	13	13
EMQ	0,2396	0,2280
$\sigma$	0,2554	0,2943
$\alpha$	12,40	-

Como pode ser visto na Figura 12, a falha de San Andreas, como as outras falhas, são próximas da linearidade. Assim, o processo de EMV apresentado neste trabalho não difere muito do processo de estimação por Mínimos Quadrados.

## 4.2 Conclusões

*Pela construção do processo  $N'$ , o erro aleatório embora seja unidimensional, produz um erro nas duas coordenadas. Uma outra característica desse processo é que os pontos vieram de um processo de Poisson homogêneo sobre a curva, espera-se que se aglomerem mais pontos no interior das parábolas da curva.*

*Devido a essas configurações do processo  $N'$ , a estimação por Mínimos Quadrados falha, pois esta assume que o erro ocorre somente nas ordenadas, quando na verdade os pontos vêm da direção normal à curva. Assim, pontos aninhados nos interiores das parábolas da curva verdadeira fazem com que a estimação por Mínimos Quadrados fique muito penalizada nos pontos da parábola, passando bem abaixo do vértice quando a parábola é convexa ( para baixo) e passando bem acima do vértice quando a parábola é côncava ( para cima). Veja Figuras 8 e 10.*

*Note que analisar a conjunto de dados através de EMV é utilizar o modelo correto. A intensidade e a medida média do processo  $N'$  são calculadas exatamente em função de  $f$ . Pelos exemplos dados, pode-se concluir que a EMV consegue obter bons resultados, bem melhor que a estimação por Mínimos Quadrados, na existência e localização da curvatura da função verdadeira.*

*A grande desvantagem da EMV é o custo computacional. Para cada iteração da verossimilhança, o cálculo da intensidade tem de ser feito para cada ponto do processo, pois a função  $h(u, x)$  dada em (2.22) precisa ser calculada também sobre os pontos do processo para depois serem calculados os valores dados em (2.23). Para se ter uma idéia, o tempo gasto para estimar uma curva com  $m$  bases varia entre  $3600 \text{ seg} * m$  e  $15000 \text{ seg} * m$ . Essa variação ocorre em função da curvatura da função a ser estimada, além de toda a estrutura do processo de Poisson  $N'$ .*

*Outra desvantagem que também aumenta, e muito, o custo computacional é a inexistência, no problema de EMV, da matriz de suavização  $H$ , dada em (B.2) ( ver Apêndice B), utilizada no GCV na estimação por Mínimos Quadrados. É impraticável usar Validação Cruzada, que para cada ponto do processo, deveria estimar a curva  $f$  sem este ponto ( a  $i$ -ésima observação) ( Método leave-one-out). Impraticável porque acarretaria um custo  $n$  vezes maior,  $n$  o número de pontos.*

*Com todos esses problemas de velocidade na EMVP, não pode deixar de ser mencionado a eficiência, em termos de estimação da curva que melhor representa o processo  $N'$ . O Método de Mínimos Quadrados não consegue estimar bem a curva. Esta estimação piora quando a curvatura da função é pouco suave, pois pontos do processo próximos deste local tendem a confundir sua origem em relação à curva para o modelo de Mínimos Quadrados.*

# APÊNDICE A – Splines

**Definição A.1.** *Seja*

$$\mathbf{P}_k = \left\{ p(x) : p(x) = \sum_{i=1}^k c_i x^{i-1}, \quad c_1, \dots, c_k, x \in \mathbb{R} \right\}.$$

Assim,  $\mathbf{P}_k$  é chamado de **espaço de polinômios de ordem  $k$**

com as seguintes características:

1.  $\mathbf{P}_k$  é um espaço linear de dimensão finita com uma base, o que é conveniente.
2. Polinômios são funções suaves.
3. Polinômios são fáceis de armazenar, manipular e calcular em computadores.
4. As derivadas e antiderivadas de um polinômio são polinômios cujos coeficientes podem ser encontrados algebricamente.
5. O número de zeros (raízes) de um polinômio de ordem  $k$  não pode exceder  $k - 1$ .
6. Taxas precisas de convergência podem ser dadas por aproximação de funções suaves por polinômios.

**Definição A.2.** *Sejam  $t := t_1 < \dots < t_{l+1}$ , uma seqüência estritamente crescente de pontos,  $k$  um inteiro positivo e  $P_1, \dots, P_l$ , uma seqüência de polinômios, todos de ordem  $k$ . O **polinômio por partes  $f$  de ordem  $k$**  é definido da seguinte forma:*

$$f(x) = P_i(x), \text{ se } t_i < x < t_{i+1}; \quad i = 1, \dots, l. \quad (\text{A.1})$$

Os pontos  $t_i$  são chamados os nós de  $f$ .

**Definição A.3.** *O espaço de polinômios por partes de ordem  $k$  com nós  $t_1, \dots, t_{l+1}$ , (denotado por  $\mathbf{P}_{k,t}$ ) é a coleção de todos os polinômios por partes definidos em (A.1).*

*Observação:*  $\mathbf{P}_{k,t}$  é um espaço linear de dimensão  $kl$ , pois cada um de seus elementos consiste de  $l$  polinômios por partes de ordem  $k$ .

Considere a seguinte relação:

$$S_{t_i} D^{j-1} f = 0 \text{ para } j = 1, \dots, \nu_i \text{ e } i = 2, \dots, l, \quad (\text{A.2})$$

para algum vetor  $\nu := (\nu_i)_{i=2}^l$  de inteiros não negativos, onde

$$S_\alpha f = f(\alpha^+) - f(\alpha^-), \quad (\text{A.3})$$

é a função salto.

O valor  $\nu_i$  conta o número de continuidades exigidas para  $\xi_i$ . Assim,  $\nu_i = 0$  significa que nenhuma condição de continuidade é imposta para  $t_i$ .

O subconjunto de todas  $f \in \mathbf{P}_{k,t}$  satisfazendo (A.2) para um dado vetor  $\nu$ , que é um subespaço linear de  $\mathbf{P}_{k,t}$ , é denotado por  $\mathbf{P}_{k,t,\nu}$ .

**Teorema A.1.** (Curry e Schoenberg) Para uma dada seqüência estritamente crescente  $t = (t_i)_{i=1}^{l+1}$ , e uma dada seqüência de inteiros não negativos  $\nu = (\nu_i)_{i=2}^l$  com  $\nu_i \leq k$ , para todo  $i$ , seja

$$m := k + \sum_{i=2}^l (k - \nu_i) = kl - \sum_{i=2}^l \nu_i = \dim(\mathbf{P}_{k,t,\nu})$$

e seja  $\xi := (\xi_i)_{i=1}^{m+k}$  uma seqüência não decrescente tal que

- (i)  $\xi_1 \leq \xi_2 \leq \dots \leq \xi_k \leq t_1$  e  $t_{l+1} \leq \xi_{m+1} \leq \dots \leq \xi_{m+k}$ ;
- (ii) para  $i = 2, \dots, l$ , o número  $t_i$  ocorre exatamente  $k - \nu_i$  vezes em  $\xi$ .

Então a seqüência  $B_{1,k,\xi}, \dots, B_{m,k,\xi}$  de B-splines de ordem  $k$  para a seqüência de nós  $\xi$  é uma base para  $\mathbf{P}_{k,t,\nu}$ , consideradas como funções em  $[\xi_k, \xi_{m+1}]$ .

Assim, tem-se que:

$$\text{número de condições de continuidade para } t_i + \text{número de nós para } t_i = k. \quad (\text{A.4})$$

O teorema deixa aberto a escolha dos  $k$  primeiros e os  $k$  últimos nós. Uma escolha conveniente é:

$$\xi_1 = \dots = \xi_k = t_1, \quad \xi_{m+1} = \dots = \xi_{m+k} = t_{l+1}, \quad (\text{A.5})$$

sendo  $\nu_1 = \nu_{l+1} = 0$ . Ou seja, nenhuma condição de continuidade é exigida para os pontos  $t_1$  e  $t_{l+1}$ . Isto é consistente com o fato de que as bases de B-splines fornecem uma representação válida para elementos de  $\mathbf{P}_{k,t,\nu}$  somente no intervalo  $[\xi_k, \xi_{m+1}]$ .

Geralmente os nós  $\xi$  são parcialmente coincidentes e construídos a partir de  $t$  e  $\nu$ . Maiores detalhes, veja de Boor (1978).

## A.1 Derivadas de B-splines

Os resultados dados nesta seção podem ser encontrados em Bartels (1987) ou de Boor (1978).

Inicialmente, dado um vetor de nós  $\xi := (\xi_i)$ , tem-se a primeira derivada de um B-spline:

$$\frac{d^{(1)}}{dx} B_{i,k,\xi}(x) = (k-1) \left[ \frac{B_{i,k-1,\xi}(x)}{\xi_{i+k-1} - \xi_i} - \frac{B_{i+1,k-1,\xi}(x)}{\xi_{i+k} - \xi_{i+1}} \right]. \quad (\text{A.6})$$

Aplicando (A.6), tem-se a segunda derivada de B-splines

$$\begin{aligned} \frac{d^{(2)}}{dx} B_{i,k,\xi}(x) &= (k-1)(k-2) \left[ \frac{1}{\xi_{i+k-1} - \xi_i} \left( \frac{B_{i,k-2,\xi}(x)}{\xi_{i+k-2} - \xi_i} - \frac{B_{i+1,k-2,\xi}(x)}{\xi_{i+k-1} - \xi_{i+1}} \right) \right. \\ &\quad \left. - \frac{1}{\xi_{i+k} - \xi_{i+1}} \left( \frac{B_{i+1,k-2,\xi}(x)}{\xi_{i+k-1} - \xi_{i+1}} - \frac{B_{i+2,k-2,\xi}(x)}{\xi_{i+k} - \xi_{i+2}} \right) \right] \end{aligned}$$

Supondo os nós igualmente espaçados, seja

$$d\xi = \xi_{i+1} - \xi_i. \quad (\text{A.7})$$

Assim  $\xi_{i+h} - \xi_i = hd\xi$ ,  $h$  um inteiro positivo, e as primeira e segunda derivadas de um B-spline cúbico são:

$$\begin{aligned} \frac{d^{(1)}}{dx} B_{i,4,\xi}(x) &= 3 \left[ \frac{B_{i,3,\xi}(x)}{\xi_{i+3} - \xi_i} - \frac{B_{i+1,3,\xi}(x)}{\xi_{i+4} - \xi_{i+1}} \right] \\ &= \frac{1}{d\xi} [B_{i,3,\xi}(x) - B_{i+1,3,\xi}(x)]. \end{aligned}$$

$$\begin{aligned} \frac{d^{(2)}}{dx} B_{i,4,\xi}(x) &= \frac{1}{d\xi} \left[ \frac{1}{d\xi} [2B_{i,2,\xi}(x) - B_{i+1,2,\xi}(x)] - \frac{1}{d\xi} [B_{i+1,2,\xi}(x) - B_{i+2,2,\xi}(x)] \right] \\ &= \frac{1}{(d\xi)^2} [B_{i,2,\xi}(x) - 2B_{i+1,2,\xi}(x) + B_{i+2,2,\xi}(x)]. \end{aligned}$$

Considere um vetor de nós  $\xi := (\xi_i)$  e  $f(x) = \sum_{i=0}^n E_i B_{i,K,\xi}(x)$ , segue o seguinte resultado sobre derivadas de ordens menores qua a ordem do B-spline.

Seja  $0 \leq k < K$ . Então

$$\frac{d^k}{dx^k} f(x) = \sum_{i=0}^{n+k} E_i^{[k]} B_{i,K-k,\xi}(x)$$

onde

$$E_i^{[0]} = E_i, \text{ para } i = 0, \dots, n$$

e para cada  $r = 1, \dots, k$ ,

$$E_i^{[r]} = (K - r) \frac{E_i^{[r-1]} - E_{i-1}^{[r-1]}}{\xi_{i+K-r} - \xi_i}, \text{ para } i = 0, \dots, n + r.$$

É convencional que  $E_{-1}^{[0]} = E_{-1}^1 = \dots = 0$  e  $E_{n+1}^{[0]} = E_{n+2}^1 = \dots = 0$ . Também  $E_i^{[r]} = 0$  se  $\xi_{i+K-r} = \xi_i$ .

Se os nós são igualmente espaçados, sejam  $d\xi = \xi_{i+1} - \xi_i$  e  $\xi_{i+h} - \xi_i = hd\xi$ , onde  $h$  é um inteiro positivo e  $f(x) = \sum_{i=1}^n E_i B_{i,4,\xi}(x)$ . Assim,

$$\begin{aligned} f^{(2)}(x) &= \sum_{i=1}^{n+2} E_i^{[2]} B_{i,2,\xi}(x) \\ &= \sum_{i=1}^{n+2} 2 \frac{E_i^{[1]} - E_{i-1}^{[1]}}{\xi_{i+2} - \xi_i} B_{i,2,\xi}(x) \\ &= \sum_{i=1}^{n+2} 2 \frac{3 \frac{E_i^{[0]} - E_{i-1}^{[0]}}{\xi_{i+3} - \xi_i} - 3 \frac{E_{i-1}^{[0]} - E_{i-2}^{[0]}}{\xi_{i+2} - \xi_{i-1}}}{2d\xi} B_{i,2,\xi}(x) \\ &= \sum_{i=1}^{n+2} \frac{E_i - E_{i-1}}{d\xi} \frac{E_{i-1} - E_{i-2}}{d\xi} B_{i,2,\xi}(x) \\ &= \frac{1}{(d\xi)^2} \sum_{i=1}^{n+2} (E_i - 2E_{i-1} + E_{i-2}) B_{i,2,\xi}(x). \end{aligned}$$



## APÊNDICE B – Validação cruzada

A estimação do parâmetro de suavização  $\lambda'$  em (3.10) pode ser feita por validação cruzada. Sua motivação básica se baseia em termos de predição, pois assumindo que os erros têm média zero, a verdadeira curva de regressão tem a propriedade de que, se uma observação  $Y$  é feita num ponto  $t$ , o valor de  $f(t)$  é o melhor preditor de  $Y$  em termos do erro médio quadrático.

A técnica de validação cruzada é a seguinte: para  $\lambda'$  fixo, seja  $\hat{f}_{\lambda'}^{(-i)}(x)$  a solução de (3.10) sem utilizar a  $i$ -ésima observação. A qualidade de  $\hat{f}_{\lambda'}^{(-i)}(x)$  como preditor da nova observação pode ser julgada através de como o valor de  $\hat{f}_{\lambda'}^{(-i)}(x_i)$  prediz  $Y_i$ . A função de validação cruzada é definida por

$$CV(\lambda') = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{f}_{\lambda'}^{(-i)}(x_i))^2 \quad (\text{B.1})$$

e então escolhe-se  $\lambda'$  que minimiza  $CV(\lambda')$ . Sabendo que  $\hat{f}_{\lambda'} = H(\lambda')Y$  onde

$$H(\lambda') = X(X'X + \lambda'\Phi)^{-1}X', \quad (\text{B.2})$$

sendo  $X_{i,j} = B_{i,k,\xi}(x_j)$  e a matriz  $\Phi$  dada por (3.12).

Pode-se mostrar que, sendo  $h_{ii}$  a diagonal de  $H(\lambda')$ ,

$$CV(\lambda') = \frac{1}{n} \sum_{i=1}^n \left( \frac{Y_i - \hat{f}_{\lambda'}(x_i)}{1 - h_{ii}(\lambda')} \right)^2.$$

Craven e Wahba(1979) obtiveram uma modificação do princípio de validação cruzada, a Validação Cruzada Generalizada(GCV). A idéia básica do GCV é substituir os fatores  $h_{ii}(\lambda')$  por suas médias  $\frac{1}{n} \sum_{i=1}^n h_{ii}$ , obtendo assim

$$GCV(\lambda') = \frac{1}{n} \frac{\sum_{i=1}^n (Y_i - \hat{f}_{\lambda'}(t_i))^2}{(1 - \frac{1}{n} \text{Tr}(H(\lambda')))^2} \quad (\text{B.3})$$

ou

$$GCV(\lambda') = EQM(\lambda')/[Tr(I - H(\lambda'))/n]^2$$

onde  $EQM(\lambda') = \sum_{i=1}^n (y_i - \hat{f}_{\lambda'}(x_i))^2/n$  é o erro quadrático médio,  $Tr(H(\lambda'))$  é o traço da matriz  $H(\lambda')$ , isto é, a soma dos elementos da diagonal principal de  $H$ .

## APÊNDICE C – Método de Monte Carlo

Seja  $h : \mathbb{R} \rightarrow \mathbb{R}$  e

$$I = \int_0^1 h(x) dx.$$

Para calcular o valor de  $I$ , observe que se  $U \sim U(0, 1)$  então

$$I = \mathbb{E}[h(U)]. \tag{C.1}$$

Se  $U_1, \dots, U_k$  são variáveis aleatórias com distribuição  $U(0, 1)$ , então segue que  $h(U_1), \dots, h(U_k)$  são variáveis aleatórias com média  $I$  por (C.1). Pela Lei Forte dos Grandes Números, tem-se, com probabilidade 1, que

$$\sum_{i=1}^k \frac{h(U_i)}{k} \rightarrow \mathbb{E}[h(U)] = I \text{ quando } k \rightarrow \infty.$$

Portanto, para se obter uma boa aproximação de  $I$ , é suficiente gerar números aleatórios  $U_i$  e tomar como aproximação a média observada dos valores  $h(U_i)$ . Este é o **método de Monte Carlo**, para o caso unidimensional.

Para o caso multidimensional, a idéia é análoga. Suponha  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  e seja

$$I = \int_0^1 \dots \int_0^1 h(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Da mesma forma que no caso unidimensional, a idéia de utilizar o Método de Monte Carlo para estimar a integral acima baseia-se no fato que

$$I = \mathbb{E}[h(U_1, \dots, U_n)],$$

onde  $U_1, \dots, U_n$  são variáveis aleatórias independentes com distribuição  $U(0, 1)$ .

Assim, gerando seqüências independentes  $U_1^i, \dots, U_n^i$ , com  $i = 1, \dots, k$ , cada uma com

distribuição  $U(0, 1)$  estima-se o valor de  $I$  através da média

$$\sum_{i=1}^k \frac{h(U_1^i, \dots, U_n^i)}{k}.$$

Quando os intervalos de integração não forem  $[0, 1]$ , basta aplicar o teorema de mudança de variáveis. Para ver mais detalhes e exemplos, consulte Machado (1998).

# APÊNDICE D – Processos Pontuais Univariados

**Definição D.1.** Dado um espaço de probabilidade  $(\Omega, \Phi, P)$ , todo evento  $A$  tal que  $P(A) = 0$  é chamado **nulo**.

**Definição D.2.** Um espaço de probabilidade  $(\Omega, \Phi, P)$  é chamado de **completo** se todos os subconjuntos nulos são eventos.

**Definição D.3.** Dado um espaço de probabilidade completo  $(\Omega, \Phi, P)$ , uma família  $\Phi_t, t \geq 0$  de sub- $\sigma$ -álgebras de  $\Phi$  é dita ser uma **filtragem** se esta é crescente, isto é,  $\Phi_t$  é  $\sigma$ -álgebra,  $\Phi_t \subseteq \Phi$ , para todo  $t \geq 0$  e  $\Phi_s \subseteq \Phi_t$ , para todo  $s < t$ .

**Definição D.4.** Seja  $(\Omega, \Phi, P)$  um espaço de probabilidade completo e  $\{\Phi_t, t \geq 0\}$  uma filtragem. Uma variável aleatória  $\tau : \Omega \rightarrow [0, \infty]$  é chamada de **tempo de Markov (ou tempo de parada)** relativo à filtragem  $\{\Phi_t, t \geq 0\}$  se, para todo  $t \geq 0$ , tivermos

$$\{w : \tau(w) \leq t\} \in \Phi_t.$$

**Definição D.5.** Seja  $(\Omega, \Phi, P)$  um espaço de probabilidade completo,  $\{\Phi_t, t \geq 0\}$  uma filtragem contínua à direita (i. é,  $\Phi_t = \bigcap_{s>t} \Phi_s$ ) tal que  $\Phi_0$  contém todos os conjuntos de probabilidade zero de  $\Omega$ . Um **processo pontual simples** em  $[0, \infty)$  é uma sequência  $\mathbb{T} = \{\tau_n, n \geq 1\}$  de tempos de Markov com relação à filtragem  $\{\Phi_t, t \geq 0\}$ , que satisfaz, com probabilidade 1, as seguintes propriedades:

1.  $\tau_1 > 0$ ;
2.  $\tau_n < \tau_{n+1}$ , se  $\tau_n < \infty$ ;
3.  $\tau_n = \tau_{n+1}$ , se  $\tau_n = \infty$ .

O processo  $\mathbb{T} = \{\tau_n, n \geq 1\}$  é completamente caracterizado pelo processo de contagem  $N = \{N(t) = N((0, t]), t \geq 0\}$ , onde

$$N(t) = \sum_{n=1}^{\infty} \mathbb{I}_{[\tau_n, \infty)}(t), \quad t \geq 0,$$

pois  $\{N(t) \geq n\} = \{\tau_n \leq t\}$  para todo  $n \in \mathbb{N}$  e todo  $t \geq 0$ , ou seja, é possível reconstruir um processo pontual a partir de seu processo de contagem associado, e vice-versa.

Uma forma mais intuitiva de se definir processos pontuais é a seguinte: Suponha um experimento começando no instante  $t = 0$ . Eventos de uma certa classe ocorrem aleatoriamente, o primeiro ocorrendo em  $\tau_1$ , o segundo em  $\tau_2$ , etc., onde  $\tau_1, \tau_2$ , etc. são variáveis aleatórias não negativas. Os valores  $t_i$  de  $\tau_i$ ,  $i = 1, 2, \dots$  serão chamados **pontos de ocorrência** ou eventos. Seja  $(s_1, s_2]$  um subintervalo do intervalo  $[0, s]$  onde  $s < \infty$ . Denote por  $N(s_1, s_2]$  o número de pontos de ocorrência em  $(s_1, s_2]$ . Então  $N(s_1, s_2]$  é uma variável aleatória e a coleção de tais variáveis aleatórias, abreviadas por  $N$ , para todos subintervalos de  $[0, s]$  é chamada de **processo de contagem** em  $[0, s]$ .

Observe que as trajetórias destes processos são, com probabilidade 1, contínuas à direita, possuem limites à esquerda e são funções escadas com saltos de comprimento 1.

As  $\sigma$ -álgebras  $\Phi_t$  que aparecem na definição do processo de contagem podem ser interpretadas como toda a informação disponível até o instante  $t$ , isto é, elas contém todos os eventos observáveis em  $[0, t]$ .

Define-se a seguir os parâmetros de processos pontuais.

$$M(t) := \mathbb{E}(N(t))$$

onde  $\mathbb{E}$  denota esperança com respeito à  $\mathbb{P}$ ;  $M(t)$  será uma função não decrescente contínua à direita. Se a derivada de  $M(t)$  existe, define

$$\mu_1(t) := M'(t).$$

Assim,  $\mu_1(t)$  pode ser interpretado como a razão instantânea de mudança do número esperado de ocorrências com respeito ao tempo ou **razão de chegadas do processo**.

Outro parâmetro de interesse para processos pontuais é a chamada intensidade

$$\mu(t) = \lim_{\Delta \rightarrow 0} \frac{P(N(t - \Delta, t] \geq 1)}{\Delta}.$$

Thompson (1988) afirma que se  $\mu(t)$  existe, então para  $\Delta$  pequeno,  $\Delta > 0$ ,  $\lambda(t)\Delta$  é aproximadamente a probabilidade de uma ou mais chegadas no intervalo  $(t - \Delta, t]$ . Além disso, se não existem chegadas simultâneas no processo, então a razão de ocorrências é igual

à intensidade, isto é,  $\mu_1(t) = \mu(t)$ , pois

$$M(t) = \int_0^t \mu(x) dx.$$

## D.1 Processos Pontuais de Poisson

**Definição D.6.** Um processo pontual tem **incrementos independentes** se para todo  $l_1 < u_1 \leq l_2 < u_2 \leq \dots \leq l_n < u_n$  e para todo  $n$ , tem-se

$$P(N(l_1, u_1] = k_1, \dots, N(l_n, u_n] = k_n) = \prod_{i=1}^n P(N(l_i, u_i] = k_i).$$

**Definição D.7.** Um processo pontual é dito ser **homogêneo** ou ter **incrementos estacionários** se  $P(N(t, t+x] = k)$  não depende de  $t$ .

**Definição D.8.** Um processo com incrementos independentes satisfazendo

$$P(N(t, t+x] = k) = \frac{(\alpha x)^k e^{-\alpha x}}{k!} \quad k = 0, 1, 2, \dots$$

com  $0 < \alpha < \infty$ , é dito ser um **processo de Poisson homogêneo** com intensidade  $\alpha$ . Para este processo,  $M(t) = \alpha t$  e  $\mu(t) = \alpha$ .

Existem outras duas caracterizações de processos de Poisson. Provas desses resultados podem ser encontradas em Thompson (1988).

**Teorema D.1.** Somente processos pontuais não-degenerados com incrementos estacionários e independentes e sem ocorrências simultâneas são processos de Poisson homogêneos.

**Teorema D.2.** Um processo pontual é um processo de Poisson homogêneo com intensidade  $\alpha$  se, e somente se, os intervalos dos tempos de ocorrência são exponencialmente distribuídos com média  $1/\alpha$ .

Seguem dois resultados interessantes de processos de Poisson que podem auxiliar na compreensão e em sua simulação.

**Teorema D.3.** Um processo de Poisson homogêneo em um intervalo finito de comprimento  $l$  pode ser realizado em dois estágios: 1. gerar o número de ocorrências,  $m$ , em um intervalo de acordo com a distribuição de Poisson com parâmetro  $\alpha l$ ; 2. escolher  $m$  pontos no intervalo independentemente de acordo com a distribuição uniforme nesse intervalo.

**Corolário D.1.** Seja  $N$  um processo de Poisson homogêneo, denote  $Y_1, \dots, Y_k$   $k$  ocorrências de  $N$  em um intervalo fixo. Então o vetor  $Y_1, \dots, Y_k$  tem a mesma distribuição das estatísticas de ordem para a amostra de tamanho  $k$  de uma distribuição uniforme naquele intervalo.

## D.2 Processos de Poisson não Homogêneos

**Definição D.9.** Um processo de Poisson não homogêneo é um processo pontual  $\{N(t), t \geq 0\}$ , definido em um espaço de probabilidade  $(\Omega, \Phi, P)$  que satisfaz as seguintes condições:

1.  $P(N(0) = 0) = 1$ ;
2.  $\{N(t), t \geq 0\}$  possui incrementos independentes;
3.  $P(N(t+h) - N(t) = 1 | F_{t-}^N) = \mu(t)h + o(h)$ ; e
4.  $P(N(t+h) - N(t) \geq 2 | F_{t-}^N) = o(h)$ ,

onde  $\mu : [0, \infty) \rightarrow [0, \infty)$  é a função intensidade do processo e

$$F_{t-}^N = \sigma(N(s) : s < t).$$

Note que se  $\mu(t) = \mu$ ,  $\{N(t), F_t^N, t \geq 0\}$  é um processo de Poisson homogêneo.

A Condição 2 significa que  $N(b) - N(a)$  e  $N(d) - N(c)$  (para  $a < b$  e  $c < d$ ) são variáveis aleatórias independentes se os intervalos  $(a, b]$  e  $(c, d]$  forem disjuntos, ou ainda, que  $N(t) - N(s)$  é independente de  $F_s^N = \sigma\{N(u) : u \leq s\}$ , isto é,  $P(A \cap B) = P(A)P(B)$ ,  $\forall A \in \sigma(N(t) - N(s))$ ,  $s < t$  e  $B \in F_s^N$ .

A Condição 4 significa que a probabilidade de ocorrerem pelo menos 2 eventos simultaneamente é zero (isto é,  $N(t) - N(t-) = 0$  ou 1). Por este motivo,  $N$  é dito ser **simples**.

A seguir são dados alguns resultados sobre processos de Poisson não homogêneos. Demonstrações podem ser encontradas em Gomes (1991).

1. Um processo de Poisson não homogêneo satisfaz

$$P(N(t) - N(s) = k) = \frac{\exp\left(-\int_s^t \mu(u)du\right) \left[\int_s^t \mu(u)du\right]^k}{k!}, k = 0, 1, \dots$$

e  $0 \leq s \leq t$ .

2. Seja  $N = \{N(t), t \geq 0\}$  um processo de Poisson não homogêneo com função intensidade  $\mu(\cdot)$ . Então, se

$$Z(t) = \inf\{v : \Lambda(v) > t\} \text{ e } \Lambda(t) = \int_0^t \mu(s)ds,$$

o processo  $\tilde{N} = \{N(Z(t)), t \geq 0\}$  é um processo de Poisson homogêneo com função intensidade  $\mu \equiv 1$ .



3. Sejam  $0 < T_1 < T_2, \dots$  os sucessivos tempos de ocorrência de saltos de um processo de Poisson não homogêneo com função intensidade  $\mu$ . Então para qualquer  $n$ ,

$$P(T_{n+1} - T_n > s | T_1 = t_1, \dots, T_n = t_n) = \exp[-(\Lambda(t_n + s) - \Lambda(t_n))],$$

isto é, o tempo  $X_{n+1} = T_{n+1} - T_n$  e  $T_0 = 0$  decorrido entre dois saltos consecutivos possui função densidade de probabilidade dada por

$$f_{X_{n+1}|T_1, \dots, T_n}(x_{n+1}|t_1, \dots, t_n) = \mu(t_{n+1}) \exp\left[-\int_{t_n}^{t_{n+1}} \mu(u) du\right]$$

4. Dado que  $N(1) = n$ , os tempos de ocorrência dos  $n$  primeiros eventos  $0 < T_1 < \dots < T_n \leq 1$  são distribuídos como  $n$  estatísticas de ordem de uma amostra de  $n$  observações da densidade  $f(u) = \frac{\mu(u)}{\Lambda(1)}$ ,  $0 \leq u \leq 1$ . Ou seja,

$$f_{T_1, \dots, T_n | N(1)=n}(t_1, \dots, t_n) = \frac{n! \prod_{i=1}^n \mu(t_i)}{\left[\int_0^1 \mu(s) ds\right]^n}, \quad 0 < t_1 < \dots < t_n \leq 1.$$

# *APÊNDICE E – Estimação de Máxima Verossimilhança Penalizada*

**Definição E.1.** Dada a amostra  $X_1, \dots, X_n$  em  $\Phi$ , a log-verossimilhança  $\phi$ -penalizada de  $\theta \in \Theta$  é definida por:

$$\ell_{n,\lambda}(\theta) = \ell_n(\theta) - \lambda\phi(\theta), \quad \lambda > 0 \tag{E.1}$$

onde

$$\ell_n(\theta) = \log \prod_{i=1}^n \frac{\partial P_\theta}{\partial \mu}(x_i) = \log \prod_{i=1}^n p(x_i, \theta) = \sum_{i=1}^n \log p(x_i, \theta).$$

**Definição E.2.** Seja o seguinte problema de otimização, para cada  $\lambda > 0$

$$\begin{cases} \text{maximizar } \ell_{n,\lambda}(\theta), \text{ sujeito a} \\ \theta \in \Theta, \Theta \subset H \end{cases} \tag{E.2}$$

Uma solução de (2.3) será dita um estimador de máxima verossimilhança penalizada (EMVP) baseada na amostra  $X_1, \dots, X_n$  correspondente à  $\theta$  e à função de penalidade  $\phi$ , e denotada por  $\theta_\lambda$ .

Um EMVP será então uma curva que maximize a log-verossimilhança penalizada sobre a classe de todas as curvas satisfazendo condições de regularidade onde  $\phi(\theta)$  é definida. Quando  $H$  for espaço de Hilbert, uma função de penalidade natural é  $\phi(\theta) = \|\cdot\|^2$ , onde  $\|\cdot\|$  denota a norma em  $H$ .

A idéia básica para se utilizar verossimilhança penalizada na estimação de uma curva é composta de dois objetivos: maximizar a adequação aos dados medido pela verossimilhança e evitar curvas "pouco suaves" ou que mostrem uma variabilidade muito rápida.

A "suavidade" de uma curva  $f$  pode ser medida por um funcional  $\phi(f)$  de várias formas. A mais usada é fazer  $\phi(f) = \int f''(t)^2 dt$ , que dará um valor grande se  $f$  tiver uma grande

curvatura local e um valor zero será dado caso  $f$  seja uma reta. Maximizar  $\ell_{n,\lambda}(\theta)$  representa então um critério de compromisso entre os dois objetivos. Assim o possível problema de obter um valor grande de  $\ell(\theta)$  e também valores excessivos de  $\phi(\theta)$  é resolvido subtraindo da verossimilhança um múltiplo de  $\phi(\theta)$ . O parâmetro  $\lambda$  é chamado de parâmetro de suavização e controla o balanço entre suavidade e bondade de ajuste. O grau de suavização cresce quando  $\lambda$  cresce, resultando um EMVP mais suave.

Em alguns casos, o estimador de máxima verossimilhança não existe. Para o caso de EMVP, existem alguns teoremas que, sobre certas condições, garantem a existência e até a unicidade de maximizadores de  $\ell_{n,\lambda}(\theta)$ ,  $\lambda > 0$ . Veja mais detalhes em Gimenez (1993) que exhibe estes teoremas.

## APÊNDICE F – Espaço de Funções

No contexto de regressão de curvas, paramétrica ou não paramétrica, supõe-se que a curva pertença a um espaço de funções. Geralmente assume-se que a curva  $f$  é uma combinação linear de funções que formam uma base desse espaço.

**Definição F.1.** Denomina-se  $L_2[a, b]$  o espaço vetorial das funções definidas em  $[a, b]$  de quadrado integrável, isto é,  $\int f^2(t)dt < \infty, \forall f \in L_2[a, b]$ . Seu produto interno é definido por

$$\langle f, g \rangle = \int_a^b f(t)g(t)dt, \forall f, g \in L_2[a, b],$$

sendo portanto sua norma

$$\| f \| = \left( \int_a^b f^2(t)dt \right)^{1/2}.$$

O espaço  $L_2[a, b]$  representa uma rica coleção de funções. Entretanto, ele não é totalmente apropriado no contexto de regressão. Duas funções,  $f, g \in L_2[a, b]$  são consideradas idênticas em  $L_2[a, b]$  se  $\| f - g \| = 0$ . Assim,  $f$  e  $g$  são idênticas se elas diferem apenas num conjunto de pontos em  $[a, b]$  cuja a medida de Lebesgue seja nula. Isso significa que os elementos de  $L_2[a, b]$  são classes equivalentes de funções que são iguais quase em toda a parte com respeito à medida de Lebesgue. Conseqüentemente, a avaliação de um elemento de  $L_2[a, b]$  para pontos em  $[a, b]$  não é uma operação bem definida.

No contexto de regressão, é razoável assumir que a função  $f$  em (3.8) seja uma função suave. Logo, uma melhor escolha para a função de regressão seria o espaço de Sobolev de  $m$ -ésima ordem

$$W_2^k[a, b] = \left\{ \begin{array}{l} f : f^{(j)} \text{ é absolutamente contínua,} \\ j = 0, \dots, k-1 \text{ e } f^{(k)} \in L_2[a, b] \end{array} \right\},$$

que é um subconjunto de  $L_2[a, b]$ .

**Definição F.2.** Duas funções  $f, g \in L_2[a, b]$  são **ortogonais** se o produto interno entre  $f$  e  $g$  é zero, isto é,  $\langle f, g \rangle = 0$ .

**Definição F.3.** Uma seqüência de funções  $\{\phi_j\}$  é dita ser **ortonormal** se

$$\langle \phi_i, \phi_j \rangle = \begin{cases} 0 & \text{se } i \neq j \\ 1 & \text{se } i = j \end{cases}$$

e  $\|\phi_j\| = 1$  para todo  $j$ , isto é, uma seqüência ortonormal é também ortogonal.

**Definição F.4.** Uma seqüência de funções  $\{\phi_j\}$  é dita ser **completa** se a única função  $g$  que satisfaz  $\langle g, \phi_j \rangle = 0$ , para todo  $j$  é a função nula, isto é,  $g \equiv 0$ .

**Proposição F.1.** : Seja  $\{\phi_j\}_{j=1}^{\infty}$  uma seqüência completa e seja  $f \in L_2[a, b]$ . Defina  $\beta_j = \langle f, \phi_j \rangle$ ,  $j = 1, 2, \dots$ , então  $\sum_{j=1}^K \beta_j \phi_j$  é a melhor aproximação de  $f$  no sentido que  $\|f - \sum_{j=1}^K \beta_j \phi_j\| \leq \|f - \sum_{j=1}^K \alpha_j \phi_j\|$  para todo  $\alpha = (\alpha_1, \dots, \alpha_K)^t \in \mathbb{R}^K$ . Mais ainda,  $\|f - \sum_{j=1}^K \beta_j \phi_j\| \rightarrow 0$ , quando  $k \rightarrow \infty$ .

Esta proposição afirma que pode-se aproximar  $f$  por  $\sum_{j=1}^k \beta_j \phi_j$  quando  $k \rightarrow \infty$  e como a norma (em  $L_2[a, b]$ ) entre  $f$  e  $\sum_{j=1}^k \beta_j \phi_j$  converge para zero, pode-se dizer que esse dois elementos são idênticos em  $L_2[a, b]$ . Devido à igualdade de Parseval,  $\sum_{j=0}^{\infty} |\beta_j|^2 = \|f\|^2$ , pode-se dizer que se  $\{\beta_j\}$  é uma seqüência tal que  $\sum_{j=0}^{\infty} \beta_j^2 < \infty$  então qualquer  $f \in L_2[a, b]$  pode ser escrita como  $f = \sum_{j=0}^{\infty} \beta_j \phi_j$ , uma combinação linear de funções  $\{\phi_j\}$  que formam uma seqüência ortonormal completa para  $L_2[a, b]$ .

Com isto em mente, basta escolher uma base e estimar, por algum critério, os valores dos parâmetros  $\beta$  e então determinar a estimativa da função de regressão  $f$ .

## APÊNDICE G – Programas

```
function result=GCV
% Programa principal, calcula curva de regressão pelos métodos de spline (usual) e por
máxima verossimilhança

global x y n a b;

tic % inicia contador de tempo
% Leitura dos Dados

A=load('dados.txt'); % matriz 2Xn, onde a primeira linha são as abscissas dos dados e
a segunda linha são as ordenadas

A=A';

x=A(1,:);
y=A(2,:);

n=length(x); % Número de pontos do processo N
a=x(1); % Extremo esquerdo do intervalo
b=x(n); % Extremo direito do intervalo

% Estimação da curva usando número de bases inicial
d=.995; % Valor mínimo da Distância de Hellinger
h2=0;
ndx=4;
knot=knots(x,a,b,ndx); % numero de nós=ndx+8 ; numero de bases=ndx+4
m=length(knot);

beta01=gcv2(ndx); % chute inicial do vetor  $\beta$  ( a solução por Mínimos Quadrados) para
o determinado número de bases (ndx + 4)
```

```

beta0=beta01(1:end-1);
lsp=beta01(end);
spt=spmak(knot,beta0');
sigma0=vari(spt);
L=quadr(@raiz,a,b,[],[],spt); % Comprimento do arco da função definida no intervalo
[a,b]
alfa0=n/L;
lmin=lambdamin(spt,lsp,alfa0,sigma0) % Calcula o valor de  $\lambda$  na EMV do processo  $N'$ ,
a partir do  $\lambda$  encontrado no processo de estimação por Mínimos Quadrados
beta0=[beta0;alfa0;sigma0];
beta10=maxvero(lmin,beta0,knot,m); % EMVP da função de verossimilhança
beta1=beta10(1:end-2);
sp=spmak(knot,beta1');
f1=fnval(sp,x);
i1=norm(f1);
while h2 ≤ d % condição de parada: distância de Hellinger ≥ d
mk=ndx+1; Estimação da curva usando número de bases aumentado de 1 (uma) unidade
knot=knots(x,a,b,mk);
m=length(knot);
beta01=gcv2(mk);
beta0=beta01(1:end-1);
lsp=beta01(end);
spt=spmak(knot,beta0');
sigma0=vari(spt);
L=quadr(@raiz,a,b,[],[],spt); % Comprimento do arco da função definida no intervalo
[a,b]
alfa0=n/L;
lmin=lambdamin(spt,lsp,alfa0,sigma0); beta0=[beta0;alfa0;sigma0];
beta10=maxvero(lmin,beta0,knot,m); % EMVP da função de verossimilhança

```

```

beta=beta10(1:end-2);
sp=spmak(knot,beta');
f2=fnval(sp,x);
i2=norm(f2);

% distância de Hellinger entre as duas funções estimadas usando número de bases atual
e número de bases no passo anterior
h2=sum(abs(f1.*f2))/(i1*i2)
ndx=ndx+1;
f1=f2;
i1=i2;
end
knot Poisson=m
x1=linspace(a,b,500);
% curva estimada usando EMV do processo de Poisson N'
fpois=fnval(sp,x1);
% Estimação por Mínimos Quadrados utilizando H-splines: núm. de nós,  $\lambda$ , Dist. de
Hellinger e curva estimada
[kmq lmq h2mq fmq]=gcvs(x1,h);
plot(x1,fpois,'r-',x1,fspli','g-',x,y,'b*')
h2 MinQuad=h2mq
lambda MinQuad=lmq
num nós MinQuad=kmq
[kmq lmq h2mq fmq]=gcvs(x,h);
EQMMinQuad = (norm(y - fmq'))^2/n % EQM pelo método de Mínimos Quadrados
h2 Poisson=h2
lambda Poisson=lmin
num de nós Poisson=m
fpois=fnval(sp,x);

```



```
EQMPoisson = (norm(y - fpois))^2/n; % EQM pelo método de EMV do processo de Poisson N'
```

```
toc % finalizador de contagem de tempo de execução do programa principal
```

```
function r=raiz(t,sp)
```

```
% Comprimento do arco da função definida no intervalo [a,b]
```

```
f1=fnval(fnder(sp,2),t);
```

```
r=sqrt(1+(f1).^2); %  $\sqrt{1 + f'(x)^2}$ 
```

```
function s=vvari(sp)
```

```
% Calcula a variância do processo pontual
```

```
global a b y n
```

```
z=zeros(1,n);
```

```
for i=1:n
```

```
[xm,z(i)]=fminbnd(@dist,a,b,[],i,sp); % xm seria a abscissa do PP homogêneo
```

```
end
```

```
s=sqrt(sum(z.^2)/(n - 1));
```

```
function d=dist(t,i,sp)
```

```
% Distância entre os pontos do processo e a curva
```

```
global x y
```

```
yt=fnval(sp,t);
```

```
d=sqrt((x(i)-t).^2 + (y(i) - yt).^2);
```

```
function m=knots(x,a,b,ndx)
```

```
% Cálculo do vetor de nós
```

```
c=1/(ndx+1);
```

```
knot=linspace(c,1-c,ndx);
```

```
knot=prctile(x',100*knot);
```

```

m=[a a a a knot b b b b];

function l=lambdamin(sp0,p,lambda0,sigma0)
    % Calcula o valor de  $\lambda$  na EMV do processo  $N'$ , a partir do  $\lambda$  encontrado no processo
    de estimação por Mínimos Quadrados
    global x y
    p0=penalty(sp0);
    SP=(1 - p) * sum((y - fnval(sp0,x)).^2) + p * p0;
    vero=abs(intlamb(sp0,lambda0,sigma0)-soma(sp0,lambda0,sigma0));
    l=p*vero/(SP+p*vero);

function betamin=maxvero(p,beta0,knot,m)
    % Maximização da função de verossimilhança
    options=optimset('MaxIter',300*(m-4),'MaxFunEvals',5000*(m-4));
    betamin=fminsearch(@vero,beta0,options,p,Om,knot);

function logf=vero(beta,p,knot)
    % função de verossimilhança a ser maximizada
    betat=beta(1:end-2);
    lambda=beta(end-1);
    sigma=beta(end);
    sp=spmak(knot,betat');
    logf=-(1-p)*(intlamb(sp,lambda,sigma)-soma(sp,lambda,sigma))+p*penalty(sp);

function s=soma(sp,lambda,sigma)
    % Medida média do Processo
    global a b
    warning off;
    u=b+5;

```

```

z=sort(unifrnd(a,b,1,1000));
f=fnval(sp,z);
der=fnval(fnder(sp),z);
a1=find(der > 0);
a2=find(der ≤ 0);
g=sqrt(1+der.^2);
v=max(f)+4*sigma;
zu=(z-u).*g./der;
zv=(v-f).*g;
c=min(zu,zv);
f2=zeros(1,1000);
f2a=normcdf(zv(a1),0,sigma)-normcdf(zu(a1),0,sigma);
f2(a1)=max(f2a,0);
f2(a2)=normcdf(c(a2),0,sigma);
s=lambda/(sqrt(2*pi)*sigma)*(sum(g.*f2))/1000;

```

```

function m=intlamb(sp,lambda,sigma)
% funcao intensidade do processo
global a b x y n
x00=mean(fnzeros(fnder(sp),[a,b])); % Aj: vetor de derivadas nulas de f
x0=[a x00 b];
n2=length(x0)-1;
a1=x0(1:n2)+diff(x0)/2;
fa1=fnval(fnder(sp),a1);
ind1=fa1 > 0; % funcao indicadora: f'(Aj > 0) (vetor)
mu1=zeros(1,n);
mut=interv(a,b,sp);
for i=1:n

```

```

ind3=x0(1:end-1) < x(i); % funcao indicadora: x0(k) < x(i)
ind4=x0(2:end) > x(i); % funcao indicadora: x0(k+1) > x(i)
mu0=find(mut(2,:)==i); % interv(a,b,sp,x(i),y(i));
if length(mu0)==1
mu=mut(1,mu0);
mu1(i)=sum(g(mu,sp,i,sigma));
elseif length(mu0) > 1
mu=mut(1,mu0);
mu00=zeros(1,n2);
for k=1:n2
c=find(x0(k) < mu & mu ≤ x0(k+1));
if length(c)==1
g0=g(mu(c),sp,i,sigma);
mu00(k)=g0*(1-ind1(k)*(1-ind3(k))); % 1-ind(f'(A_j > 0))*ind(x0(k) > x(i))
elseif length(c) > 1
g0=g(mu(c),sp,i,sigma);
a11=mu(c(1:end-1))+diff(mu(c))/2; % ponto medio de mu
h1=hu(a11,sp,x(i),y(i));
ind2=h1 > 0; % funcao indicadora: h_x(A_j k(x(i), y(i))) > y(i)
ind5=ind2*(1-ind1(k))+(1-ind2)*ind1(k);
mu00(k)=sum(diff(g0).*ind5)+(g0(1)-g0(end))*(1-ind4(k))+g0(1)*ind3(k)*ind4(k);
end
end
mu1(i)=sum(mu00);
end
end
par=lambda/(sqrt(2*pi)*sigma);
t=find(mu1 > 0);

```

```

mu1=mu1(t);
m=sum(log(par*mu1));

```

```

function p=penalty(sp)
% Penalty: integral do quadrado da segunda derivada de 'f'
global a b
t=sort(unifrnd(a,b,1,500));
p=(b-a)*sum((fnval(fnder(sp,2),t).^2))/500;

```

```

function lam=g(mu,sp,i,sigma)
% Calcula segundas derivadas da medida media
global y
warning off;
f11=fnval(sp,mu);
der1=1+(fnval(fnder(sp),mu)).^2;
der2=fnval(fnder(sp,2),mu);
derparc=der1+der2.*(f11-y(i));
lam=der1.*exp(-(f11-y(i)).^2.*(der1/(2*sigma^2)))./derparc;

```

```

function h=hu(t,sp,x,y)
% calcula a altura da normal no vetor ponto t relativa ao eixo vertical x quando essa toca
o eixo horizontal y

```

```

f=ones(length(x),1)*fnval(sp,t);
der=ones(length(x),1)*fnval(fnder(sp),t);
h=zeros(length(x),length(t));
h=f+(ones(length(x),1)*t'*ones(1,length(t)))./(der+eps)-y'*ones(1,length(t));

```

```

function g=gcv2(ndx)

```

```

% Programa para estimar a curva f, por Mínimos Quadrados, para um determinado
número de bases ndx

```

```

global x y n a b
knot=knots(x,a,b,ndx);
m=length(knot);
l=1/ndx^4;
X=spcol(knot,4,x);
Om=omega(ndx,a,b);
beta0=(X'*X+l*Om)\X'*y';
y1=X*beta0;
lambda1=fminbnd(@gcv1,0.001,.999,[],y1',X,Om,y);
beta0=(X'*X+lambda1*Om)\X'*y';
g=[beta0;lambda1];

```

```

function g=gcv(x1,h)
% Programa para estimar a curva f, por Míminos cuadrados via H-splines
global x y n a b
h2=0;
ndx=3;
knot=knots(x,a,b,ndx);
m=length(knot);
l=1/ndx^4;
X=spcol(knot,4,x);
Om=omega(ndx,a,b);
beta0=(X'*X+l*Om)\X'*y';
y1=X*beta0;
lambda1=fminbnd(@gcv1,0.001,.999,[],y1',X,Om,y);
beta0=(X'*X+lambda1*Om)\X'*y';
y1=X*beta0;
i1=norm(y1);

```

```

while h2 ≤ h
mk=ndx+1;
knot1=knots(x,a,b,mk);
m=length(knot1);
X=spcol(knot1,4,x);
Om=omega(mk,a,b);
beta0=(X'*X+l*Om)X' * y');
y2=X*beta0;
lambda1=fminbnd(@gcv1,0.001,.999,[],y2',X,Om,y);
beta=(X'*X+lambda1*Om)X' * y');
y2=X*beta;
i2=norm(y2);
h2=sum(abs(y1.*y2))/(i1*i2);
ndx=ndx+1;
y1=y2;
i1=i2;
end
knot spline=m
lambda spline=lambda1
h2 spline=h2
X1=spcol(knot1,4,x1);
g=X1*beta;

```

```

function l=gcv1(lam,ya,X,Om,y) % GCV
n=length(y);
A=pinv(X'*X+lam*Om); % n * l^2, nao faz diferenca na otimiz.
l=norm(y-ya)/(n-trace(X*A*X'));
function omega=omega(ndx,a,b)
% Matriz Omega da solucao por splines
c=1/(ndx+1);
knot=linspace(c,1-c,ndx);
t=linspace(a,b,500);
knot=prctile(t',100*knot);
knots=[a a a a knot b b b b];
e=spcol(knots,2,t);
m=length(knots);
base=zeros(500,m-4);
base(2:499,2)=-6*(e(2:499,3)/(knots(5)-knots(3))-e(2:499,4)/(knots(6)-knots(4)))
/(knots(6)-knots(3));
for j=3:m-6
base(2:499,j)=6*((e(2:499,j)/(knots(j+2)-knots(j))-e(2:499,j+1)/(knots(j+3)-knots(j+1)))
/(knots(j+3)-knots(j))-
(e(2:499,j+1)/(knots(j+3)-knots(j+1))-e(2:499,j+2)/(knots(j+4)-knots(j+2)))
/(knots(j+4)-knots(j+1)));
end
base(2:499,m-5)=6*((e(2:499,m-5)/(knots(m-3)-knots(m-5))-e(2:499,m-4)
/(knots(m-2)-knots(m-4)))/(knots(m-2)-knots(m-5))-
e(2:499,m-4)/(knots(m-2)-knots(m-4))/(knots(m-1)-knots(m-4)));
omega=(b-a)*base'*base/500;

function beta=betai(k,X,Om,y)

```



```
% Cálcula o vetor  $\beta$  da solução por splines, pelo GCV  
b0=(X'*X+k*Om)X' * y');  
y1=X*b0;  
lambda1=fminbnd(@gcv1,0,k,[],y1',X,Om,y);  
beta=(X'*X+lambda1*Om)X' * y');
```

# Referências Bibliográficas

- [1] Bartels, R. H., Beatty, J. C., Barsky, B. A. (1987). An Introduction to Splines for use in Computer Graphics and Geometric Modeling, *Morgan Kaufmann Publishers, Inc., California*.
- [2] Bustos, O. H. e Orgambide, A. C. F. (1992). Simulação Estocástica: Teoria e Algoritmos, 10º SINAPE, Rio de Janeiro.
- [3] Choi, E. and Hall, P. (1999). Nonparametric Approach to Analysis of Space-time Data on Earthquake Occurrences, *Journal Computation and Graphical Statistics, Volume 8, Number 4, 733-748*.
- [4] Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions, *Numerische Mathematik*, 8: 377-403.
- [5] Daley, D. J. and Vere-Jones, D. (1988). An Introduction to the Theory of Point Process, *Springer Verlag, New York*.
- [6] de Boor, C. (1978). A Pratical Guide to Splines, *Springer Verlag, New York*.
- [7] Dias, R. (1994). Density Estimation via H-splines, *University of wisconsin-Madison, Ph.D. dissertation*.
- [8] Dias, R. (1999). Sequential Adaptive Nonparametric Regression via H-splines, *Communications and Simulations*, 28, 501-515.
- [9] Dias, R. (2001). Regressão Não Paramétrica, *Unicamp, Campinas*.
- [10] Dudewicz, E. J. and Mishra, S. N. (1988). Modern Mathematical Statistics, *John Wiley & Sons, New York*.
- [11] Eubank, R. L. (1988). Spline Smoothing and Nonparametric Regression, *Marcel Dekker, Inc.*
- [12] Garcia, N. L. (1995). Maximum Likelihood Estimation and Suboptimal Stopping Time for a Poisson Point Process, *REBRAPE*, 9 , 67-82.
- [13] Gimenez, P. C. (1993). Estimacão de Máxima Verossimilhança não Paramétrica pelos Métodos de Grenander e de Máxima verossimilhança Penalizada, *Unicamp*.
- [14] Gomes, A. E. (1991). Estimacão da Função Intensidade do Processo de Poisson não Homogêneo pelo Método de Grenander, *Unicamp*.

- [15] *Kutoyants, Yu. A. (1979)*. Intensity parameter estimation of an inhomogeneous Poisson process, *Prob. Control Inform. Theory*, 8, 137-149.
- [16] *Machado, F. (1998)*. *Mathematica para a Probabilidade e os Sistemas de Partículas*, IME-USP, São Paulo.
- [17] *Ogata, Y. (1988)*. Statistical Models for Earthquake Occurrences and Residuals Analysis for Point Process, *Journal of American Statistical Association*, 83, 9-27.
- [18] *Olivares, B. O. A. (1991)*. O Método de Validação Cruzada na Estimação de Processos de Poisson não Homogêneos, *Unicamp*.
- [19] *Ramlau-Hansen, H. (1983)*. Smoothing counting process intensities by means of Kernel function. *Ann. Statist.* 11, 453-466.
- [20] *Rathbun, S. L. (1996)*. Estimation of Poisson Intensity Using Partially Observed Concomitant Variables, *Biometrics*, 52, 226-242.
- [21] *Resnick, S. I. (1992)*. *Adventures in Stochastic Processes*, *Birkhäuser, Boston*.
- [22] *Schoenberg, I. J. (1964)*. Spline functions and the problem of graduation, *Proc. Nat. Acad. Sci., USA* 52, 947-950.
- [23] *Thompson, W. A., Jr (1988)*. *Point Process Models with Applications to Safety and Reliability*, *Chapman and Hall, New York*.