

ESTIMAÇÃO DA FUNÇÃO INTENSIDADE DO PROCESSO DE POISSON NÃO
HOMOGÊNEO PELO MÉTODO DE GRENANDER

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. Antonio Eduardo Gomes e aprovada pela Comissão Julgadora.

Campinas, 13 de dezembro de 1991.



Prof. Dr. Mauro Sérgio de Freitas Marques

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

G586e

15425/BC

UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Ciência da Computação - IMECC

**ESTIMAÇÃO DA FUNÇÃO INTENSIDADE DE PROCESSOS DE POISSON NÃO
HOMOGÊNEOS PELO MÉTODO DE GRENANDER**

Antonio Eduardo Gomes

Prof. Dr. Mauro S. de F. Marques

Orientador

Campinas - São Paulo

1991



Aos meus pais.

Agradecimentos

Ao prof. Dr. Mauro Sérgio de Freitas Marques, pela orientação e aprendizado durante a elaboração deste trabalho.

Ao prof. Dr. Dalton Francisco de Andrade, pelas correções e sugestões, e também pelo auxílio na parte computacional.

Ao prof. Dr. José Antonio Cordeiro, pelas correções e sugestões.

À Cibeles, pelo apoio nas horas difíceis.

E a todos que direta ou indiretamente contribuíram para a realização deste trabalho.

Índice

Capítulo 1 - Introdução.....	1
Capítulo 2 - O Método de "Sieves".....	8
2.1 - Experimento Estatístico Abstrato.....	8
2.2 - Descrição do Método de "Sieves".....	13
Capítulo 3 - Aplicação do Método de "Sieves" ao Problema de Estimação da Função Intensidade de um Processo de Poisson Não Homogêneo.....	27
3.1 - Processos Pontuais.....	27
3.2 - Processos de Poisson Não Homogêneos.....	36
3.2.1 - Descrição do Experimento Estatístico.....	44
3.2.2 - Decomposição de Doob-Meyer para um Processo	
de Poisson Não Homogêneo.....	47
3.2.3 - Derivada de Radon-Nikodym.....	49
3.2.4 - Métrica Adotada para o Espaço de Parâmetros.....	52
3.2.5 - Escolha do "Sieve".....	54
3.2.5.1 - "Sieve" de Histograma.....	55
3.2.5.2 - "Sieve" de Fourier.....	59
Capítulo 4 - Aspectos Práticos na Aplicação do Método de "Sieves".....	68
4.1 - Algoritmo para Simulação de Processos de	

Poisson Não Homogêneos.....	68
4.2 - Procedimentos Práticos para Obtenção das estimativas das funções intensidade λ	71
4.3 - Análise dos Resultados.....	75
4.4 - Sugestões.....	76
 Apêndice I.....	 78
 Apêndice II.....	 80
 Apêndice III.....	 105
 Referências Bibliográficas.....	 112

CAPÍTULO 1

INTRODUÇÃO

Um processo pontual é um modelo matemático de pontos distribuídos aleatoriamente em algum espaço, freqüentemente utilizado para descrever fenômenos físicos ou biológicos caracterizados por eventos que ocorrem de forma acidental no tempo ou no espaço. Os pontos podem representar tempos de ocorrência de eventos (como falha ou troca de equipamentos, chegadas de pacientes a uma unidade de pronto socorro ou de clientes a uma loja) ou localizações de objetos, por exemplo.

Como se vê, os modelos de processos pontuais podem ser utilizados na descrição e estudo de muitos fenômenos em diversas áreas. Para tal, basta que seqüências de tempos (cada tempo correspondendo à ocorrência de algum evento) constituam os dados observados, e é este o caso em muitos problemas de teoria de comunicações, pesquisa operacional, medicina nuclear, neurofisiologia, biologia, geografia, meteorologia, física, etc. Note que "tempo" aqui pode ser substituído por "localização".

Em um modelo de processo pontual, o objeto de estudo é, geralmente, a taxa ou intensidade com que um evento ocorre. A intensidade resume, em um dado instante, o potencial para geração de um evento no futuro próximo, dada a observação do passado (incluindo

o registro completo de todos os tempos de ocorrência anteriores). Sendo assim, um modelo estatístico para um processo pontual deve incluir uma especificação de como a intensidade depende do tempo, já que o desenvolvimento de um tal processo é governado pelo seu "processo de intensidade" $\left\{ \lambda(t) , t \geq 0 \right\}$, onde o valor $\lambda(t)dt$ pode ser interpretado como sendo a probabilidade de que um evento ocorra no intervalo $[t, t+dt]$, condicionada a tudo que ocorreu até o instante anterior a t .

As suposições sobre a intensidade determinam a classificação do problema de sua estimação em dois casos distintos. No caso paramétrico, a intensidade tem sua forma funcional especificada exceto por um número finito de valores $\theta_1, \dots, \theta_m$. Se isto não ocorrer a intensidade é considerada um parâmetro de dimensão infinita, e temos, então, um problema de estimação não paramétrica. Neste caso, geralmente são feitas apenas suposições de ordem qualitativa sobre a função intensidade, como: continuidade, simetria, etc.

O exemplo a seguir ilustra a aplicabilidade dos modelos de processos pontuais.

Exemplo 1.1:

Considere n pacientes que sofrem de determinada doença letal sendo acompanhados em um hospital a partir do instante em que a doença é diagnosticada até a morte do paciente ou até uma data limite de

observação. Seja $N(t)$ o número de mortes ocorridas no intervalo $[0, t]$. Então, a intensidade do processo pode ser expressa por $\lambda(t) = \alpha(t)Y(t)$ onde $\alpha(t)$ é tal que a probabilidade de um determinado paciente vir a falecer no intervalo $[t, t+dt]$, dado que ele esteja vivo até o instante imediatamente anterior a t , é dada por $\alpha(t)dt$, e $Y(t)$ representa o número de pacientes vivos até o instante imediatamente anterior a t . Aqui foi suposto que a função $\alpha(t)$ é idêntica para todos os pacientes (i.e., a população é homogênea). Quando a intensidade tem a forma acima ela dita *multiplicativa* (formalmente $\lambda(t) = \alpha(t)Y(t)$, onde α é uma função determinística desconhecida e Y é um processo estocástico previsível, definido no capítulo 3).

Na verdade, a teoria e os exemplos aqui expostos dizem respeito a processos pontuais univariados. Sua generalização, os processos pontuais multivariados (i. e., processos $N = (N_1, \dots, N_k)$, onde N_i , para $i = 1, 2, \dots, k$, são processos pontuais univariados) englobam muitos modelos de processos estocásticos conhecidos, como processos de nascimento e morte (introduzido por Feller (1939)), cadeias de Markov em tempo contínuo, processos de Poisson e modelos para análise de sobrevivência.

A maneira como um processo pontual é observado pode variar quanto ao intervalo de observação ($[0, \infty)$ ou $[0, T]$, no caso de uma censura simples pelo tempo), número de trajetórias (uma ou mais) e continuidade ou não da observação (feita em todo o intervalo ou apenas

nos instantes $0, t, 2t, \dots$). O estudo de propriedades assintóticas pode, então, ser feito considerando $t \rightarrow \infty$ ou $n \rightarrow \infty$. Neste trabalho, observamos n trajetórias de um processo de Poisson não homogêneo (i.e., quando a intensidade λ é determinística) no intervalo $[0, 1]$ (geradas por simulação) e fazemos n variável.

Em muitos casos práticos, como por exemplo, na análise de tempos de falha censurados que surgem em problemas de engenharia, os métodos paramétricos têm encontrado larga utilização. Entre os autores que empregam métodos paramétricos, os primeiros trabalhos abordavam a estimação dos parâmetros de um processo de nascimento e morte, como Kendall (1949), Moran (1951, 1953), Anscombe (1953), Darwin (1956) e, mais recentemente, Keiding (1975), que consideram processos onde as taxas de nascimento (λ) e morte (μ) são constantes. Wolff (1963) supõe um modelo onde λ e μ dependem da diferença entre o número de nascimentos e de mortes de acordo com uma relação funcional que pode ser descrita em termos de um número finito de parâmetros independentes.

Halperin (1952) e Epstein e Sobel (1953) deram importantes contribuições com seus estudos sobre estimação de máxima verossimilhança para modelos paramétricos de análise de sobrevivência com dados censurados.

Para um processo de Poisson não homogêneo, Kutoyants (1979)

estudou a estimação da função intensidade quando esta depende de um parâmetro real unidimensional, i. e. , $\lambda(t) = \lambda(t, \theta)$, $t \in [0, T]$, $\theta \in \Theta \subset \mathbb{R}$. O autor aponta, ainda, propriedades assintóticas para grandes intervalos de observação e dá exemplos. Antes deste trabalho, MacLean (1974) considerou o caso em que

$$\lambda(t) = \exp \left(\sum_{m=0}^r \alpha_m t^m \right).$$

onde $\alpha_1, \dots, \alpha_r$ são constantes.

Mais recentemente, alguns autores abordaram a estimação de processos pontuais gerais, como Borgan (1984) (que obteve propriedades de um estimador de máxima verossimilhança para tempos de falha sob diferentes padrões de censura dos dados, com resultados aplicáveis à análise de modelos de cadeias de Markov em tempo contínuo), e Svensson(1990).

Os trabalhos mencionados acima tratam da estimação paramétrica da função intensidade de processos pontuais. Há situações, no entanto, onde se torna difícil encontrar uma justificativa para a utilização de um modelo paramétrico de processo pontual (como, por exemplo, aplicações em bioestatística). Neste caso, utiliza-se o enfoque não paramétrico para estimação, abordado primeiramente por Kaplan e Meyer (1958), que propuseram seu conhecido estimador de produto limite para a proporção $P(t)$ de elementos da população cujo tempo de vida excede t .

Nelson (1969,1972) e Altshuler(1970) propuseram o estimador empírico cumulativo da intensidade, mais tarde generalizado por Aalen (1978), e que é conhecido como estimador de Nelson-Aalen. Através da suavização dos incrementos desse estimador, Ramlau-Hansen (1983 a, b) obteve um estimador para a função intensidade de um processo pontual (aspectos práticos da utilização do estimador proposto por Ramlau-Hansen foram estudados por Crestana (1991), para o caso de um processo de Poisson não homogêneo).

Cox (1972) propôs um modelo semi-paramétrico de regressão para dados de sobrevivência censurados, no qual a intensidade do processo é dada pelo produto entre um termo paramétrico e um não paramétrico. Andersen e Gill (1982) estudaram tal modelo, que é muito utilizado em pesquisas na área de bioestatística.

Aalen (1978) também deu importante contribuição na área de estimação não paramétrica de processos pontuais, mostrando como a teoria de processos de contagem multivariados serve como estrutura geral na qual tanto dados sobre tempos de falha censurados quanto observações censuradas de cadeias de Markov não homogêneas podem ser analisados. Neste trabalho, o autor propôs, ainda, o modelo de intensidade multiplicativa exposto no exemplo 1.1, ilustrando sua utilização para dados sobre sobrevivência, cadeias de Markov e processos de nascimento e morte.

Entre os trabalhos envolvendo modelos de cadeias de Markov, podemos citar, ainda, Aalen e Johansen (1978), Fleming(1978a,b) e Aalen et al. (1980).

Neste trabalho, estudaremos os aspectos práticos de um método para obtenção de estimadores de máxima verossimilhança da função intensidade de um modelo particular de processo pontual: o processo de Poisson não homogêneo. Tal método, conhecido como método de "sieves", foi proposto por Grenander (1981). Na verdade, trata-se de um método de estimação bastante genérico para a solução de problemas não paramétricos, e sua utilização neste trabalho é apenas uma de suas muitas aplicações possíveis (Geman e Hwang (1982) dão exemplos variados de sua aplicação).

O capítulo 2 contém uma descrição pormenorizada do método de "sieves" e alguns resultados sobre existência e consistência de estimadores obtidos através de sua aplicação. O capítulo seguinte traz resultados sobre a teoria de processos pontuais e particulariza tais resultados para um processo de Poisson não homogêneo, além de descrever este tipo de processo. No capítulo 4 são abordados os aspectos práticos da aplicação do método, através da descrição das estimativas para algumas funções intensidades utilizadas para geração de trajetórias por simulação.

Capítulo 2

O Método de "Sieves"

Neste capítulo, descreveremos o que vem a ser um experimento estatístico abstrato. Posteriormente, o método de "sieves", proposto por Grenander (1981), será descrito e alguns resultados sobre existência e consistência de estimadores de máxima verossimilhança obtidos por este método serão apresentados.

2.1 - Experimento Estatístico Abstrato

Um *experimento estatístico* consiste de um espaço de observações Ω' , uma σ -álgebra $\mathcal{F}'$ de subconjuntos de Ω' , e uma família de medidas de probabilidade $\left\{ P_{\theta} , \theta \in \Theta \right\}$, definidas no espaço mensurável $(\Omega', \mathcal{F}')$. Θ é chamado de *espaço paramétrico*.

As observações constituem a base de um experimento estatístico, e estas podem ser valores numéricos, vetores em $\mathbb{R}^n$, trajetórias de processos estocásticos, etc, como veremos em alguns exemplos.

Exemplo 2.1:

Quando nossa observação X é o valor de uma variável aleatória, temos $\Omega' = \mathbb{R}$, $\mathcal{F}' = \mathcal{B}(\mathbb{R})$ (σ -álgebra de Borel de $\mathbb{R}$), e $\left\{ P_{\theta} , \theta \in \Theta \right\}$ como sendo a classe das possíveis distribuições de probabilidade da

variável aleatória X .

Exemplo 2.2:

Quando observamos n variáveis aleatórias $X_1, \dots, X_n$, temos $\Omega' = \mathbb{R}^n$, $\mathcal{F}' = \mathcal{B}(\mathbb{R}^n)$ (σ -álgebra de Borel de $\mathbb{R}^n$), e $\left\{ P_\theta, \theta \in \Theta \right\}$ como sendo a classe das possíveis distribuições conjuntas de probabilidade das variáveis aleatórias $X_1, \dots, X_n$.

Exemplo 2.3:

Se nossa observação é uma trajetória de um processo estocástico $\left\{ X(t), t \in \mathbb{T} \right\}$, temos: $\Omega' = \mathbb{R}^{\mathbb{T}}$ (espaço das funções $x: \mathbb{T} \rightarrow \mathbb{R}$), $\mathcal{F}' = \sigma(\mathcal{C})$ (σ -álgebra gerada pela classe $\mathcal{C}$ dos conjuntos cilíndricos

$$C(t_1, \dots, t_n, B) = \left\{ x: (x(t_1), \dots, x(t_n)) \in B \right\}, \quad B \in \mathcal{B}(\mathbb{R}^n), \quad e$$

$\left\{ P_\theta, \theta \in \Theta \right\}$ representa a classe das possíveis medidas de probabilidade associadas ao processo $\left\{ X(t), t \in \mathbb{T} \right\}$.

A diversidade de espaços de observações Ω' obtida nos exemplos acima nos mostra a necessidade de definirmos cada observação X como sendo um "elemento aleatório" definido em um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ e tomando valores em $(\Omega', \mathcal{F}', P^X)$, onde $P^X(A) = P(X \in A)$, para qualquer $A \in \mathcal{F}'$. Portanto, X é uma função $\mathcal{F}/\mathcal{F}'$ -mensurável (i.e., $\left\{ \omega : X(\omega) \in B \right\} \in \mathcal{F}$, para qualquer $B \in \mathcal{F}'$).

A característica básica de um elemento aleatório X é sua distribuição de probabilidade associada P^X , definida acima. Nos problemas práticos, no entanto, tudo o que se sabe sobre tal distribuição é que ela pertence a uma determinada classe de distribuições $\mathcal{P}$. O interesse do estatístico é fazer inferências sobre P^X , baseado na observação X , trabalhando com o espaço de valores $(\Omega, \mathcal{F})$ e com a família de distribuições de probabilidade $\mathcal{P}$. O espaço de probabilidade $(\Omega, \mathcal{F}, P)$ onde a observação X está definida, tem um papel auxiliar e pode ser construído de diferentes maneiras. Por exemplo, fazendo $\Omega = \Omega'$, $\mathcal{F} = \mathcal{F}'$, $P = P^X$ e definindo X de tal forma que $X(\omega) = \omega$. Claramente, X é um elemento aleatório com valores em Ω' e distribuição P^X definida em $(\Omega', \mathcal{F}')$. Quando o experimento é construído dessa forma, diz-se que ele é gerado pela observação X .

A família de distribuições $\mathcal{P}$ pode sempre ser parametrizada por meio da escolha adequada de um parâmetro, e representada na forma

$$\mathcal{P} = \left\{ P_{\theta} , \theta \in \Theta \right\}.$$

Além disso, $\mathcal{P}$ deve ser tal que duas de suas distribuições de probabilidade P_{θ_1} e P_{θ_2} distintas, definidas em $\mathcal{F}'$, correspondam a distintos elementos θ_1 e θ_2 de Θ , respectivamente, e vice-versa. Portanto, qualquer inferência estatística feita sobre as distribuições $\left\{ P_{\theta} , \theta \in \Theta \right\}$ é equivalente àquela feita com relação aos valores dos parâmetros $\theta \in \Theta$.

Quando em um experimento estatístico existe uma medida μ

σ -finita, definida em $(\Omega', \mathcal{F}')$ tal que $P_\theta \ll \mu$, para qualquer $\theta \in \Theta$ (i.e., P é absolutamente contínua com relação à medida μ), dizemos que o experimento é *dominado*. Neste caso, em consequência do teorema de Radon-Nikodym (ver Bartle (1966) ou Royden (1968), p. ex.), existem "densidades" relativas a μ para cada valor de θ , i.e.,

$$p(\omega, \theta) = \frac{dP_\theta}{d\mu}(\omega),$$

e para todo $A \in \mathcal{F}'$,

$$P_\theta(A) = \int_A p(\omega, \theta) \mu(d\omega) = \int_A p(\theta) d\mu.$$

A medida dominante μ pode ser escolhida de diversas maneiras, mas para o propósito da investigação estatística, tal escolha é irrelevante, já que, se μ_1 e μ_2 forem duas medidas dominantes distintas, as densidades $p_1(\omega, \theta)$ e $p_2(\omega, \theta)$ diferirão apenas por um fator multiplicativo independente de θ , pois

$$\frac{dP_\theta}{d(\mu_1 + \mu_2)}(\omega) = p_1(\omega, \theta) \frac{d\mu_1}{d(\mu_1 + \mu_2)}(\omega) = p_2(\omega, \theta) \frac{d\mu_2}{d(\mu_1 + \mu_2)}(\omega).$$

O experimento $\varepsilon_1 = \left\{ \Omega'_1, \mathcal{F}'_1, \left\{ P_{1\theta}, \theta \in \Theta \right\} \right\}$ é dito ser uma projeção do experimento $\varepsilon = \left\{ \Omega', \mathcal{F}', \left\{ P_\theta, \theta \in \Theta \right\} \right\}$ se Ω' e $\mathcal{F}'$ puderem ser escritos na forma $\Omega' = \Omega'_1 \times \Omega'_2$, $\mathcal{F}' = \mathcal{F}'_1 \otimes \mathcal{F}'_2$ (σ -álgebra produto), e, além disso, para todo $A_1 \in \mathcal{F}'_1$, $P_{1\theta}(A_1) = P_\theta(A_1 \times \Omega'_2)$.

Quando trabalharmos com uma família $\{ \varepsilon_i, i \in \mathbb{N} \}$ de experimentos estatísticos (por exemplo, quando temos n observações de um experimento ε_i), devemos assumir que todos os experimentos ε_i são projeções de um experimento comum ε . Por exemplo, sejam $\varepsilon_i, i \in \mathbb{N}$, os experimentos gerados pelas observações X_i , e ε o experimento gerado pelo sistema de observações $\{ X_i, i \in \mathbb{N} \}$. Neste caso, todos os experimentos ε_i são projeções de ε .

O experimento $\varepsilon = \left\{ \Omega', \mathcal{F}', \left\{ P_\theta, \theta \in \Theta \right\} \right\}$ é chamado "produto" dos experimentos $\varepsilon_i = \left\{ \Omega'_i, \mathcal{F}'_i, \left\{ P_{i\theta}, \theta \in \Theta \right\} \right\}, i=1, \dots, n$, se $\Omega = \Omega'_1 \times \dots \times \Omega'_n$, $\mathcal{F}' = \mathcal{F}'_1 \otimes \dots \otimes \mathcal{F}'_n$ e $P_\theta = P_{1\theta} \times \dots \times P_{n\theta}$. Claramente, todos os experimentos ε_i são projeções de ε . Se o experimento ε_i é gerado por uma observação X_i , então o experimento ε é gerado por n observações independentes $X_1, \dots, X_n$.

Quando consideramos um experimento estatístico onde Θ tem dimensão infinita, (i.e., quando $\theta \in \Theta$ é um parâmetro de dimensão infinita) a técnica convencional utilizada para obtenção de estimadores de máxima verossimilhança, geralmente, apresenta problemas, como ocorre no exemplo a seguir.

Exemplo 2.4:

Seja $X_1, \dots, X_n$ uma amostra i.i.d. de uma distribuição absolutamente contínua, cuja densidade f_0 é totalmente desconhecida,

ou seja,

$$\Theta = \left\{ f : \mathbb{R} \rightarrow \mathbb{R}^+ : \int_{-\infty}^{+\infty} f(x)dx = 1 \right\}.$$

O estimador de máxima verossimilhança f^* de f_0 maximiza

$$L(\underline{X}, f) = \prod_{i=1}^n f(X_i),$$

onde $\underline{X} = (X_1, \dots, X_n)^t$.

Como nenhuma suposição foi feita sobre a forma funcional da densidade f_0 , podemos fazer com que ela assuma valores cada vez maiores em vizinhanças cada vez menores dos pontos $X_1, \dots, X_n$ de tal forma que $\int_{-\infty}^{+\infty} f(x)dx = 1$. Obviamente, $L(\underline{X}, f)$ é não limitada quando utilizamos tal procedimento, ou seja, não é possível encontrar uma densidade que maximize $L(\underline{X}, f)$.

Este exemplo mostra a necessidade de métodos alternativos de estimação quando Θ possui dimensão infinita. Para o problema específico de estimação não-paramétrica de densidades, ver Tapia e Thompson (1978). Para problemas mais gerais, o método descrito neste trabalho (método de "sieves", proposto por Grenander (1981)) é uma alternativa e foi proposto por Grenander (1981).

2.2 - Descrição do Método de "Sieves"

O método de "sieves" é mais uma técnica não paramétrica, proposta para solucionar problemas como o descrito no exemplo apresentado, e

sua idéia básica consiste em realizar o processo de otimização (em nosso caso, maximizar a função de verossimilhança) considerando como espaço paramétrico um subconjunto do espaço de parâmetros Θ verdadeiro, e repetindo o processo para subconjuntos cada vez menos restritos de Θ , conforme o tamanho da amostra cresça.

O que obtemos, então, é uma seqüência de estimadores. A definição da seqüência de subconjuntos (que vem a ser o "sieve") não pode, no entanto, ser feita livremente. Ela deve ser tal que, para cada subconjunto S_m do espaço paramétrico Θ , exista um estimador de máxima verossimilhança. Além disso, qualquer que seja o verdadeiro parâmetro $\theta_0 \in \Theta$, ele deve poder ser tão aproximado quanto se queira pela seqüência de estimadores de máxima verossimilhança obtidos em cada subconjunto S_m do espaço de parâmetros Θ . É natural que se exija portanto, que $\bigcup_{m=1}^{\infty} S_m$ seja denso em Θ .

Exemplos de utilização deste método podem ser encontrados em Geman e Hwang (1982), Nguyen e Pham (1982) e McKeague (1986).

Geman e Hwang (1982) estabeleceram alguns resultados sobre consistência dos estimadores obtidos por este método, seguindo uma idéia proposta por Wald (1949), que demonstrou a consistência dos estimadores de máxima verossimilhança baseado na lei forte dos grandes números e em uma desigualdade envolvendo as entropias das distribuições de probabilidade (informação de Kullback-Leibler),

observando não ser necessário que o espaço paramétrico tenha dimensão finita.

Sendo $f(x, \theta)$ a densidade da variável aleatória X e θ um elemento do espaço paramétrico,

$$H(\theta_0, \theta) = E_{\theta_0} \left[\ln f(X, \theta) \right]$$

é a *entropia formal* da distribuição de probabilidade da variável aleatória X , onde θ_0 designa o verdadeiro valor do parâmetro. Para uma amostra $X_1, \dots, X_n$, i.i.d., seja

$$Y_1 = \ln[f(X_1, \theta)].$$

Temos, então, que

$$\bar{Y} = \frac{1}{n} \ln[L(\underline{X}, \theta)] = \frac{1}{n} \sum_{i=1}^n \ln[f(X_i, \theta)],$$

onde L é a função de verossimilhança. Como

$$E[Y_1] = H(\theta_0, \theta)$$

e

$$H(\theta_0, \theta) < H(\theta_0, \theta_0)$$

(ver Rényi (1966), p. ex.), se maximizarmos $L(\underline{X}, \theta)$ e utilizarmos a lei forte dos grandes números, espera-se que o valor de θ obtido aproxime-se daquele que maximiza $H(\theta_0, \theta)$, ou seja, do verdadeiro parâmetro θ_0 .

A seguir é descrita a notação, bem como algumas suposições necessárias para exposição dos resultados obtidos por Geman e Hwang (1982).

Dado um experimento estatístico abstrato $\left\{ \Omega, \mathcal{F}, \left\{ P_{\theta} : \theta \in \Theta \right\} \right\}$ dominado por uma medida σ -finita μ , $p(\omega, \theta)$ designará a derivada de Radon-Nikodym $\frac{dP_{\theta}}{d\mu}(\omega)$ e θ_0 o verdadeiro parâmetro.

Para $\theta \in S_m$,

$$B_m(\theta, \varepsilon) = \left\{ \beta : \beta \in S_m, d(\theta, \beta) < \varepsilon \right\}$$

onde $d(.,.)$ é uma métrica de Θ .

$$E_{\theta}(g) = \int g dP_{\theta} = \int g \frac{dP_{\theta}}{d\mu} d\mu.$$

Para qualquer função real estendida g em Θ e qualquer $B \subset \Theta$,

$$g[B] = \sup_{\beta \in B} g(\beta).$$

$$M_m^n = \left\{ \theta \in S_m : L_n(\underline{\omega}, \theta) = L_n\left[\underline{\omega}, S_m\right] \right\}$$

é o conjunto de estimadores de máxima verossimilhança, onde:

$$L_n(\underline{\omega}, \theta) = \prod_{i=1}^n \frac{dP_{\theta}}{d\mu}(\omega_i) \quad e$$

$\underline{\omega} = (\omega_1, \dots, \omega_n)$ representa a amostra.

$$A_m = \left\{ \theta : \theta \in S_m, H(\theta_0, \theta) = H\left[\theta_0, S_m\right] \right\}$$

representa o conjunto de entropia máxima em S_m .

Para $C_m \subseteq \Theta$, $C_m \rightarrow \Theta$, quando $m \rightarrow \infty$, significa que $\sup_{\alpha \in C_m} d(\theta, \alpha) \rightarrow 0$,

quando $m \rightarrow \infty$.

A medida produto $P_{\theta_0} \times P_{\theta_0} \times \dots$ em $(\Omega^*, \mathcal{F}^*)$, será designada por P , onde $\Omega^* = \Omega' \times \Omega' \times \dots$ e $\mathcal{F}^*$ é o completamento de $\mathcal{F}' \otimes \mathcal{F}' \otimes \dots$ (σ -álgebra gerada por retângulos $A_1 \times A_2 \times \dots$ onde $A_1 \in \mathcal{F}'$) com relação à P .

O primeiro resultado obtido por Geman e Hwang (também apresentado em Grenander (1981)) trata da existência de uma sequência m_n para a qual o conjunto de máxima verossimilhança $M_{m_n}^n$ é consistente.

Teorema 2.1:

Suponha que um sieve $\{S_m\}$ seja escolhido de tal forma que:

1) para todo $m \in \mathbb{N}$, todo $\theta \in S_m$ e todo $\varepsilon > 0$,

$$p\left[\omega, B_m(\theta, \varepsilon)\right] \text{ é } \mathcal{F}'\text{-mensurável;}$$

2) para todo $m \in \mathbb{N}$ e todo $\theta \in S_m$,

$$\lim_{\varepsilon \rightarrow 0} p\left[\omega, B_m(\theta, \varepsilon)\right] = p(\omega, \theta)$$

μ -quase certamente, i.e., $p(\omega, \theta)$ é semicontinua superior em θ , para θ

$\in S_m$;

3) para todo $m \in \mathbb{N}$ e todo $\theta \in S_m$, existe $\varepsilon > 0$ tal que

$$E_{\theta_0} \left(\ln p \left[\omega, B_m(\theta, \varepsilon) \right] \right) < \infty;$$

4) S_m é compacto para todo $m \in \mathbb{N}$;

5) $A_m \rightarrow \theta_0$, quando $m \rightarrow \infty$.

Então, para todos $n, m \in \mathbb{N}$, $M_m^n(\omega)$ é não vazio P -quase certamente, e para toda sequência m_n que cresça de forma suficientemente lenta, $M_{m_n}^n \rightarrow \theta_0$ P -quase certamente.

Obs.: Como, em geral, $H(\theta_0, \theta_m) \rightarrow H(\theta_0, \theta_0)$ implica $\theta_m \rightarrow \theta_0$, quando $m \rightarrow \infty$, se o sieve $\{S_m\}$ for escolhido de tal forma que exista $\theta_m \in S_m$ com $H(\theta_0, \theta_m) \rightarrow H(\theta_0, \theta_0)$, então $A_m \rightarrow \theta_0$. A condição 5 garante, então, que o sieve $\{S_m\}$ é denso em Θ .

Demonstração:

As condições 2 e 3 implicam que $H(\theta_0, \theta)$ e $L_n(\omega, \theta)$ são semicontínuas superiores em θ , $\theta \in S_m$. Como S_m é compacto, A_m e M_m^n são compactos e não vazios, já que uma função semicontínua superior definida em um conjunto compacto assume seu máximo.

Primeiramente mostraremos que, quando n_m , definido de tal forma que

$$n_{m_0} = \min \left\{ k : m_k \geq m_0 \right\}$$

crece de forma suficientemente rápida, $M_{m_n}^n$ eventualmente fica contido em

$$B_m(A_m, 1/m) = \left\{ \theta : \theta \in S_m, d(\theta, A_m) < \frac{1}{n} \right\},$$

onde $d(\theta, A_m) = \inf_{\alpha \in A_m} d(\theta, \alpha)$.

A condição 5, então, completa a demonstração.

Tome

$$\theta \in S_m \setminus B_m(A_m, 1/m)$$

e

$$c > 0$$

tal que

$$H(\theta_0, \theta_m) - H[\theta_0, S_m] < -c$$

e fixe $\alpha_m \in A_m$. Para todo ε suficientemente pequeno,

$$E_{\theta_0} \left[\ln \left(p \left[\omega, B_m(\theta, \varepsilon) \right] \right) \right] - E_{\theta_0} \left[\ln \left(p \left(\omega, \alpha_m \right) \right) \right] < -c.$$

Usando este resultado, temos que

$$\begin{aligned} & \frac{1}{n} \left\{ \ln \left(L_n \left[\omega, B_m(\theta, \varepsilon) \right] \right) - \ln \left(L_n(\omega, \alpha_m) \right) \right\} \\ &= \frac{1}{n} \left\{ \ln \left(\sup_{\gamma \in B_m(\theta, \varepsilon)} \left[\prod_{i=1}^n p(\omega_i, \gamma) \right] \right) - \ln \left(\prod_{i=1}^n p(\omega_i, \alpha_m) \right) \right\} \end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{n} \left\{ \ln \left(\prod_{i=1}^n p[\omega_i, B_m(\theta, \varepsilon)] \right) - \ln \left(\prod_{i=1}^n p(\omega_i, \alpha_m) \right) \right\} \\
&= \frac{1}{n} \sum_{i=1}^n \left(\ln \left(p[\omega_i, B_m(\theta, \varepsilon)] \right) - \ln \left(p(\omega_i, \alpha_m) \right) \right) \\
&E_{\theta_0} \left[\ln \left(p[\omega, B_m(\theta, \varepsilon)] \right) \right] - E_{\theta_0} \left[\ln \left(p(\omega, \alpha_m) \right) \right] < -c
\end{aligned}$$

P-q.c., quando $n \rightarrow \infty$.

Portanto,

$$\frac{L_n \left[\omega, B_m(\theta, \varepsilon) \right]}{L_n \left(\omega, \alpha_m \right)} < e^{-nc}$$

P-quase certamente, quando n é suficientemente grande (n pode depender de ω).

Para todo $\theta \in S_m \setminus B_m(A_m, 1/m)$, podemos tomar $c_\theta > 0$ e $\varepsilon_\theta > 0$ tais que, com probabilidade 1, a desigualdade acima seja válida para n grande. As bolas $B_m(\theta, \varepsilon_\theta)$ cobrem o conjunto compacto $S_m \setminus B_m(A_m, 1/m)$, de tal forma que podemos encontrar um conjunto finito $\theta_1^m, \dots, \theta_{p_m}^m$ tal que $\{B_m(\theta_k^m, \varepsilon_k^m)\}$, $k=1, \dots, p_m$, cobre $S_m \setminus B_m(A_m, 1/m)$. Então, com

$$c = \min_{k=1, \dots, p_m} \{c_{\theta_k}\},$$

temos, com probabilidade 1,

$$\sup_{k=1, \dots, p_m} \frac{L_n \left[\underline{\omega}, B_m(\theta_k^m, \varepsilon_k^m) \right]}{L_n \left(\underline{\omega}, \alpha_m \right)} < e^{-nc},$$

para todo $n \in \mathbb{N}$ suficientemente grande.

Portanto, para todo $m \in \mathbb{N}$, existe G_m tal que, sempre que $n \geq G_m$,

$$P \left\{ \sup_{k=1, \dots, p_m} \frac{L_n \left[\underline{\omega}, B_m(\theta_k^m, \varepsilon_k^m) \right]}{L_n \left(\underline{\omega}, \alpha_m \right)} \geq 1 \right\} \leq \frac{1}{m^2}.$$

Finalmente, para qualquer seqüência $\{n_m\}$ tal que $n_m \geq G_m$,

para todo $m \in \mathbb{N}$,

$$\begin{aligned} & \left\{ M_m^{n_m} \subseteq B_m(A_m, 1/m) \text{ i.o.} \right\} \\ & \subseteq \left\{ \sup_{\theta \in S_m \setminus B_m(A_m, 1/m)} \frac{L_n(\underline{\omega}, \theta)}{L_n(\underline{\omega}, \alpha_m)} \geq 1 \text{ i.o.} \right\} \\ & \subseteq \left\{ \sup_{k=1, \dots, p_m} \frac{L_n \left[\underline{\omega}, B_m(\theta_k^m, \varepsilon_k^m) \right]}{L_n \left(\underline{\omega}, \alpha_m \right)} \geq 1 \text{ i.o.} \right\} \end{aligned}$$

e este último conjunto possui probabilidade zero, pelo lema de Borel-Cantelli. ■

Este teorema fornece condições para que um sieve produza estimadores consistentes. No entanto, nenhuma informação sobre o quão rápido n_m (ou o quão lento m_n) deve tender a infinito pode ser extraída de seu enunciado. O teorema a seguir fornece um meio para a identificação de uma tal seqüência.

Teorema 2.2:

Suponha um sieve $\{S_m\}$ tal que as condições sejam satisfeitas.

- 1) para todos $m, n \in \mathbb{N}$, M_m^n é não vazio P -quase certamente;
- 2) se, para alguma seqüência $\theta_m \in S_m$,

$$H(\theta_0, \theta_m) \rightarrow H(\theta_0, \theta_0) ,$$

então

$$\theta_m \rightarrow \theta_0;$$

- 3) existe uma seqüência $\theta_m \in S_m$ tal que

$$H(\theta_0, \theta_m) \rightarrow H(\theta_0, \theta_0).$$

Para cada $\delta > 0$ e cada m , seja

$$D_m = \left\{ \theta \in S_m : H(\theta_0, \theta) \leq H(\theta_0, \theta_m) - \delta \right\},$$

onde $\{\theta_m\}$ é a seqüência da condição 3. Dados l conjuntos $O_1, \dots, O_l$ em

S_m tais que $p[., O_k]$ seja $\mathcal{F}'$ -mensurável para cada k , seja

$$\rho_m = \sup_k \inf_{t \geq 0} E_{\theta_0} \left[\exp \left\{ t \ln \left(\frac{p[\omega, O_k]}{p(\omega, \alpha_m)} \right) \right\} \right].$$

Considere uma seqüência $m_n \rightarrow \infty$, e suponha que para cada $\delta > 0$ seja possível encontrar $O_1^m, \dots, O_{\ell_m}^m$ em S_m , $m \in \mathbb{N}$, tais que:

$$4) D_m \subseteq \bigcup_{k=1}^{\ell_m} O_k^m;$$

$$5) p[., O_k^m] \text{ é } \mathcal{F}'\text{-mensurável};$$

$$6) \sum_{n=1}^{\infty} \ell_{m_n} (\rho_{m_n})^n < \infty.$$

Então,

$$M_{m_n}^n \rightarrow \theta_0 \text{ P-quase certamente.}$$

Obs.1: É de se esperar que $\rho_m < 1$, já que, se O_k é pequeno e existe algum $\theta \in O_k \cap D_m$, então a função

$$\phi(t) = E_{\theta_0} \left[\exp \left\{ t \ln \left(\frac{p[\omega, O_k]}{p(\omega, \alpha_m)} \right) \right\} \right]$$

satisfaz $\phi(0) = 1$, e

$$\phi'(0) = E_{\theta_0} \left[\ln(p[\omega, O_k]) \right] - H(\theta_0, \theta_m) \approx$$

$$\approx H(\theta_0, \theta) - H(\theta_0, \theta_m) \leq -\delta.$$

Note, ainda, que conjuntos $O_1, \dots, O_{\ell_m}$ implicam ρ_m menor, mas ℓ_m maior.

Obs.2: A métrica d a ser escolhida para o espaço paramétrico Θ , e em termos da qual é expressa a consistência dos estimadores, geralmente é sugerida pela condição 2. Ela deve ser tal que

$$H(\theta_0, \theta_m) \rightarrow H(\theta_0, \theta_0) \Rightarrow d(\theta_m, \theta_0) \rightarrow 0.$$

Demonstração:

Fixemos $\delta > 0$. O que queremos é mostrar que

$$P\left(D_{m_n} \cap M_{m_n}^n \neq \emptyset \text{ i.o.}\right) = 0, \quad (1)$$

pois, isto implica, com probabilidade 1, que

$$\inf_{\theta \in M_{m_n}^n} H(\theta_0, \theta) \geq H(\theta_0, \theta_{m_n}) - \delta$$

para todo n suficientemente grande. Como δ é arbitrário e

$$H(\theta_0, \theta_m) \rightarrow H(\theta_0, \theta_0)$$

pela condição 3,

$$\lim_{n \rightarrow \infty} \inf_{\theta \in M_{m_n}^n} H(\theta_0, \theta) \geq H(\theta_0, \theta_0)$$

P-quase certamente.

Como $H(\theta_0, \theta) \leq H(\theta_0, \theta_0)$,

$$\lim_{n \rightarrow \infty} \sup_{\theta \in M_{m_n}^n} |H(\theta_0, \theta) - H(\theta_0, \theta_0)| = 0 \text{ P-quase certamente.} \quad (2)$$

Fixemos $\varepsilon > 0$, e para cada n , escolhamos $\alpha_n \in M_{m_n}^n$ tal que

$$\frac{d(\theta_0, \alpha_n)}{1 + d(\theta_0, \alpha_n)} > \sup_{\theta \in M_{m_n}^n} \frac{d(\theta_0, \theta)}{1 + d(\theta_0, \theta)} - \varepsilon.$$

A condição 2, combinada com o resultado (2), implica que

$$d(\theta_0, \alpha_n) \rightarrow 0 \quad P\text{-quase certamente.}$$

Portanto,

$$\limsup \sup_{\theta \in M_{m_n}^n} \frac{d(\theta_0, \theta)}{1 + d(\theta_0, \theta)} \leq \varepsilon \quad P\text{-quase certamente.}$$

Como ε é arbitrário,

$$M_{m_n}^n \rightarrow \theta_0 \quad P\text{-quase certamente.}$$

Consequentemente, é suficiente provar o resultado (1).

Fixemos n e m . Então,

$$\begin{aligned} & \left\{ D_m \cap M_m^n \neq \emptyset \right\} \\ & \subseteq \left\{ \sup_{\theta \in D_m} L_n(\omega, \theta) \geq L_n(\omega, \theta_m) \right\} \\ & \subseteq \bigcup_{k=1}^{\ell_m} \left\{ \sup_{\theta \in O_k^n} \left[\prod_{i=1}^n P(\omega_i, \theta) \right] \geq \prod_{i=1}^n P(\omega_i, \theta_m) \right\} \end{aligned}$$

$$\leq \bigcup_{k=1}^{\ell_m} \left\{ \prod_{l=1}^n P(\omega_l, \theta_k^m) \geq \prod_{l=1}^n P(\omega_l, \theta_m) \right\}.$$

Limitamos, então, a probabilidade deste último conjunto, designada por π .

$$\begin{aligned} \pi &\leq \sum_{k=1}^{\ell_m} P \left\{ \prod_{l=1}^n P[\omega_l, \theta_k^m] \geq \prod_{l=1}^n P(\omega_l, \theta_m) \right\} \\ &= \sum_{k=1}^{\ell_m} P \left[\exp \left[\sum_{l=1}^n t_k \ln \left\{ \frac{P[\omega_l, \theta_k^m]}{P(\omega_l, \theta_m)} \right\} \right] \geq 1 \right] \\ &\leq \sum_{k=1}^{\ell_m} \left\{ E_{\theta_0} \left[\exp \left\{ t_k \ln \left(\frac{P[\omega_l, \theta_k^m]}{P(\omega_l, \theta_m)} \right) \right\} \right] \right\}^n. \end{aligned}$$

para quaisquer $t_1, \dots, t_k$ não negativos. Como $t_1, \dots, t_k$ são arbitrários, $\pi \leq \ell_m (\rho_m)^n$, e (1) segue do lema de Borel-Cantelli. ■

Obviamente o teorema não fornece nenhuma regra simples para o cálculo de m_n . A etapa mais difícil na aplicação do método está no estabelecimento de condições específicas para cada exemplo.

Capítulo 3

Aplicação do Método de "Sieves" ao Problema de Estimação da Função Intensidade de Processos de Poisson Não Homogêneos

Como os aspectos práticos da aplicação do método de "sieves" têm sido pouco estudados, nos restringiremos apenas ao mais simples dos processos pontuais onde o método faz sentido, o processo de Poisson. Neste capítulo, algumas definições da teoria de processos pontuais serão expostos e particularizados para o caso de um processo de Poisson não homogêneo. Feito isto, o experimento estatístico com o qual estamos trabalhando será especificado. Além disso, descreveremos os "sieves" utilizados para estimação da função intensidade.

3.1 - Processos Pontuais

Antes de darmos a definição de processo pontual, os conceitos nela envolvidos serão expostos. O primeiro deles diz respeito a um tipo especial de processo estocástico, chamado *martingale*. A teoria sobre este tipo de processo é a estrutura matemática sobre a qual a noção de intensidade é formalizada, e sua utilização em processos pontuais tornou-se usual a partir do trabalho de Brémaud (1972).

Definição 3.1:

Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade completo (i.e., $\mathcal{F}$ contém

todos os subconjuntos de conjuntos de Ω de probabilidade zero),

$\left\{ \mathcal{F}_t, t \geq 0 \right\}$ uma família crescente de sub- σ -álgebras de $\mathcal{F}$ (i.e., $\mathcal{F}_t$ é σ -álgebra e $\mathcal{F}_t \subseteq \mathcal{F}$, para todo $t \geq 0$, onde $\mathcal{F}_s \subseteq \mathcal{F}_t$, para $s < t$). Tal família é denominada *filtragem*. Então, $M = \left\{ M(t), t \geq 0 \right\}$ é um *martingale* se as duas condições abaixo forem satisfeitas:

1) $M(t)$ é $\mathcal{F}_t$ -mensurável (i.e., $M(t)$ é uma variável aleatória com relação à σ -álgebra $\mathcal{F}_t$) e integrável ($E(|M(t)|) < \infty$), para todo $t \geq 0$;

2) Para $0 \leq s < t$, $E \left(M(t) \mid \mathcal{F}_s \right) = M(s)$, P-quase certamente.

No item (2) acima, se o sinal de igualdade for substituído pelo sinal $\leq$, M será chamada de *supermartingale*; e se for substituído pelo sinal $\geq$, M será chamada de *submartingale*.

Quando consideramos $\mathcal{F}_t = \mathcal{F}_t^X = \sigma \left\{ X(s), s \in [0, t] \right\}$, a filtragem é também chamada de *história interna do processo* $X = \left\{ X(t), t \geq 0 \right\}$.

Outro conceito envolvido na definição de um processo pontual é o de *tempo de Markov*.

Definição 3.2:

Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade completo e $\left\{ \mathcal{F}_t, t \geq 0 \right\}$ uma filtragem. A variável aleatória $\tau = \tau(\omega)$ que assume valores em

$[0, \infty]$ é chamada de *tempo de Markov* relativo à filtragem $\left\{ \mathcal{F}_t, t \geq 0 \right\}$

se, para todo $t \geq 0$, tivermos

$$\left\{ \omega : \tau(\omega) \leq t \right\} \in \mathcal{F}_t.$$

Com estes conceitos em mente, podemos, agora, definir o que vem a ser um processo pontual.

Definição 3.3:

Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade completo, $\left\{ \mathcal{F}_t, t \geq 0 \right\}$ uma filtragem contínua à direita (i.e., $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$) tal que $\mathcal{F}_0$ contém todos os conjuntos de probabilidade zero de Ω . Um *processo pontual* é uma sequência $T = \left\{ \tau_n, n \geq 1 \right\}$ de tempos de Markov com relação à filtragem $\left\{ \mathcal{F}_t, t \geq 0 \right\}$, que satisfaz:

- 1) $\tau_1 > 0$ (P-quase certamente) ;
- 2) $\tau_n < \tau_{n+1}$ (P-quase certamente) , se $\tau_n < \infty$;
- 3) $\tau_n = \tau_{n+1}$ (P-quase certamente) , se $\tau_n = \infty$.

Um processo pontual $T = \left\{ \tau_n, n \geq 1 \right\}$ é completamente caracterizado pelo *processo de contagem* $N = \left\{ N(t), t \geq 0 \right\}$, onde:

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

ou seja,

$$N(t) = \begin{cases} n, & \text{se } t \in [\tau_n, \tau_{n+1}) \\ +\infty, & \text{se } t \geq \tau_\infty \end{cases}$$

para $n \geq 0$, onde $\tau_\infty = \lim_{n \rightarrow \infty} \tau_n$.

Aqui, consideraremos apenas os casos em que $\tau_\infty = \infty$.

Como fica claro na definição acima, $N = \{ N(t), t \geq 0 \}$ é um processo estocástico que pode ser definido como sendo a contagem do número de ocorrências do evento sob estudo no intervalo $[0, t]$.

Na verdade, muitas vezes $\{ N(t), t \geq 0 \}$ também costuma ser chamado de processo pontual, já que o estudo da sequência $T = \{ \tau_n, n \geq 1 \}$ e do processo $N = \{ N(t), t \geq 0 \}$ são equivalentes 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.

Convém salientar que as trajetórias destes processos são P-quase certamente contínuas à direita, possuem limites à esquerda e são funções escada com saltos de magnitude 1.

Note, ainda, que N é um submartingale, pois $N(t)$, $t \geq 0$, é não decrescente P -quase certamente e $N(s)$ é $\mathcal{F}_s$ -mensurável, e portanto, temos que $E[N(t) | \mathcal{F}_s] \geq E[N(s) | \mathcal{F}_s] = N(s)$.

A seguir, são dados alguns exemplos de processos pontuais.

Exemplo 3.1:

$N(t) = 1_{[\tau, \infty)}(t)$, $t \geq 0$, onde τ é tempo de Markov com $P(\tau > 0) = 1$. Neste caso, $\tau_1 = \tau$, e $\tau_n = \infty$, para $n \geq 2$. Aqui, podemos considerar a filtragem na qual $\mathcal{F}_t = \sigma\{N(s), s \leq t\}$.

Exemplo 3.2:

Processo de Poisson $\pi = \{\pi(t), t \geq 0\}$, com parâmetro λ , isto é, um processo com incrementos estacionários independentes (também dito homogêneo, pois a distribuição de probabilidade de $\pi(t) - \pi(s)$ depende apenas de $t - s$) tal que $\pi(0) = 0$ e

$$P(\pi(t) - \pi(s) = k) = \frac{e^{-\lambda(t-s)} [\lambda(t-s)]^k}{k!}, \quad s \leq t, \quad k = 0, 1, 2, \dots$$

Este é um processo pontual com respeito à filtragem na qual

$$\mathcal{F}_t = \mathcal{F}_t^\pi = \sigma\{\pi(s), s \leq t\}, \quad t \geq 0.$$

Exemplo 3.3:

Seja $N = \left\{ N(t) , t \geq 0 \right\}$ um processo pontual, e σ um tempo de Markov com respeito a $F = \left\{ \mathcal{F}_t , t \geq 0 \right\}$. O processo $\left\{ N(t \wedge \sigma), t \geq 0 \right\}$ é um processo pontual, onde:

$$N(t \wedge \sigma) = \sum_{n \geq 1} 1_{[\tau_n^\sigma, \infty)}(t)$$

e

$$\tau_n^\sigma = \begin{cases} \tau_n , & \text{se } \tau_n \leq \sigma , \\ \infty , & \text{se } \tau_n > \sigma . \end{cases}$$

Definido o que vem a ser um processo pontual, podemos, agora, expor um resultado bastante importante sobre tais processos, o teorema da decomposição de Doob-Meyer. Antes de enunciá-lo, no entanto, os conceitos nele envolvidos serão expostos.

Definição 3.4:

Um submartingale $M = \left\{ M(t) , t \geq 0 \right\}$ com respeito à filtragem $\left\{ \mathcal{F}_t , t \geq 0 \right\}$ é contínuo à direita se suas trajetórias $M(t, \omega)$, $\omega \in \Omega$, são contínuas à direita (P-quase certamente), e se a família $\left\{ \mathcal{F}_t , t \geq 0 \right\}$, é contínua à direita.

Definição 3.5:

Um conjunto de variáveis aleatórias $\{ \xi_\alpha : \alpha \in \mathcal{U} \}$ é dito ser *uniformemente integrável* se

$$\lim_{x \rightarrow \infty} \sup_{\alpha \in \mathcal{U}} \int \{ |\xi_\alpha| > x \} |\xi_\alpha| dP = 0 ,$$

ou equivalentemente,

$$\sup_{\alpha \in \mathcal{U}} E(|\xi_\alpha|) < \infty$$

e

$$\lim_{P(A) \rightarrow 0} \sup_{\alpha \in \mathcal{U}} \int_A |\xi_\alpha| dP = 0 , \text{ para } \forall A \in \mathcal{F} .$$

Definição 3.6:

Um (sub, super) martingale $M = \{ M(t) , t \geq 0 \}$ com respeito à filtragem $\{ \mathcal{F}_t , t \geq 0 \}$, com trajetórias $M(t, \omega)$, $\omega \in \Omega$, contínuas à direita, *pertence à classe D* se a família de variáveis aleatórias $\{ M(\tau) , \tau \in \mathcal{T} \}$, onde $\mathcal{T}$ é o conjunto de tempos de Markov τ com relação à mesma filtragem para os quais $P(\tau < \infty) = 1$, é uniformemente integrável.

Definição 3.7:

O processo estocástico $A = \{ A(t) , t \geq 0 \}$ é dito ser *crescente* se as variáveis $A(t)$ são $\mathcal{F}_t$ -mensuráveis, $A(0) = 0$ e $A(s) \leq A(t)$

(P-quase certamente), para $s \leq t$. Tal processo é dito ser *previsível* se a função $A : [0, \infty) \times \Omega \rightarrow \mathbb{R}$ é $\mathcal{W}(\mathcal{F}_t)$ -mensurável, onde $\mathcal{W}(\mathcal{F}_t)$ é a σ -álgebra de subconjuntos de $(0, \infty) \times \Omega$ gerada pelos retângulos da forma $(s, t] \times G$; $0 \leq s \leq t$, $G \in \mathcal{F}_s$, e $\{0\} \times G$, $G \in \mathcal{F}_0$. Tal processo é dito ser *integrável* se $E[A(t)] < \infty$, para todo $t > 0$.

Para efeitos práticos, um processo previsível pode ser considerado como sendo adaptado a $\left\{ \mathcal{F}_t, t \geq 0 \right\}$ (i.e., $A(t)$ é $\mathcal{F}_t$ -mensurável, para todo $t \geq 0$) e contínuo à esquerda.

Teorema 3.1 (decomposição de Doob-Meyer):

Seja $X = \left\{ X(t), t \geq 0 \right\}$ um submartingale com respeito a $\left\{ \mathcal{F}_t, t \geq 0 \right\}$ contínuo à direita e pertencente à classe D . Então existe um processo crescente previsível integrável $A = \left\{ A(t), t \geq 0 \right\}$ adaptado a $\left\{ \mathcal{F}_t, t \geq 0 \right\}$ tal que $X(t) - A(t) = M(t)$, onde $\left\{ M(t), t \geq 0 \right\}$ é um martingale contínuo à direita uniformemente integrável. Esta decomposição é única (a menos de processos indistinguíveis, i.e., processos tais que $P(X(t) = Y(t)) = 1$, para todo $t \geq 0$).

Pode ser mostrado que submartingales positivos (como por exemplo, processos de contagem) são de classe D .

Exemplo 3.4:

Seja $N = \{ N(t), t \geq 0 \}$ um processo determinístico com $N(t) = 1_{[1, \infty)}(t)$ e $\mathcal{F}_t = \left\{ \left\{ \Omega, \emptyset \right\}, t \geq 0 \right\}$. Na decomposição do teorema acima, temos $M(t) = 0$, $A(t) = N(t)$.

Exemplo 3.5:

Seja $\Pi = \{ \pi(t), t \geq 0 \}$ um processo de Poisson com parâmetro $\lambda > 0$. O processo $\{ \pi(t) - \lambda t, t \geq 0 \}$ é uma martingale com respeito a $\{ \mathcal{F}_t^\pi, t \geq 0 \}$. Portanto, $M(t) = \pi(t) - \lambda t$ e $A(t) = \lambda t$ na decomposição do teorema. De fato,

$$\begin{aligned} E \left[M(t) \mid \mathcal{F}_s^\pi \right] &= E \left[\pi(t) - \lambda t \mid \mathcal{F}_s^\pi \right] = \\ &= E \left[\pi(t) - \pi(s) + \pi(s) - \lambda t + \lambda s - \lambda s \mid \mathcal{F}_s^\pi \right] = \\ &= \lambda(t-s) + \pi(s) - \lambda(t-s) - \lambda s = \pi(s) - \lambda s = M(s), \end{aligned}$$

pois π possui incrementos independentes, $\pi(s)$ é $\mathcal{F}_s^\pi$ -mensurável e λ é constante. $A(t) = \lambda t$ é integrável e previsível (λt é determinística). A demonstração de que $\pi(t) - \lambda t$ pertence à classe $\mathcal{D}$ não é difícil e pode ser feita por um processo análogo ao utilizado para o caso de um processo de Poisson não homogêneo, na seção 3.2.2.

Exemplo 3.6:

Seja $\Pi = \left\{ \pi(t) , t \geq 0 \right\}$ um processo de Poisson com parâmetro $\lambda > 0$, $\tau_1 = \inf \left\{ t \geq 0 : \pi(t) = 1 \right\}$ e $\tilde{\Pi} = \left\{ \tilde{\pi}(t) , t \geq 0 \right\}$ com

$$\tilde{\pi}(t) = \pi(t \wedge \tau_1).$$

Neste caso, a decomposição tem a forma

$$\tilde{\pi}(t) = \left[\pi(t \wedge \tau_1) - \lambda(t \wedge \tau_1) \right] + \lambda(t \wedge \tau_1) .$$

O processo crescente previsível $A = \left\{ A(t) , t \geq 0 \right\}$ que aparece nos teoremas e exemplos acima, é chamado de *compensador* do processo pontual $N = \left\{ N(t) , t \geq 0 \right\}$.

3.2 - Processo de Poisson Não Homogêneo

Definição 3.8:

Um processo de Poisson não homogêneo é um processo pontual $\left\{ N(t) , t \geq 0 \right\}$, definido em um espaço de probabilidades $(\Omega, \mathcal{F}, P)$ que satisfaz as seguintes condições:

- 1) $P \left(N(0) = 0 \right) = 1;$
- 2) $\left\{ N(t) , t \geq 0 \right\}$ possui incrementos independentes;

$$3) P\left(N(t+h) - N(t) = 1 \mid \mathcal{F}_{t-}^N \right) = \lambda(t)h + o(h); e$$

$$4) P\left(N(t+h) - N(t) \geq 2 \mid \mathcal{F}_{t-}^N \right) = o(h),$$

onde λ é a função intensidade do processo e

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

Note que, se tomarmos $\lambda(t) = \lambda$, $\left\{ N(t), \mathcal{F}_t^N, t \geq 0 \right\}$ passa a ser 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 $\mathcal{F}_s^N = \sigma\left\{ N(u) : u \leq s \right\}$, i.e., $P(A \cap B) = P(A)P(B)$, $\forall A \in \sigma(N(t) - N(s))$, $s < t$, e $\forall B \in \mathcal{F}_s^N$.

Já a condição 4 significa que a probabilidade de ocorrerem pelo menos 2 eventos simultaneamente é zero, i.e., $N(t) - N(t-) = 0$ ou 1 (por este motivo N é dito ser *simples*).

Os processos de Poisson são caracterizados, também, por possuírem intensidades determinísticas, i.e., num modelo de intensidade multiplicativa, onde $\lambda(t, \omega) = \alpha(t)Y(t, \omega)$, o fato de $Y(t, \omega) \equiv 1$, para $\forall \omega \in \Omega$, é equivalente ao processo pontual ser um processo de Poisson (ver Jacod e Shiryaev (1987), p. ex.).

A seguir, são dados alguns resultados sobre processos de Poisson não homogêneos.

Proposição 3.1:

Um processo de Poisson não homogêneo é tal que

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

e $0 \leq s \leq t$.

Demonstração:

Como a função geratriz de probabilidade de uma variável aleatória inteira não negativa determina unicamente a distribuição de probabilidade desta variável, vamos calcular $E\left[z^{N(t)-N(s)} \right]$, $|z| \leq 1$.

Seja $\psi(z, t) = E\left[z^{N(t)} \right]$. Como N possui incrementos independentes,

temos

$$\begin{aligned} \psi(z, t+h) &= E\left[z^{N(t+h)} \right] = E\left[z^{N(t+h)-N(t)+N(t)} \right] = \\ &= \psi(z, t) E\left[z^{N(t+h)-N(t)} \right]. \end{aligned}$$

Como consequência da definição de função geratriz de probabilidade,

$$\begin{aligned} \frac{1}{h} \left\{ E \left[z^{N(t+h)-N(t)} \right] - 1 \right\} &= \frac{1}{h} \left\{ P \left(N(t+h) - N(t) = 0 \right) - 1 \right\} + \\ + z \frac{1}{h} P \left(N(t+h) - N(t) = 1 \right) &+ \frac{1}{h} \sum_{n=2}^{\infty} z^n P \left(N(t+h) - N(t) = n \right). \end{aligned}$$

As condições 3 e 4 da definição 3.8 implicam que

$$\lim_{h \rightarrow 0} \left\{ \frac{1}{h} P \left(N(t+h) - N(t) = 0 \right) - 1 \right\} = -\lambda(t),$$

$$\lim_{h \rightarrow 0} \frac{1}{h} P \left(N(t+h) - N(t) = 1 \right) = \lambda(t) \text{ e}$$

$$\lim_{h \rightarrow 0} \sum_{n=2}^{\infty} z^n \frac{1}{h} P \left(N(t+h) - N(t) = n \right) = 0,$$

pela condição 4 e porque $|z| \leq 1$.

Conseqüentemente, temos que

$$\lim_{h \rightarrow 0} \frac{1}{h} \left\{ E \left[z^{N(t+h)-N(t)} \right] - 1 \right\} = -\lambda(t) + z\lambda(t) = (z-1)\lambda(t).$$

Como $\psi(z, t+h) = \psi(z, t) \left\{ E \left[z^{N(t+h)-N(t)} \right] - 1 \right\} + \psi(z, t)$ temos

$$\frac{\psi(z, t+h) - \psi(z, t)}{h} = \psi(z, t) \frac{1}{h} \left\{ E \left[z^{N(t+h)-N(t)} \right] - 1 \right\} \rightarrow$$

$$\Rightarrow \frac{\partial \psi(z, t)}{\partial t} = \psi(z, t)(z-1)\lambda(t), \text{ para } |z| < 1, t > 0.$$

Sob a condição inicial $\psi(z, 0) = 1$, a solução da equação diferencial acima é dada por:

$$\psi(z, t) = \exp \left[(z-1) \int_0^t \lambda(u) du \right] .$$

Para $0 < s < t$,

$$E \left[z^{N(t)-N(s)} \right] = \frac{\psi(z, t)}{\psi(z, s)} = \exp \left[(z-1) \int_s^t \lambda(u) du \right] .$$

Portanto,

$$N(t) - N(s) \sim P \left(\int_s^t \lambda(u) du \right) .$$

Uma demonstração alternativa pode ser encontrada em Ross (1983).

Proposição 3.2:

Seja $N = \{ N(t) , t \geq 0 \}$ um processo de Poisson não homogêneo com função intensidade λ . Então o processo $\tilde{N} = \{ N(Z(t)) , t \geq 0 \}$ é um processo de Poisson homogêneo com função intensidade $\lambda \equiv 1$, onde:

$$Z(t) = \inf \left\{ v : \Lambda(v) > t \right\} \quad \text{e} \quad \Lambda(t) = \int_0^t \lambda(s) ds.$$

Demonstração:

$$\begin{aligned} E \left[\tilde{N}(t+s) - \tilde{N}(t) \mid \tilde{N}(u) , u \leq t \right] &= \\ = E \left[N(Z(t+s)) - N(Z(t)) \mid N(Z(u)) , u \leq t \right] &= \end{aligned}$$

$$\begin{aligned}
&= E \left[N(Z(t+s)) - N(Z(t)) \right] = \Lambda(Z(t+s)) - \Lambda(Z(t)) = \\
&= \Lambda \left(\inf \left\{ v : \Lambda(v) > t+s \right\} \right) - \Lambda \left(\inf \left\{ v : \Lambda(v) > t \right\} \right) = t + s - t = s,
\end{aligned}$$

pois Z é uma função não decrescente, $\{ N(t) , t \geq 0 \}$ possui incrementos independentes e Λ é uma função contínua. ■

Proposição 3.3:

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

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

i.e., o tempo transcorrido entre dois saltos consecutivos possui densidade dada por

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

onde $X_{n+1} = T_{n+1} - T_n$ e $T_0 = 0$.

Demonstração:

O evento $A = \left\{ \Lambda(T_{n+1}) > \Lambda(T_n + s) \right\}$ está contido no evento

$\left\{ T_{n+1} - T_n > s \right\}$, que, por sua vez, está contido no evento

$B = \left\{ \Lambda(T_{n+1}) \geq \Lambda(T_n + s) \right\}$. Então,

$$\begin{aligned} P\left(A \mid T_1=t_1, \dots, T_n=t_n \right) &\leq P\left(X_{n+1} > s \mid T_1=t_1, \dots, T_n=t_n \right) \leq \\ &\leq P\left(B \mid T_1=t_1, \dots, T_n=t_n \right). \end{aligned}$$

Os valores $t_1, \dots, t_n$ determinam $\Lambda(t_1), \dots, \Lambda(t_n)$. Portanto, pela proposição anterior, $\Lambda(t_{n+1}) - \Lambda(t_n)$ é independente de $t_1, \dots, t_n$, e possui distribuição exponencial com parâmetro 1. Então,

$$\begin{aligned} P\left(A \mid T_1=t_1, \dots, T_n=t_n \right) &= \\ &= P\left\{ \Lambda(T_{n+1}) - \Lambda(T_n) > \Lambda(T_n + s) - \Lambda(T_n) \mid T_1=t_1, \dots, T_n=t_n \right\} = \\ &= \exp\left[- \left(\Lambda(t_n + s) - \Lambda(t_n) \right) \right] = P\left(B \mid T_1=t_1, \dots, T_n=t_n \right), \end{aligned}$$

pois $P\left(Y \geq t \right) = P\left(Y > t \right)$ para qualquer variável aleatória Y com distribuição exponencial. ■

Proposição 3.4:

Conicionados a $N(1) = n$, os tempos de ocorrência dos eventos, i.e., $0 < T_1 < \dots < T_n \leq 1$, são distribuídos como n estatísticas de ordem de uma amostra aleatória de n observações da densidade

$$f(t) = \frac{\lambda(t)}{\int_0^1 \lambda(s) ds}, \quad 0 \leq t \leq 1. \quad \text{Ou seja,}$$

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

$$0 < t_1 < \dots < t_n \leq 1.$$

Demonstração:

Sejam $X_{n+1} = T_{n+1} - T_n$ e $X_1 = T_1$. Então, $|J| = 1$ e

$$\begin{aligned} f_{T_1, \dots, T_n}(t_1, \dots, t_n) &= f_{X_1, \dots, X_n}(t_1, \dots, t_n) = \\ &= f_{X_n | X_1, \dots, X_{n-1}}(t_n | t_1, \dots, t_{n-1}) \dots f_{X_2 | X_1}(t_2 | t_1) f_{X_1}(t_1) = \\ &= \lambda(t_n) \exp \left[- \int_{t_{n-1}}^{t_n} \lambda(s) ds \right] \dots \lambda(t_1) \exp \left[- \int_{t_0}^{t_1} \lambda(s) ds \right] = \\ &= \left[\prod_{i=1}^n \lambda(t_i) \right] \exp \left[- \int_0^{t_n} \lambda(s) ds \right], \text{ para } 0 < t_1 < \dots < t_n. \end{aligned}$$

$$\begin{aligned} f_{T_1, \dots, T_n, N(1)}(t_1, \dots, t_n, n) &= \\ &= P \left(N(1) = n \mid T_1 = t_1, \dots, T_n = t_n \right) f_{T_1, \dots, T_n}(t_1, \dots, t_n) = \\ &= P \left(T_{n+1} > 1 \mid T_1 = t_1, \dots, T_n = t_n \right) f_{T_1, \dots, T_n}(t_1, \dots, t_n) = \\ &= \exp \left[- \int_{t_n}^1 \lambda(s) ds \right] \left[\prod_{i=1}^n \lambda(t_i) \right] \exp \left[- \int_0^{t_n} \lambda(s) ds \right] = \end{aligned}$$

$$= \left[\prod_{i=1}^n \lambda(t_i) \right] \exp \left[- \int_0^1 \lambda(s) ds \right].$$

Portanto,

$$\begin{aligned} f_{T_1, \dots, T_n | N(1)}(t_1, \dots, t_n | n) &= \frac{f_{T_1, \dots, T_n, N(1)}(t_1, \dots, t_n, n)}{P[N(1) = n]} = \\ &= \frac{n! \prod_{i=1}^n \lambda(t_i)}{\left[\int_0^1 \lambda(s) ds \right]^n}, \text{ para } 0 < t_1 < \dots < t_n \leq 1. \quad \blacksquare \end{aligned}$$

Exemplos de aplicações de processos de Poisson podem ser encontrados em Cox e Lewis (1966), Lewis (1972), Lewis e Shedler (1976a), Steinijs (1976) e Reymont (1976).

3.2.1 - Descrição do Experimento Estatístico

As trajetórias de um processo pontual são funções escada com saltos de magnitude 1, contínuas à direita com limite à esquerda em cada ponto $t \in [0, \infty)$. Podemos, então, tomar como espaço de observações Ω' o espaço

$$\mathcal{D}[0,1] = \left\{ g: [0,1] \rightarrow \mathbb{R} : g \text{ é contínua à direita c/ limites à esquerda} \right\}.$$

Claramente, este espaço contém muitos elementos além das trajetórias de um processo pontual. No entanto, sua adoção como espaço de observações é útil por sua estrutura de espaço métrico, descrita a

seguir.

Nosso espaço mensurável será dado por $\left(\mathbb{D}[0,1] , B(\mathbb{D}[0,1]) \right)$, onde $B(\mathbb{D}[0,1])$ é a σ -álgebra de Borel de $\mathbb{D}[0,1]$, i.e., a σ -álgebra gerada pelos conjuntos abertos do espaço $\mathbb{D}[0,1]$, segundo a métrica de Skorohod, dada pela expressão:

$$d_S(x,y) = \inf_{h \in H} \left[\sup_{t \in [0,1]} |x(t) - y(h(t))| + \sup_{t \in [0,1]} |t - h(t)| \right],$$

onde

$$H = \left\{ h:[0,1] \rightarrow [0,1] : h(0) = 0, h(1) = 1, h \text{ é contínua e estritamente crescente} \right\}.$$

Parthasarathy (1967) e Billingsley(1968), por exemplo, descrevem detalhadamente tal métrica, que torna o espaço $\mathbb{D}[0,1]$ separável, ao contrário da métrica uniforme

$$d(x,y) = \sup_{t \in [0,1]} |x(t) - y(t)|.$$

Observe que, para duas trajetórias ω e $\omega' \in \mathbb{D}[0,1]$, que contenham o mesmo número de saltos, temos que

$$d_S(\omega, \omega') = \max_{i=1, \dots, N(1, \omega)} \left\{ |\tau_i - \tau'_i| \right\},$$

onde τ_i e τ'_i representam o i -ésimo salto dentro das trajetórias ω e

ω' , respectivamente.

A σ -álgebra $B(D[0,1])$ coincide com aquela gerada pelos conjuntos cilíndricos

$$I(t_1, \dots, t_n) = \left\{ x \in \Omega' : (x(t_1), \dots, x(t_n)) \in B, B \in B(\mathbb{R}^n) \right\}$$

(ver Shiriyayev (1984)).

Neste espaço mensurável, definimos as medidas de probabilidade P_λ , $\lambda \in \Theta$, de tal forma que

$$P_\lambda(A) = P_\lambda \left\{ \omega : N(., \omega) \in A \right\}, A \in B(D[0,1]),$$

onde $\left\{ N(t), t \geq 0 \right\}$ tem λ como sua função intensidade.

Nosso experimento estatístico será considerado como sendo o "produto" dos experimentos $\varepsilon_i = \left\{ \Omega'_i, \mathcal{F}'_i, \left\{ P_{i\lambda}, \lambda \in \Theta \right\} \right\}$, $i \in \mathbb{N}$, onde $\Omega'_i = D[0,1]$, $\mathcal{F}'_i = B(D[0,1])$ e $P_{i\lambda} = P_\lambda$.

Aqui, adotaremos como espaço paramétrico um subconjunto do espaço $L_2([0,1], dt)$, i.e., o espaço formado pelas funções $f: [0,1] \rightarrow \mathbb{R}$ que satisfazem $\int_0^1 [f(t)]^2 dt < \infty$.

$$\Theta = \left\{ \lambda : \lambda \in L_2([0,1],dt) , \inf_{t \in [0,1]} \lambda(t) > 0 , \sup_{t \in [0,1]} \lambda(t) < \infty \right\}.$$

A razão para se definir Θ desta forma reside no fato de $L_2([0,1],dt)$ ser um espaço de Hilbert, o que possibilita a definição de diferentes "sieves" a partir de diferentes bases deste espaço.

3.2.2 - Decomposição de Doob-Meyer para um Processo de Poisson não Homogêneo

Como já vimos, $\{ N(t) , t \geq 0 \}$ é um submartingale. Podemos, então, obter a decomposição de Doob-Meyer para o processo de Poisson não homogêneo. Neste caso, $N(t) = M(t) + A(t)$, onde $M(t) = N(t) - \int_0^t \lambda(s)ds$ é uma martingale contínua à direita e uniformemente integrável. De fato,

$$\begin{aligned} E \left[M(t) \mid \mathcal{F}_s \right] &= E \left[N(t) - \int_0^t \lambda(u)du \mid \mathcal{F}_s \right] = \\ &= E \left[N(t) - N(s) \mid \mathcal{F}_s \right] + E \left[N(s) \mid \mathcal{F}_s \right] - \int_0^t \lambda(u)du = \\ &= \int_0^t \lambda(u)du + N(s) - \int_0^t \lambda(u)du = N(s) - \int_0^s \lambda(u)du = M(s) , \end{aligned}$$

pois λ é determinística.

Além disso, pela própria definição do espaço Ω' , as trajetórias de $M(t)$ são contínuas à direita.

A filtragem $\left\{ \mathcal{F}_t, t \geq 0 \right\}$ é contínua à direita em decorrência do seguinte lema (ver Liptser e Shiriyayev (1978)).

Lema 3.1:

Seja um espaço de eventos elementares Ω tal que, para todo $t \geq 0$ e todo $\omega \in \Omega$, existe $\omega' \in \Omega$ tal que $N(s, \omega') = N(t \wedge s, \omega)$, para todo $s > 0$. Então, a família de σ -álgebras $\left\{ \mathcal{F}_t^N, t \geq 0 \right\}$ é contínua à direita.

Claramente, $\Omega' = D[0,1]$ satisfaz a condição deste lema.

Note, ainda, que $N(t) - \int_0^t \lambda(s)ds$ é uniformemente integrável.

Definindo

$$C_x = \left\{ \left| N(t) - \int_0^t \lambda(s)ds \right| > x \right\},$$

temos

$$\begin{aligned} & \sup_{t \in [0,1]} \int_{C_x} \left| N(t) - \int_0^t \lambda(s)ds \right| dP_\lambda \\ & \leq \sup_{t \in [0,1]} \left[\int_{C_x} N(t) dP_\lambda + \int_{C_x} \left[\int_0^t \lambda(s)ds \right] dP_\lambda \right] \end{aligned}$$

$$\leq \sup_{t \in [0,1]} \int_{C_x} N(t) dP_\lambda + \sup_{t \in [0,1]} \left[\int_0^t \lambda(s) ds \right] P_\lambda(C_x).$$

Como

$$\lim_{x \rightarrow \infty} \int_{C_x} N(t) dP_\lambda = 0, \text{ para todo } t \in [0,1],$$

pois $E[N(t)] < \infty$, para todo $t \in [0,1]$, e $C_x \rightarrow \emptyset$, quando $x \rightarrow \infty$, então

$$\lim_{x \rightarrow \infty} \sup_{t \in [0,1]} \int_{C_x} N(t) dP_\lambda = 0.$$

Além disso,

$$\lim_{x \rightarrow \infty} \sup_{t \in [0,1]} \left[\int_0^t \lambda(s) ds \right] P_\lambda(C_x) = \left[\int_0^1 \lambda(s) ds \right] \lim_{x \rightarrow \infty} P_\lambda(C_x) = 0.$$

$A(t) = \int_0^t \lambda(s) ds$ é, então, o compensador do processo de Poisson não homogêneo (já que, para este tipo de processo, $\Lambda(t)$ é determinístico e, portanto, previsível).

3.2.3 - Derivada de Radon-Nikodym

O que nos interessa, de fato, é a obtenção de estimadores de máxima verossimilhança para a função $\lambda(t)$. Devemos, portanto, determinar a função que desejamos maximizar. Como nosso espaço de observações é um espaço abstrato, não temos nenhuma medida que seja diretamente equivalente à medida de Lebesgue em $\mathbb{R}^n$. Neste caso, adota-se a derivada de Radon-Nikodym como função de verossimilhança

caso tenhamos uma medida dominante em nosso experimento. Como foi visto no capítulo anterior, a escolha da medida dominante é irrelevante para a obtenção do estimador de máxima verossimilhança (caso haja mais de uma medida nesta condição).

Sabe-se que, para qualquer processo de Poisson com função intensidade λ , $P_\lambda \ll P_1$, onde P_1 é a medida de probabilidade induzida sobre $\left(\mathbb{D}[0,1], B(\mathbb{D}[0,1]) \right)$ pelas funções amostrais de um processo de Poisson homogêneo com $\lambda \equiv 1$ (ver Karr (1987), p. ex.).

Pelo teorema de Radon-Nikodym existe, então, $p(\omega, \lambda) = \frac{dP_\lambda}{dP_1}(\omega)$ tal

que $P_\lambda(A) = \int_A p(\lambda) dP_1$, para todo $A \in \mathcal{F}' = B(\mathbb{D}[0,1])$.

Proposição 3.5:

Para um processo de Poisson, temos

$$p(\omega, \lambda) = \frac{dP_\lambda}{dP_1}(\omega) = \exp \left[\int_0^1 \ln[\lambda(t)] dN(t, \omega) + \int_0^1 [1 - \lambda(t)] dt \right].$$

Uma maneira não rigorosa de se chegar a esta expressão é a seguinte: fixada uma trajetória ω^* cujos pontos de salto sejam

$t_1, \dots, t_k$, calculemos o quociente $\frac{P_\lambda(A_n)}{P_1(A_n)}$, onde

$$A_n = \left\{ \omega : T_1(\omega) \in \left(t_1 - \frac{\varepsilon}{n}, t_1 + \frac{\varepsilon}{n} \right), \dots, T_k(\omega) \in \left(t_k - \frac{\varepsilon}{n}, t_k + \frac{\varepsilon}{n} \right), \right. \\ \left. N(1, \omega) = k \right\}$$

e T_j é a variável aleatória que designa o j -ésimo salto de cada trajetória ω .

$$\varepsilon \text{ deve ser tal que } \varepsilon < \frac{1}{2} \left[\min_{i=1, \dots, k} \left\{ t_i - t_{i-1}, 1 - t_k \right\} \right].$$

Temos, então,

$$\begin{aligned} P_\lambda(A_n) &= P\left(N\left(t_1 - \frac{\varepsilon}{n}\right) = 0 \right) P\left(N\left(t_1 + \frac{\varepsilon}{n}\right) - N\left(t_1 - \frac{\varepsilon}{n}\right) = 1 \right) \\ &P\left(N\left(t_2 - \frac{\varepsilon}{n}\right) - N\left(t_1 + \frac{\varepsilon}{n}\right) = 0 \right) \dots P\left(N\left(t_k + \frac{\varepsilon}{n}\right) - N\left(t_k - \frac{\varepsilon}{n}\right) = 1 \right) \\ &P\left(N(1) - N\left(t_k + \frac{\varepsilon}{n}\right) = 0 \right) = \\ &= \exp\left[- \int_0^{t_1 - \varepsilon/n} \lambda(s) ds \right] \exp\left[- \int_{t_1 - \varepsilon/n}^{t_1 + \varepsilon/n} \lambda(s) ds \right] \int_{t_1 - \varepsilon/n}^{t_1 + \varepsilon/n} \lambda(s) ds \\ &\exp\left[- \int_{t_1 + \varepsilon/n}^{t_2 - \varepsilon/n} \lambda(s) ds \right] \dots \exp\left[- \int_{t_k - \varepsilon/n}^{t_k + \varepsilon/n} \lambda(s) ds \right] \int_{t_k - \varepsilon/n}^{t_k + \varepsilon/n} \lambda(s) ds \\ &\exp\left[- \int_{t_k + \varepsilon/n}^1 \lambda(s) ds \right] = \\ &= \exp\left[- \int_0^1 \lambda(s) ds \right] \prod_{i=1}^k \int_{t_i - \varepsilon/n}^{t_i + \varepsilon/n} \lambda(s) ds. \end{aligned}$$

Fazendo $\lambda \equiv 1$, temos

$$P_1(A_n) = e^{-1} \prod_{i=1}^k (2\varepsilon/n).$$

Portanto,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{P_\lambda(A_n)}{P_1(A_n)} &= \exp \left[- \int_0^1 [1 - \lambda(s)] ds \right] \prod_{i=1}^k \lim_{n \rightarrow \infty} \frac{1}{2\varepsilon/n} \int_{t_i - \varepsilon/n}^{t_i + \varepsilon/n} \lambda(s) ds = \\ &= \exp \left[- \int_0^1 [1 - \lambda(s)] ds \right] \prod_{i=1}^k \lambda(t_i) = \\ &= \exp \left[\sum_{i=1}^n \ln \lambda(t_i) + \int_0^1 [1 - \lambda(s)] ds \right] = \\ &= \exp \left[\int_0^1 \ln[\lambda(t)] dN(t, \omega^*) + \int_0^1 [1 - \lambda(s)] ds \right]. \end{aligned}$$

Este procedimento guarda uma certa analogia com o que é adotado por Grenander (1981), que utiliza um resultado conhecido como teorema da convergência da razão de verossimilhança, apresentado inicialmente em Grenander (1950).

3.2.4 - Métrica Adotada para o Espaço de Parâmetros Θ .

A métrica natural a ser adotada é a métrica de $L_2([0,1], dt)$, dada por:

$$d(\lambda_1, \lambda_2) = \| \lambda_1 - \lambda_2 \|_2 = \sqrt{\int_0^1 |\lambda_1(t) - \lambda_2(t)|^2 dt}.$$

No capítulo anterior, foi dito que a condição 2 do teorema 2.2 "sugere" uma métrica para Θ . Para um processo de Poisson não homogêneo, a entropia formal é dada por (ver Karr (1987)):

$$\begin{aligned} H(\lambda_0, \lambda) &= E \left[\ln(p(\omega, \lambda)) \right] = \\ &= E_{\lambda_0} \left[\int_0^1 \ln[\lambda(t)] dN(t, \omega) + \int_0^1 [1 - \lambda(t)] dt \right] = \\ &= \int_0^1 \left[1 - \lambda(s) + \lambda_0(s) \ln[\lambda(s)] \right] ds. \end{aligned}$$

$$\begin{aligned} \text{Portanto, } \lim_{m \rightarrow \infty} \left[H(\lambda_0, \lambda_0) - H(\lambda_0, \lambda_m) \right] &= 0 \Rightarrow \\ \Rightarrow \lim_{m \rightarrow \infty} \int_0^1 \left[\left(\lambda_m(s) - \lambda_0(s) \right) + \lambda_0(s) \ln \left(\lambda_0(s) / \lambda_m(s) \right) \right] ds &= 0 \Rightarrow \\ \Rightarrow \lambda_m(s) - \lambda_0(s) + \lambda_0(s) \ln \left(\lambda_0(s) / \lambda_m(s) \right) &\rightarrow 0 \end{aligned}$$

para quase todo $s \in [0,1]$, pois

$$\begin{aligned} \lambda_m(s) - \lambda_0(s) + \lambda_0(s) \ln \left(\lambda_0(s) / \lambda_m(s) \right) &\geq 0 \Rightarrow \\ \Rightarrow \left| \lambda_m(s) - \lambda_0(s) \right| &\rightarrow 0 \end{aligned}$$

para quase todo $s \in [0,1] \Rightarrow$

$$\Rightarrow k_m = \inf \left\{ M : \text{Leb} \left\{ t : \left| \lambda_m(t) - \lambda_0(t) \right| > M \right\} = 0 \right\} \rightarrow 0 \Rightarrow$$

$$\Rightarrow \int_0^1 |\lambda_0(t) - \lambda_m(t)|^2 dt < k_m^2 \rightarrow 0 \Rightarrow \|\lambda_m - \lambda_0\|_2 \rightarrow 0.$$

3.2.5 - Escolha do "Sieve".

Como definido anteriormente, o método de "sieves" consiste na obtenção de uma seqüência $\{S_m\}$ de subconjuntos de Θ com $\bigcup_{m=1}^{\infty} S_m$ denso em Θ . No caso de um processo de Poisson não homogêneo com espaço de parâmetros Θ , queremos estimar $\lambda \in \Theta$. Para isso, consideraremos aqui dois "sieves" diferentes: o "sieve" de histograma e o "sieve" de Fourier. Antes de descrevê-los, no entanto, convém justificar a necessidade da utilização de um método alternativo para a obtenção dos estimadores de máxima verossimilhança de λ .

Se possuírmos uma amostra i.i.d. de tamanho n de nosso espaço Ω' , teremos

$$\begin{aligned} L_n(\lambda; \underline{\omega}) &= \prod_{i=1}^n \frac{dP_{\lambda}(\omega_i)}{dP_1} = \prod_{i=1}^n \exp \left[\int_0^1 \ln[\lambda(t)] dN(t, \omega_i) + \int_0^1 [1 - \lambda(t)] dt \right] = \\ &= \prod_{i=1}^n \exp \left[\sum_{j=1}^{r_i} \ln[\lambda(\tau_{i,j})] + \int_0^1 [1 - \lambda(t)] dt \right]. \end{aligned}$$

onde $\tau_{i,j}$ representa o instante do j -ésimo salto dentro da i -ésima trajetória e r_i designa o número de saltos da i -ésima trajetória, i.e., $r_i = N(1, \omega_i)$.

Observe que, se fizermos $\lambda(t)$ assumir valores cada vez maiores nos pontos $\tau_{i,j}$, de tal forma que $\int_0^1 (1 - \lambda(t))dt$ permaneça constante, teremos $L_n(\lambda; \underline{\omega})$ tão grande quanto se queira, e conseqüentemente, não conseguiremos encontrar um elemento $\lambda_0 \in \Theta$ que maximize $L_n(\lambda; \underline{\omega})$, ou seja, torna-se necessária a utilização de um método alternativo para a estimação de λ_0 .

3.2.5.1 - Sieve de Histograma.

No "sieve" de histograma, temos

$$S_m = \left\{ g : [0,1] \rightarrow \mathbb{R}^+ : \text{para todo } t \in [0,1], g(t) = \sum_{k=1}^m \theta_k 1_{T_k}(t) \right\},$$

onde $T_k = [(k-1)/m, k/m)$, $k = 1, \dots, m-1$ e $T_m = [(m-1)/m, 1]$.

Para aplicação do método de "sieves", o que devemos fazer, então, é obter, para cada S_m , um elemento $\hat{\lambda}_m \in S_m$ (o que equivale a obter um vetor $\hat{\underline{\theta}}_m$) que maximize

$$\begin{aligned} \mathcal{L}_n(\lambda; \underline{\omega}) &= \mathcal{L}_n(\underline{\theta}_m; \underline{\omega}) = \ln \left(L_n(\underline{\theta}_m; \underline{\omega}) \right) \\ &= \ln \left[\prod_{i=1}^n \exp \left[\int_0^1 \ln[\lambda(t)] dN(t, \omega_i) + \int_0^1 [1 - \lambda(t)] dt \right] \right] \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \sum_{j=1}^{r_i} \ln[\lambda(\tau_{i,j})] + n \int_0^1 [1 - \lambda(t)] dt \\
&= \sum_{i=1}^n \sum_{j=1}^{r_i} \ln \left[\sum_{k=1}^m \theta_{m,k} 1_{T_{m,k}}(\tau_{i,j}) \right] + n \int_0^1 \left[1 - \sum_{k=1}^m \theta_{m,k} 1_{T_{m,k}}(t) \right] dt \\
&= \sum_{i=1}^n \sum_{j=1}^{r_i} \ln \left[\sum_{k=1}^m \theta_{m,k} 1_{T_{m,k}}(\tau_{i,j}) \right] + n \left(1 - \sum_{k=1}^m \frac{\theta_{m,k}}{m} \right).
\end{aligned}$$

Neste caso, podemos encontrar o vetor $\hat{\underline{\theta}}_m$ que maximiza $\mathcal{L}_n(\underline{\theta}_m; \omega)$ analiticamente, calculando o vetor de primeiras derivadas e a matriz de segundas derivadas (matriz Hessiana) de $\mathcal{L}_n(\underline{\theta}_m; \omega)$ com relação a $\underline{\theta}_m$.

$$\begin{aligned}
\frac{\partial \mathcal{L}_n(\underline{\theta}_m; \omega)}{\partial \theta_{m,s}} &= \sum_{i=1}^n \sum_{j=1}^{r_i} \frac{\partial}{\partial \theta_{m,s}} \left(\ln \left[\sum_{k=1}^m \theta_{m,k} 1_{T_{m,k}}(\tau_{i,j}) \right] \right) - \frac{n}{m} = \\
&= \sum_{i=1}^n \sum_{j=1}^{r_i} \left(\frac{1_{T_{m,s}}(\tau_{i,j})}{\sum_{k=1}^m \theta_{m,k} 1_{T_{m,k}}(\tau_{i,j})} \right) - \frac{n}{m}.
\end{aligned}$$

$$\frac{\partial^2 \mathcal{L}_n(\underline{\theta}_m; \omega)}{\partial \theta_s \partial \theta_v} = \frac{\partial}{\partial \theta_v} \left(\frac{\partial \mathcal{L}_n(\underline{\theta}_m; \omega)}{\partial \theta_s} \right) = - \sum_{i=1}^n \sum_{j=1}^{r_i} \left(\frac{1_{T_{m,s}}(\tau_{i,j}) 1_{T_{m,v}}(\tau_{i,j})}{\left[\sum_{k=1}^m \theta_{m,k} 1_{T_{m,k}}(\tau_{i,j}) \right]^2} \right) \rightarrow$$

$$\Rightarrow = \frac{\partial^2 \mathcal{L}_n(\underline{\theta}_{-m}; \underline{\omega})}{\partial \theta_s \partial \theta_v} = \text{diag} \left(- \sum_{l=1}^n \sum_{j=1}^{r_l} \left(\frac{1_{T_{m,s}}(\tau_{l,j})}{\left[\sum_{k=1}^m \theta_{m,k} 1_{T_{m,k}}(\tau_{l,j}) \right]^2} \right), 1 \leq s \leq m \right).$$

Como a matriz Hessiana é negativa definida (é uma matriz diagonal com elementos negativos na diagonal principal) o vetor $\hat{\underline{\theta}}_{-m}$ que satisfizer

$$\left(\frac{\partial \mathcal{L}_n(\underline{\theta}_{-m}, \underline{\omega})}{\partial \theta_{m,s}}, 1 \leq s \leq m \right) = \underline{0}$$

maximizará a função $\mathcal{L}_n(\underline{\theta}_{-m}; \underline{\omega})$ (ver Lima (1981)).

$$\frac{\partial \mathcal{L}_n(\underline{\theta}_{-m}, \underline{\omega})}{\partial \theta_{m,s}} = 0 \Rightarrow \sum_{l=1}^n \sum_{j=1}^{r_l} \left(\frac{1_{T_{m,s}}(\tau_{l,j})}{\sum_{k=1}^m \theta_{m,k} 1_{T_{m,k}}(\tau_{l,j})} \right) - \frac{n}{m} = 0 \Rightarrow$$

$$\Rightarrow \hat{\theta}_{m,s} = \frac{m}{n} \# \left[\tau_{l,j} : \tau_{l,j} \in T_{m,s} \right].$$

$\bigcup_{m=1}^{\infty} S_m$ é denso em Θ , pela seguinte proposição (ver Royden(1968),p.244).

Proposição 3.6:

Seja $f \in L_p(\mu)$, $1 \leq p \leq \infty$. Então, dado $\varepsilon > 0$, existe uma função simples φ , nula fora de um conjunto de medida finita, tal que $\|f - \varphi\|_p < \varepsilon$.

$L_p(\mu) = \left\{ f : \Omega \rightarrow \mathbb{R} : \int |f|^p d\mu < \infty \right\}$, funções simples são aquelas que assumem apenas um número finito de valores e $\|\cdot\|_p$ designa a norma do espaço $L_p(\mu)$.

Karr (1987) obteve um importante resultado sobre a consistência do estimador de λ obtido utilizando-se o "sieve" de histograma.

Teorema 3.2:

Se m_n (dimensão do elemento S_m do "sieve") é escolhido de tal forma que $\sum_{n=1}^{\infty} \left(\frac{m_n^4}{n^2} \right) < \infty$, então $\|\hat{\lambda}_n - \lambda_0\|_1 \rightarrow 0$ P_{λ_0} -quase certamente, onde $\|\cdot\|_1$ designa a métrica do espaço $L_1([0,1], dt)$.

Isto implica que devemos ter $m_n < \sqrt[4]{n}$ para que

$\|\hat{\lambda}_n - \lambda_0\|_1 \rightarrow 0$ P_{λ_0} -quase certamente.

Um aspecto importante na utilização do "sieve" de histograma é sua simplicidade computacional.

3.2.5.2 - "Sieve" de Fourier

Como já foi dito, a inspiração para a construção do "sieve" de Fourier se origina da estrutura matemática do espaço $L_2([0,1],dt)$. A construção deste "sieve" é feita tomando S_m como um subconjunto do espaço vetorial gerado pelos primeiros m elementos de um sistema completo ortonormal de $L_2([0,1],dt)$, i.e., um sistema $\left\{ \varphi_k, k \in \mathbb{N} \right\}$ tal que

$$\langle \varphi_i, \varphi_j \rangle = \int_0^1 \varphi_i(t) \varphi_j(t) dt = \begin{cases} 0, & \text{se } i \neq j \\ 1, & \text{se } i = j \end{cases},$$

e onde $\langle z, \varphi_i \rangle = 0$, para todo $i \in \mathbb{N}$, implica $z = 0$.

Tal sistema é também conhecido como base de $L_2([0,1],dt)$, e cada elemento de $L_2([0,1],dt)$ pode ser expresso como combinação linear dos elementos do sistema $\left\{ \varphi_k, k \in \mathbb{N} \right\}$, i.e.,

$$\lambda(t) = \sum_{k=1}^{\infty} \theta_k \varphi_k(t), \text{ onde } \theta_k = \langle \lambda, \varphi_k \rangle = \int_0^1 \lambda(t) \varphi_k(t) dt,$$

são chamados de *coeficientes de Fourier* do elemento λ com respeito ao sistema $\left\{ \varphi_k, k \in \mathbb{N} \right\}$.

Observe que, para qualquer $\lambda \in L_2([0,1],dt)$, temos

$$\lambda^2 = \sum_{k=1}^{\infty} \theta_k^2 = \int_0^1 (\lambda(t))^2 dt < \infty \Rightarrow \theta_k \rightarrow 0, \text{ quando } k \rightarrow \infty.$$

Definimos, então,

$$S_m = \left\{ \lambda \in \Theta : \lambda(t) = \sum_{k=1}^m \theta_k \varphi_k(t) , \frac{1}{c_m} \leq \lambda(t) \leq c_m \right\} ,$$

onde $c_m \nearrow \infty$ tão rápido quanto se queira, e

$$\begin{aligned} \varphi_1(t) &= 1 , t \in [0,1] ; \\ \varphi_k(t) &= \begin{cases} \sqrt{2} \cos(2k\pi t) & , k \text{ par} \\ \sqrt{2} \sin(2(k-1)\pi t) & , k \text{ ímpar} \end{cases} . \end{aligned} \quad (1)$$

Para mostrar que $\bigcup_{m=1}^{\infty} S_m$ é denso em Θ , devemos mostrar que, dado

$\varepsilon > 0$, existe $m_0 \in \mathbb{N}$ tal que

$$\left\| \sum_{k=1}^{m_0} \theta_k^* \varphi_k - \lambda_0 \right\|_2 < \varepsilon$$

$$\text{com } \frac{1}{c_{m_0}} \leq \sum_{k=1}^{m_0} \theta_k^* \varphi_k(t) \leq c_{m_0} .$$

Como a série de Fourier de funções contínuas converge uniformemente e, além disso, para qualquer função $f \in L_2([0,1], dt)$, existe uma função contínua $\xi \in L_2([0,1], dt)$ tal que $\|f - \xi\|_2 < \varepsilon$ (ver Royden (1968), p.118), podemos tomar ξ e m_0 tais que:

- 1) $\xi(t) > 0$, para todo $t \in [0,1]$;
- 2) $\sup_{t \in [0,1]} |\lambda_0(t) - \xi(t)| < \delta$;
- 3) $\|\lambda_0 - \xi\|_2 < \varepsilon/2$;

$$4) \sup_{t \in [0,1]} \left| \sum_{k=1}^{m_0} \theta_k^* \varphi_k(t) - \xi(t) \right| < \delta ;$$

$$5) \left\| \sum_{k=1}^{m_0} \theta_k^* \varphi_k - \xi \right\|_2 < \varepsilon/2 ;$$

onde: $\delta = \frac{1}{2} \inf_{t \in [0,1]} \lambda_0(t)$ e $\left\{ \theta_k^*, k \in \mathbb{N} \right\}$ são os coeficientes de Fourier da função ξ .

Temos, então,

$$\sup_{t \in [0,1]} \left| \sum_{k=1}^{m_0} \theta_k^* \varphi_k(t) - \lambda_0(t) \right| < 2\delta = \inf_{t \in [0,1]} \lambda_0(t) ,$$

pela definição de métrica uniforme (ver seção 3.2.1), ou seja,

$$0 < \sum_{k=1}^{m_0} \theta_k^* \varphi_k(t) < \infty ,$$

para todo $t \in [0,1]$.

Além disso,

$$\left\| \sum_{k=1}^{m_0} \theta_k^* \varphi_k - \lambda_0 \right\|_2 < \varepsilon ,$$

pela desigualdade triangular.

Mostraremos, a seguir, que o "sieve" de Fourier satisfaz as condições do Teorema 2.1.

Para mostrarmos que, para todo $m \in \mathbb{N}$, todo $\theta \in S_m$ e todo $\varepsilon > 0$,

$$p\left[\omega, B_m(\theta, \varepsilon)\right] = \sup_{\theta \in B_m(\lambda, \varepsilon)} p(\omega, \theta)$$

é $\mathcal{F}'$ -mensurável, é suficiente mostrar que

$$\left\{ \omega : p(\omega, \theta) = \sum_{j=1}^{N(1, \omega)} \ln[\lambda(\tau_j)] > x \right\} \in \mathcal{F}', \text{ para todo } x \in \mathbb{R},$$

pois a função exponencial é contínua, $\int_0^1 [1 - \lambda(t)] dt$ é constante (por

λ ser determinística) e

$$\left\{ \omega : \sup_{\theta \in B_m(\lambda, \varepsilon)} p(\omega, \theta) > x \right\} = \bigcup_{\theta \in B_m(\lambda, \varepsilon)} \left\{ \omega : p(\omega, \theta) > x \right\}.$$

Como a função logarítmica é uma função contínua e $\theta \in S_m$ também é contínua, se ω for tal que

$$\sum_{j=1}^{N(1, \omega)} \ln[\theta(\tau_j)] > x,$$

então existe $\varepsilon_\omega > 0$ tal que

$$\sum_{j=1}^{N(1, \omega)} \ln[\lambda(\tau_j - \varepsilon_\omega)] > x.$$

Consequentemente, para todo ω' tal que

$$d_S(\omega, \omega') < \varepsilon_\omega,$$

temos

$$p(\omega', \theta) = \sum_{j=1}^{N(1, \omega')} \ln[\theta(\tau'_j)] > x$$

(note que, se $\varepsilon_\omega < 1$, ω' possui o mesmo número de saltos que ω).

Portanto,

$$\left\{ \omega : p(\omega, \theta) > x \right\}$$

é um subconjunto aberto de $\mathbb{D}[0,1]$, o mesmo ocorrendo com

$$\left\{ \omega : \sup_{\theta \in B_m(\lambda, \varepsilon)} p(\omega, \theta) > x \right\}.$$

A condição (1) do teorema 2.1 é, então, satisfeita.

A condição 2 será automaticamente satisfeita se provarmos que $p(\omega, \lambda)$ é contínua em λ (isto implica a semicontinuidade superior).

Novamente, como a função exponencial é contínua, é suficiente mostrar

que $\int_0^1 [1 - \lambda(t)] dt$ e $\int_0^1 \ln[\lambda(t)] dN(t, \omega)$ são contínuas em λ para cada

trajetória ω .

Seja $\left\{ \lambda_n, n \in \mathbb{N} \right\}$ uma seqüência tal que $\lambda_n \in S_m$ e $\| \lambda_n - \lambda \|_2 \rightarrow 0, \lambda \in S_m$.

$$\| \lambda_n - \lambda \|_2^2 = \sum_{k=1}^m (\theta_k^{(n)} - \theta_k)^2,$$

onde

$$\lambda_n(t) = \sum_{k=1}^m \theta_k^{(n)} \varphi_k(t) \text{ e } \lambda(t) = \sum_{k=1}^m \theta_k \varphi_k(t)$$

e $\varphi_k, k=1, \dots, m$, são dadas pela expressão (1).

Temos, então, que $\| \lambda_n - \lambda \|_2 \rightarrow 0 \Rightarrow \theta_k^{(n)} \rightarrow \theta_k, k = 1, \dots, m$,

quando $n \rightarrow \infty$, i.e., $\lambda_n(t) \rightarrow \lambda(t)$, para todo $t \in [0,1]$.

Como $\frac{1}{c_m} \leq \theta(t) \leq c_m, t \in [0,1]$, para todo $\theta \in S_m$, pelo Teorema

da Convergência Dominada

$$\lim_{n \rightarrow \infty} \int_0^1 [1 - \lambda_n(t)] dt = \int_0^1 \left[1 - \lim_{n \rightarrow \infty} \lambda_n(t) \right] dt = \int_0^1 [1 - \lambda(t)] dt.$$

Considere, agora, ω fixo e tal que $N(1, \omega) < \infty$. Temos, então, que

$$\frac{1}{c_m} < \theta(t) \leq c_m \Rightarrow \ln\left(\frac{1}{c_m}\right) \leq \ln[\theta(t)] \leq \ln[c_m] \Rightarrow$$

$$\Rightarrow |\ln[\theta(t)]| \leq \ln[c_m] \Rightarrow$$

$$\Rightarrow \int_0^1 |\ln[\lambda_n(t)]| dN(t, \omega) \leq N(1, \omega) \ln[c_m],$$

tomando $c_m > 1$.

Novamente, pelo Teorema da Convergência Dominada e pela continuidade da função logarítmica

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_0^1 \ln[\lambda_n(t)] dN(t, \omega) &= \int_0^1 \lim_{n \rightarrow \infty} \ln[\lambda_n(t)] dN(t, \omega) = \\ &= \int_0^1 \ln[\lambda(t)] dN(t, \omega). \end{aligned}$$

Isto completa a demonstração para a condição (2).

A limitação $\frac{1}{c_m} < \theta(t) \leq c_m$, $t \in [0, 1]$ e $\theta \in S_m$, implica que

$$\int_0^1 \ln[\theta(t)] dN(t, \omega) \leq N(1, \omega) \ln[c_m]$$

e que

$$\int_0^1 [1 - \theta(t)] dt \leq 1 - \frac{1}{c_m}.$$

Portanto,

$$\begin{aligned} \ln[p(\omega, \theta)] &= \int_0^1 \ln[\theta(t)] dN(t, \omega) + \int_0^1 [1 - \theta(t)] dt \leq \\ &\leq N(1, \omega) \ln[c_m] + 1 - \frac{1}{c_m}, \end{aligned}$$

para todo $\omega \in \Omega'$ tal que $N(1, \omega) < \infty$.

Consequentemente,

$$E_{\lambda_0} \left[\ln \left[p \left[\omega, B_m(\lambda, \varepsilon) \right] \right] \right] < \infty,$$

pois $P[N(1) = \infty] = 0$.

Precisamos, agora, provar que S_m é compacto.

Como os espaços L_p são completos (teorema de Riesz-Fischer) e toda seqüência (λ_i) convergente, onde $\lambda_i \in S_m$, converge para elementos de S_m , temos que S_m é completo. De fato,

$$\begin{aligned} \frac{1}{c_m} \leq \lambda_i(t) \leq c_m &\Rightarrow \frac{1}{c_m} \leq \lim_{i \rightarrow \infty} \lambda_i(t) \leq c_m \Rightarrow \\ \Rightarrow \frac{1}{c_m} \leq \sum_{k=1}^m \lim_{i \rightarrow \infty} \theta_{i,k} \varphi_k(t) \leq c_m &\Rightarrow \frac{1}{c_m} \leq \sum_{k=1}^m \varphi_k(t) \lim_{i \rightarrow \infty} \theta_{i,k} \leq c_m \Rightarrow \\ \Rightarrow \frac{1}{c_m} \leq \sum_{k=1}^m \theta_k \varphi_k(t) \leq c_m, \end{aligned}$$

onde $\theta_k = \lim_{i \rightarrow \infty} \theta_{i,k}$.

Além disso, S_m é totalmente limitado (i.e., dado $\varepsilon > 0$, existe

uma coleção finita de esferas de raio ε que cobre S_m) pois seu diâmetro é finito.

$$\begin{aligned} \text{diam } S_m &= \sup_{\lambda_1, \lambda_2 \in S_m} d(\lambda_1, \lambda_2) = \langle \lambda_1 - \lambda_2, \lambda_1 - \lambda_2 \rangle = \\ &= \int_0^1 | \lambda_1(t) - \lambda_2(t) |^2 dt \leq \int_0^1 \left(c_m - \frac{1}{c_m} \right)^2 dt < \infty . \end{aligned}$$

Portanto, por uma Proposição (Royden (1968), p.164) S_m é compacto, ou seja, satisfaz a condição 4 do teorema 2.1.

Finalmente, a condição 5.

Como

$$H(\lambda_0, \lambda_n) = E_{\lambda_0} \left[\ln \left[p(\omega, \lambda_n) \right] \right] < \infty$$

e

$$\ln[p(\omega, \lambda_n)] \leq N(1, \omega) \ln[c_m] + 1 - \frac{1}{c_m} ,$$

para $\lambda_n \in S_m$, então

$$\begin{aligned} \lim_n \lim_{\infty} E_{\lambda_0} \left[\ln \left[p(\omega, \lambda_n) \right] \right] &= E_{\lambda_0} \left[\lim_n \lim_{\infty} \ln \left[p(\omega, \lambda_n) \right] \right] = \\ &= E_{\lambda_0} \left[\ln \left[p(\omega, \lambda) \right] \right] , \end{aligned}$$

pelo Teorema da Convergência Dominada e pela continuidade da função logarítmica e de $p(\omega, \lambda)$, ou seja, $H(\lambda_0, \lambda)$ é contínua em λ .

Como $\bigcup_{m=1}^{\infty} S_m$ é denso em Θ e $\{S_m\}$ é crescente, existe $\lambda_m \in S_m$ tal

que $\| \lambda_m - \lambda_0 \|_2 \rightarrow 0$, quando $m \rightarrow \infty$. Portanto,

$$\sup_{\lambda \in S_m} H(\lambda_0, \lambda) \rightarrow H(\lambda_0, \lambda_0) , \text{ quando } m \rightarrow \infty$$

e

$$\lambda_m \rightarrow \lambda_0, \text{ quando } m \rightarrow \infty.$$

Mostramos, então, que para o "sieve" de Fourier existe uma seqüência m_n para a qual o conjunto de estimadores de máxima verossimilhança $M_{m_n}^n$ é consistente, i.e., $\lim_{n \rightarrow \infty} P(d(M_{m_n}^n, \lambda_0) > \varepsilon) = 0$, para todo $\varepsilon > 0$. Para a determinação da seqüência m_n seria necessário utilizar o teorema 2.2. No entanto, sua aplicação não é simples, segundo os próprios autores. Por este motivo, utilizamos diferentes seqüências m_n para os exemplos simulados.

CAPÍTULO 4

ASPECTOS PRÁTICOS NA APLICAÇÃO DO MÉTODO DE "SIEVES"

4.1 - Algoritmo para Simulação de Processos de Poisson Não Homogêneos

O algoritmo utilizado para simulação de um processo de Poisson não homogêneo foi o método da transformação da escala de tempo, que se baseia numa consequência da proposição 3.2:

Os pontos $t_1, t_2, \dots$ são pontos de salto de um processo de Poisson com $E[N(t)] = \Lambda(t)$ se e somente se $\Lambda(t_1), \Lambda(t_2), \dots$ são pontos de salto de um processo de Poisson com função intensidade $\lambda \equiv 1$.

Com este resultado é possível, então, gerar n trajetórias de um processo de Poisson com função intensidade $\lambda(t)$, $t \geq 0$, observadas no intervalo $[0,1]$, através do seguinte algoritmo:

- 1) faça $i = 1$;
- 2) faça $s = 0$;
- 3) gere um número aleatório u com distribuição $U[0,1]$;
- 4) faça $y = -\log(u)$;
- 5) faça $s = s + y$;
- 6) se $s > \Lambda(1)$, faça $i = i + 1$;
 se $i \leq n$, vá para 2;
 se $i > n$, vá para 7;

$$\text{se } s \leq \Lambda(1), \text{ grave } i \text{ e } Z(s) = \inf \left\{ v : \Lambda(v) > s \right\};$$

7) fim;

Obviamente, o processo de obtenção de $Z(t_i)$ a partir de Λ nem sempre é operacionalmente fácil ou mesmo possível, sendo necessário fazer uso de processos numéricos de inversão nos casos em que não seja possível obter facilmente uma expressão para $Z(t)$. Esta solução, no entanto, tornaria o processo extremamente lento.

Para contornar esse tipo de problema, Lewis e Shedler (1979) sugerem um procedimento bastante simples, baseado no seguinte

Teorema 4.1:

Considere um processo de Poisson não homogêneo $\left\{ N^*(t), t \geq 0 \right\}$ com função intensidade $\lambda^*(t)$, tal que o número de pontos $N^*(t_0)$ em um intervalo fixo $(0, t_0]$ possui distribuição de Poisson com parâmetro $\mu_0^* = \Lambda^*(t_0) - \Lambda^*(0)$. Sejam $t_1^*, t_2^*, \dots, t_{N^*(t_0)}^*$ os pontos do processo no intervalo $(0, t_0]$. Suponha que para $0 \leq t \leq t_0$, $\lambda(t) \leq \lambda^*(t)$. Para $i = 1, 2, \dots, n$, delete o ponto de salto t_i^* com probabilidade $1 - \lambda(t_i^*)/\lambda^*(t_i^*)$. Os pontos de salto restantes determinam uma trajetória de um processo de Poisson não homogêneo $\left\{ N(t), t \geq 0 \right\}$ com função intensidade λ no intervalo $(0, t_0]$.

A demonstração encontra-se em Lewis e Shedler (1979).

Obviamente, λ^* deve ser escolhida de tal forma que o cálculo de $Z(t_i^*)$, $i = 1, \dots, n$, seja fácil.

O algoritmo consiste, portanto, nos seguintes passos:

1) Gere uma trajetória de um processo de Poisson $\left\{ N^*(t), t \geq 0 \right\}$ com função intensidade $\lambda^*(t)$ num intervalo fixo $(0, t_0]$, segundo o algoritmo descrito anteriormente. Se o número de pontos gerados n^* for igual a zero, pare; não há nenhum ponto de salto na trajetória.

2) Denote os pontos (ordenados) por $t_1^*, t_2^*, \dots, t_n^*$. Faça $i = 1$ e $k = 0$.

3) Gere u_1 uniformemente distribuído entre 0 e 1.

Se $u_1 \leq \lambda(t_1^*)/\lambda^*(t_1^*)$, faça $k = k + 1$ e $t_k = t_1^*$.

4) Faça $i = i + 1$. Se $i \leq n^*$, vá para 3.

5) Se $i > n^*$, retorne $t_1, t_2, \dots, t_k$, e também k .

Estes pontos determinam uma trajetória de um processo de Poisson $\left\{ N(t), t \geq 0 \right\}$ com função intensidade $\lambda(t)$, $t \geq 0$, observada no intervalo $(0, t_0]$.

O problema de geração de trajetórias de processos de Poisson não homogêneos já foi abordado em diversos trabalhos, como Gilchrist (1976), Lewis e Shedler (1976b, 1977) e Patrow (1977), por exemplo.

4.2 - Procedimentos Práticos para Obtenção das Estimativas das Funções intensidade λ .

Obtidas as trajetórias de um processo de Poisson com a função intensidade desejada, procedeu-se a etapa seguinte: a estimação da função intensidade com base na amostra, para cada um dos dois "sieves" descritos no capítulo anterior.

Como já foi visto, este procedimento é equivalente a estimar os vetores $\theta_{-m_n} = (\theta_1, \dots, \theta_{m_n})$, onde $\hat{\lambda}(t) = \sum_{k=1}^{m_n} \theta_k \varphi_k(t)$.

Para o "sieve" de histograma é necessário, apenas, calcular o vetor

$$\left(\theta_{m_n,1}, \dots, \theta_{m_n,m_n} \right),$$

onde

$$\theta_{m_n,s} = \frac{m_n}{n} \# \left[\tau_{i,j} : \tau_{i,j} \in T_{m_n,s} \right] \text{ e}$$

$T_{m_n,s}$ foi definido no capítulo anterior.

No caso do "sieve" de Fourier, não é possível encontrar o vetor $\theta_{-m} \in S_m$ que maximize $\mathcal{L}_n(\theta_{-m}, \omega)$ analiticamente; é necessária a utilização de um método iterativo que nos leve a encontrar um vetor próximo da solução real do problema. O método utilizado foi o de

Newton-Raphson (ver Jenrich e Schluchter (1986)), onde a cada passo, temos $\theta_{-m}^{(r+1)} = \theta_{-m}^{(r)} - (1/2)^v H^{-1}(\theta_{-m}^{(r)}) S(\theta_{-m}^{(r)})$, $\theta_{-m}^{(r)}$ designando o vetor obtido no r-ésimo passo do processo, S o vetor de primeiras derivadas e H a matriz Hessiana (calculados no ponto $\theta_{-m}^{(r)}$). O número inteiro v é ajustado para se obter a convergência do processo.

Para escolha de um ponto $\theta_{-m}^{(0)} \in \mathbb{R}^n$ para iniciar o processo iterativo, foi adotado o procedimento de se calcular $\mathcal{L}_n(\theta_{-m}, \omega)$ para alguns pontos $\theta_{-m} \in \mathbb{R}^n$, escolhendo-se como $\theta_{-m}^{(0)}$ aquele onde $\mathcal{L}_n(\theta_{-m}, \omega)$ apresente o maior valor.

Como regra de parada, adota-se normalmente o fato de

$$\frac{\theta_{-m}^{(r+1)} - \theta_{-m}^{(r)}}{\theta_{-m}^{(r)}} < \delta ,$$

onde δ é um valor positivo próximo de zero (aqui adotamos $\delta = 0.001$).

Quando S_m for composto por funções positivas que sejam combinações lineares dos primeiros m elementos de uma base de $L_2([0,1], dt)$, é possível generalizar o cálculo de

$$\mathcal{L}_n(\theta_{-m}, \omega) = \sum_{i=1}^n \sum_{j=1}^r \ln \left(\sum_{k=1}^m \theta_k \varphi_k(\tau_{i,j}) \right) - n + n \sum_{k=1}^m \theta_k \int_0^1 \varphi_k(t) dt ,$$

construindo-se uma matriz A de dimensão $\left(\sum_{i=1}^n r_i \times m_n \right)$ da seguinte forma:

obtido o vetor de tempos de salto $T_i = \left(\tau_{i,1}, \dots, \tau_{i,r_i} \right)^t$ de cada

trajetória i , $i \in \{1, 2, \dots, n\}$, constroi-se um vetor $T = \left(T_1^t, \dots, T_n^t \right)^t$

e, a partir dele, a matriz A , onde $a_{u,v} = \varphi_v(\tau_{i,j})$, e o j -ésimo salto dentro da i -ésima trajetória determina a linha u da matriz.

$$A = \begin{pmatrix} \varphi_1(\tau_{1,1}) & \dots & \varphi_{m_n}(\tau_{1,1}) \\ \vdots & & \\ \varphi_1(\tau_{n,r_n}) & \dots & \varphi_{m_n}(\tau_{n,r_n}) \end{pmatrix}$$

A partir dessa matriz, o cálculo de $\mathcal{L}_n\left(\theta_{-m}, \omega\right)$ para cada vetor θ_{-m} no processo iterativo pode ser feito através de operações matriciais entre a matriz A , o vetor θ_{-m_n} e o vetor $B_{m_n} = (b_k)$,

$k=1, \dots, m_n$, onde

$$b_k = \int_0^1 \varphi_k(t) dt. \text{ De fato,}$$

$$\mathcal{L}_n\left(\theta_{-m_n}, \omega\right) = A \theta_{-m_n}^{(r)} - n + n \theta_{-m_n}^{(r)t} B_{m_n}$$

A utilização deste procedimento facilita, também, o cálculo do vetor de primeiras derivadas e da matriz Hessiana, necessários na aplicação do método de Newton-Raphson, e que são dados pelas expressões abaixo.

$$\frac{\partial \mathcal{L}_n(\underline{\theta}_{m_n}, \underline{\omega})}{\partial \theta_{m_n, u}} = \sum_{l=1}^n \sum_{j=1}^{r_l} \left(\frac{\varphi_u(\tau_{l,j})}{\sum_{k=1}^{m_n} \theta_k \varphi_k(\tau_{l,j})} \right)$$

$$\frac{\partial^2 \mathcal{L}_n(\underline{\theta}_{m_n}, \underline{\omega})}{\partial \theta_{m_n, u} \partial \theta_{m_n, v}} = - \sum_{l=1}^n \sum_{j=1}^{r_l} \left(\frac{\varphi_u(\tau_{l,j}) \varphi_v(\tau_{l,j})}{\left(\sum_{k=1}^{m_n} \theta_k \varphi_k(\tau_{l,j}) \right)^2} \right)$$

Isto pode ser observado nos programas em anexo (Apêndice II), elaborados utilizando-se o módulo de cálculo matricial **CM** do pacote estatístico **SOC** (ver NTIA (1989)).

As funções intensidade simuladas foram as seguintes:

- 1) $\lambda_1(t) = 2e^{2t}$, $t \in [0,1]$;
- 2) $\lambda_2(t) = 15e^{-2t}$, $t \in [0,1]$;
- 3) $\lambda_3(t) = \begin{cases} 7, & \text{se } t \in [0, 1/2) \\ 5, & \text{se } t \in [1/2, 1] \end{cases}$;
- 4) $\lambda_4(t) = 2 + 8t$, $t \in [0,1]$;
- 5) $\lambda_5(t) = \begin{cases} 2, & \text{se } t \in [0, 1/4) \\ 16t - 2, & \text{se } t \in [1/4, 3/4) \\ 10, & \text{se } t \in [3/4, 1] \end{cases}$;
- 6) $\lambda_6(t) = 6 + \sin(4\pi t)$, $t \in [0,1]$.

A definição da dimensão dos subconjuntos S_m de cada "sieve" foi feita tomando $m_n = [n^\epsilon]$, para $\epsilon = 1/2, 1/3$ e $1/4$. A escolha destes valores baseou-se em sugestões da literatura (MacKeague (1986) e Karr(1987)), já que os resultados de Geman e Hwang (1982) para o cálculo de m_n , segundo os próprios autores, são de difícil aplicação.

Por problemas de limitação da capacidade computacional, os tamanhos de amostra para os quais realizou-se a simulação variaram de acordo com o valor de ϵ . Para $\epsilon = 1/2$, fizemos $n = 10, 25$ e 50 ; para $\epsilon = 1/3$, fizemos $n = 10, 25, 50$ e 100 ; e finalmente, para $\epsilon = 1/4$, utilizamos $n = 25, 50, 100$ e 200 .

4.3 - Análise dos resultados.

Os gráficos (anexo II) mostram, simultaneamente, as funções intensidade e aquelas obtidas utilizando-se cada um dos "sieves", para alguns tamanhos de amostra e cada valor de ϵ .

Os gráficos mostram que, para o "sieve" de Fourier, as estimativas da função intensidade real são mais confiáveis para valores de t distantes dos extremos 0 e 1 . Isto ocorre porque, para $\lambda_m \in S_m$, temos $\lambda_m(0) = \lambda_m(1)$, e, obviamente, a função λ_0 real pode ser tal que $\lambda_0(0) \neq \lambda_0(1)$. No entanto, nota-se que, para valores não muito próximos de 0 e 1 , a estimativa obtida aproxima razoavelmente a função real, principalmente para dimensões de S_m maiores que 2 e intensidades

contínuas. Quando a intensidade é constante em intervalos de uma partição de $[0,1]$, o "sieve" de Fourier parece não fornecer bons resultados, provavelmente porque seria necessário uma dimensão maior para que a aproximação fosse razoável.

Quanto ao "sieve" de histograma, as estimativas obtidas mostram que este tipo de "sieve" fornece bons resultados para funções constantes em subintervalos de $[0,1]$, desde que cada elemento da partição de $[0,1]$ onde a função é constante seja igual à união de alguns intervalos $T_{m_s} = [(s-1)/m, s/m)$. Caso contrário, o "sieve" de histograma fornece apenas uma visualização da tendência da intensidade real, o mesmo ocorrendo para funções que não sejam constantes em subintervalos de $[0,1]$.

4.4 - Sugestões

Além de um estudo mais detalhado dos "sieves" de histograma e de Fourier (o que compreenderia a realização de simulações para tamanhos de amostra maiores em computadores de grande porte e para um número maior de funções intensidade) poder-se-ia estudar o comportamento de "sieves" formados por outras bases de $L_2([0,1], dt)$, como aquelas formadas pelas funções de Haar (ver Paley (1932) e Walsh (1923)), de Walsh (ver Walsh (1923) e Fine (1949)).

Outra possibilidade é o estudo de "sieves" que não sejam

subconjuntos de subespaços vetoriais gerados por subconjuntos de bases do espaço $L_2([0,1],dt)$.

A observação dos gráficos das estimativas obtidas pelo "sieve" de histograma sugere, ainda, que se estude o polígono de frequências obtido a partir do histograma como estimador da função intensidade.

Apêndice I

A.1 - Se f é uma função mensurável limitada em $[0,1]$, então

$$\lim_{p \rightarrow \infty} \|f\|_p = \|f\|_\infty = \text{ess sup } f(t) = \inf \left\{ M : \text{Leb}\{ t : f(t) > M \} = 0 \right\}.$$

A.2 - Uma função f é semicontínua superior se, para $\forall y$, $f(y) \neq \infty$ e

$$f(y) \geq \overline{\lim}_{x \rightarrow y} f(x) = \limsup_{x \rightarrow y} f(x).$$

A.3 - Seja $M(t)$ uma $(P, \mathcal{F}_t)$ -martingale e $\phi : \mathbb{R} \rightarrow \mathbb{R}$ uma função convexa

tal que, para $\forall t \geq 0$, $E \left[|\phi(M(t))| \right] < \infty$. Então, $\phi(M(t))$ é uma

$(P, \mathcal{F}_t)$ -martingale, pois $E \left[\phi(M(t)) \mid \mathcal{F}_s \right] \geq \phi \left(E \left[M(t) \mid \mathcal{F}_s \right] \right) = \phi(M(s))$.

A.4 - Uma $(P, \mathcal{F}_t)$ -martingale é de quadrado integrável se for tal que

$$\sup_{t \geq 0} E \left[\left(M(t) \right)^2 \right] < \infty.$$

A.5 - Um processo pontual N é um processo de Poisson não homogêneo se e somente se seu compensador é determinístico.

A.6 - Assumindo que $E(N(1)) < \infty$, existe uma martingale de quadrado

integrável $\left\{ M(t), \mathcal{F}_t, t \geq 0 \right\}$ tal que $N(t) = \int_0^t \lambda(s)ds + M(t)$,

$t \in [0,1]$. $\langle M(t) \rangle = \int_0^t \lambda(s) ds$ é o compensador de M^2 e é chamado de "variação previsível" de M .

A.7 - Propriedades de tempos de Markov:

- 1) qualquer número a em $[0, \infty]$ é um tempo de Markov para qualquer filtragem F ;
- 2) se T é um tempo de Markov, e $a \in [0, \infty]$, então $T + a$ é um tempo de Markov;
- 3) se T e S são tempos de Markov, então $S \wedge T = \inf(S, T)$ e $S \vee T = \sup(S, T)$ são tempos de Markov;
- 4) se T é um tempo de Markov, então para qualquer $a \in [0, \infty)$, $T \wedge a$ é $\mathcal{F}_a$ -mensurável;
- 5) se X é um processo estocástico adaptado à $\mathcal{F}_t$, contínuo à direita ou à esquerda, e C é um número real, então T é tempo de Markov, onde $T = \begin{cases} \inf\{ t \mid X(t) \geq C \} , & \text{se } \{...\} \neq \emptyset \\ +\infty & , \text{ caso contrário.} \end{cases}$

Apêndice II

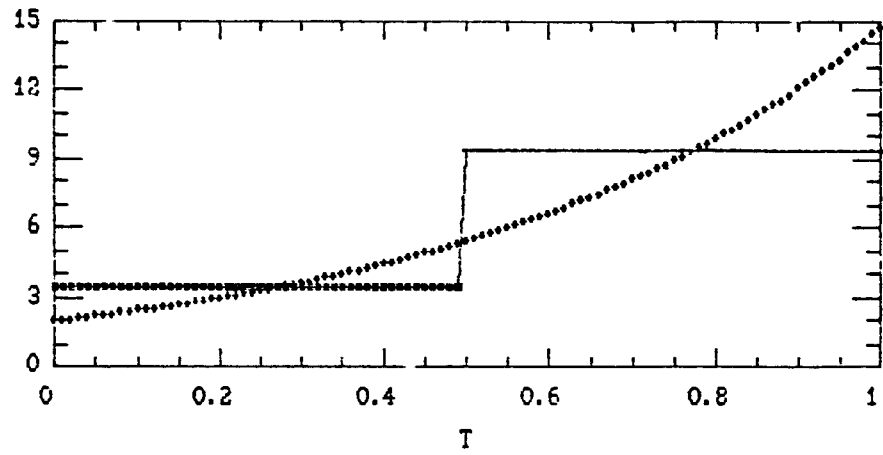
Gráficos

Gráficos para λ_1

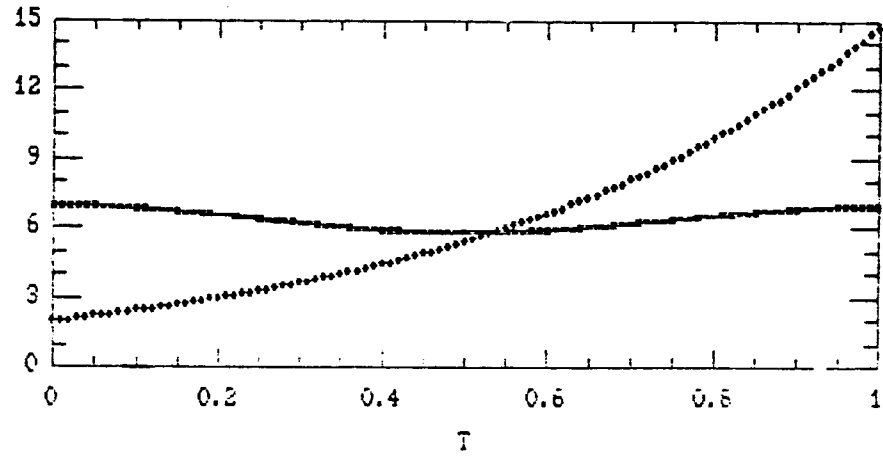
$n = 50$

$\epsilon = 1/4$

LAMBDA_1, N=50, M=N^(1/4), HISTOGRAMA $\blacksquare$ HISTOGRAMA
 $\circ$ LAMBDA_1



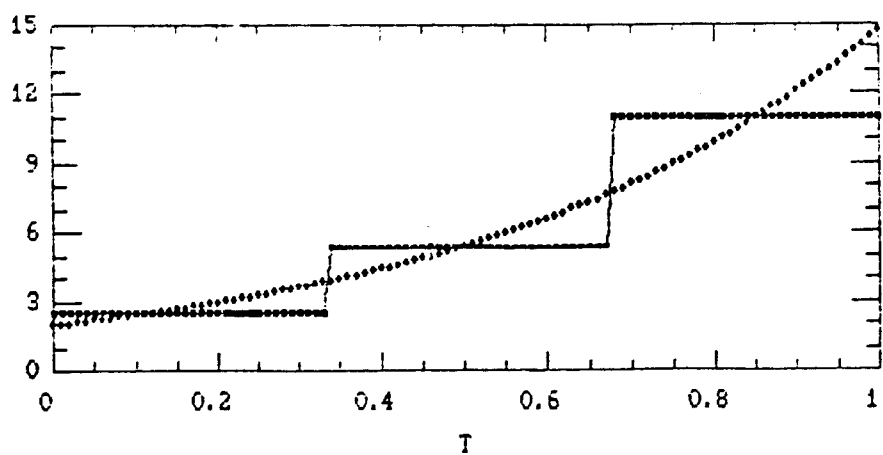
LAMBDA_1, N=50, M=N^(1/4), FOURIER $\blacksquare$ FOURIER
 $\circ$ LAMBDA_1



$n = 200$

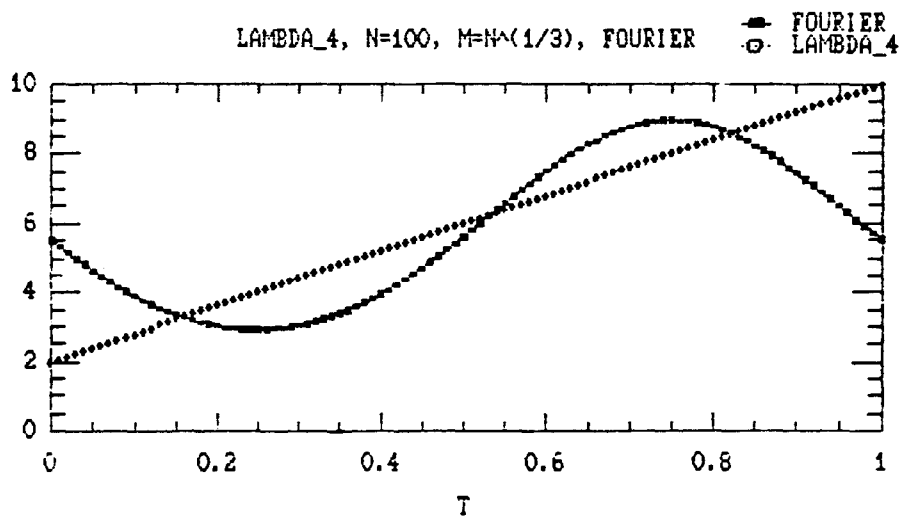
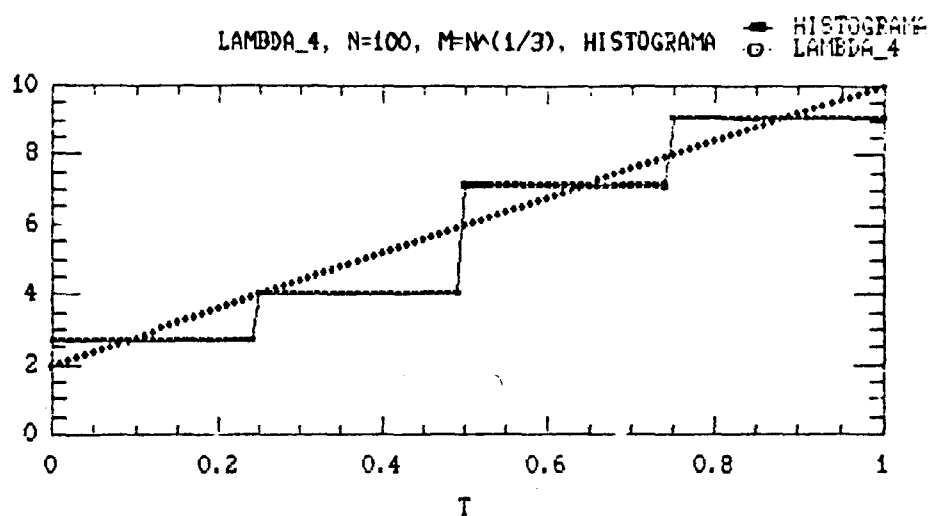
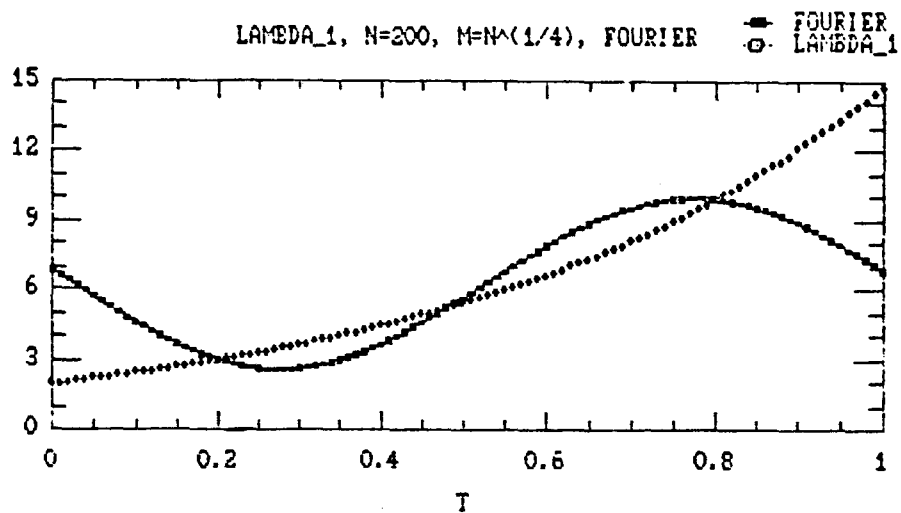
$\epsilon = 1/4$

LAMBDA_1, N=200, M=N^(1/4), HISTOGRAMA $\blacksquare$ HISTOGRAMA
 $\circ$ LAMBDA_1

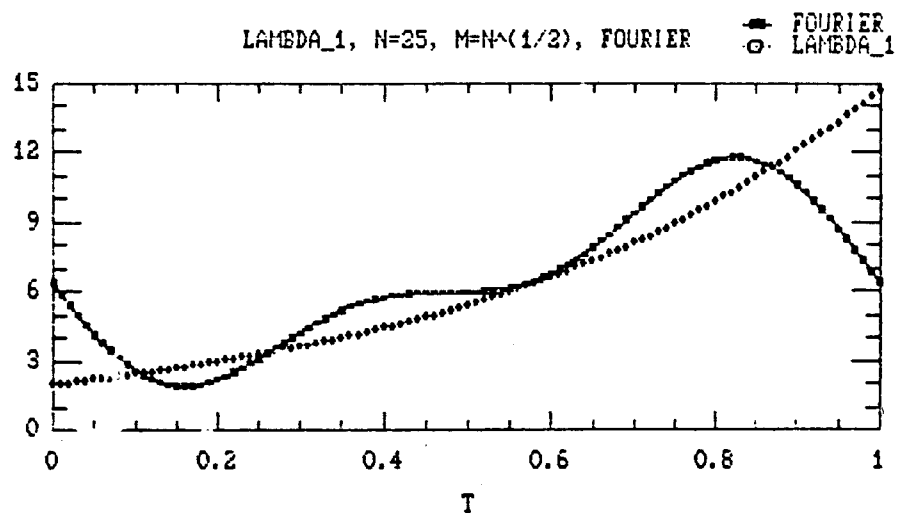
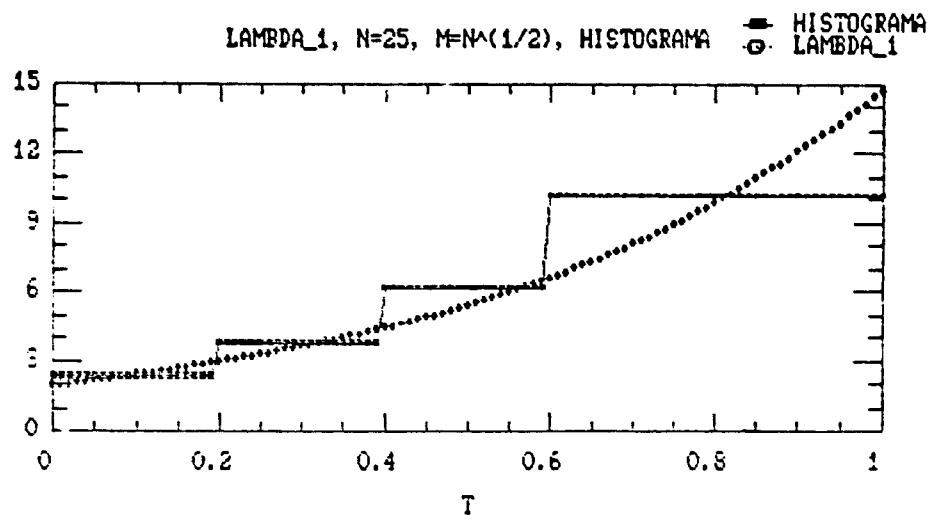


$n = 100$

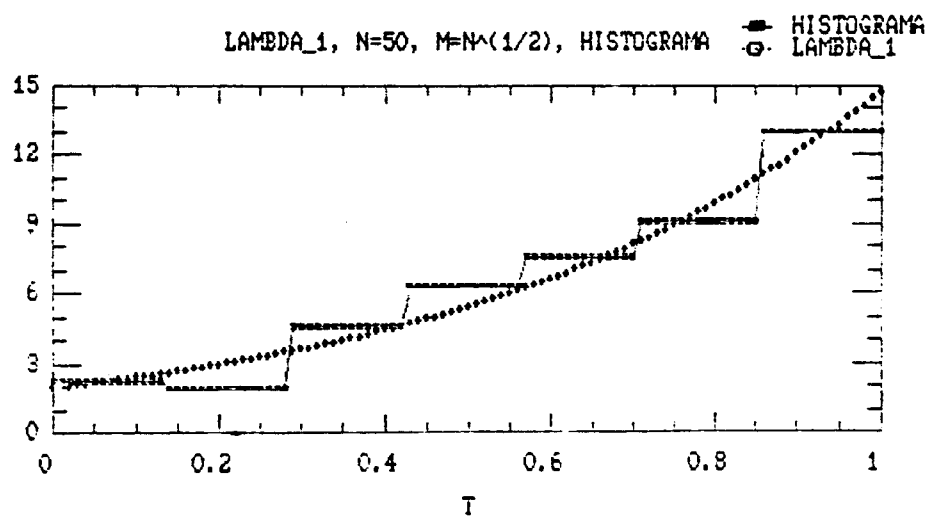
$\epsilon = 1/3$



$n = 25$
 $c = 1/2$

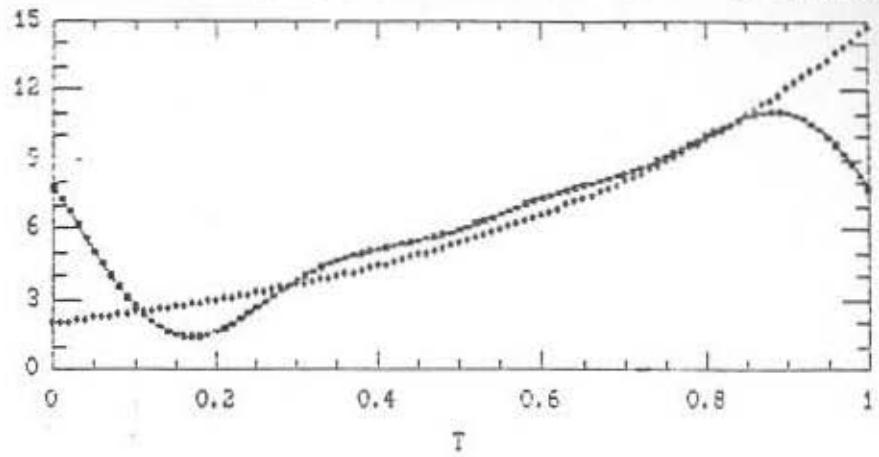


$n = 50$
 $c = 1/2$



LAMBDA_1, N=50, M=N^(1/2), FOURIER

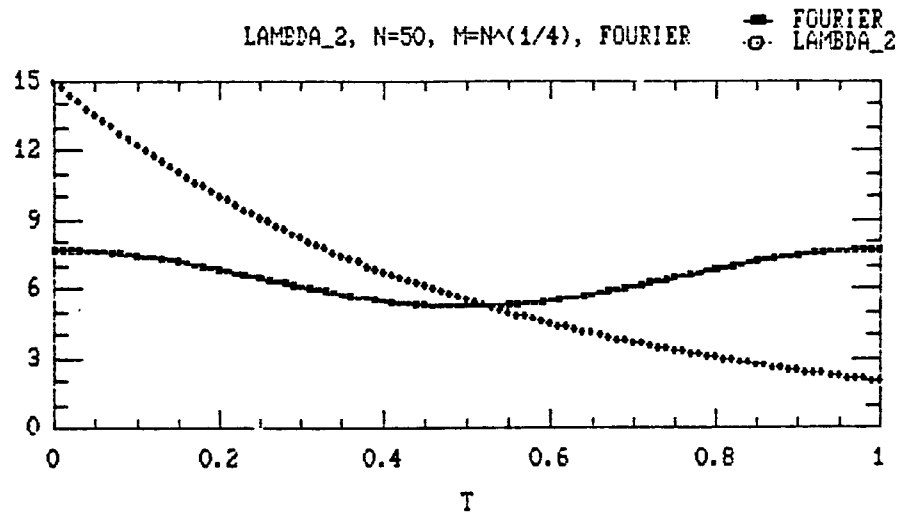
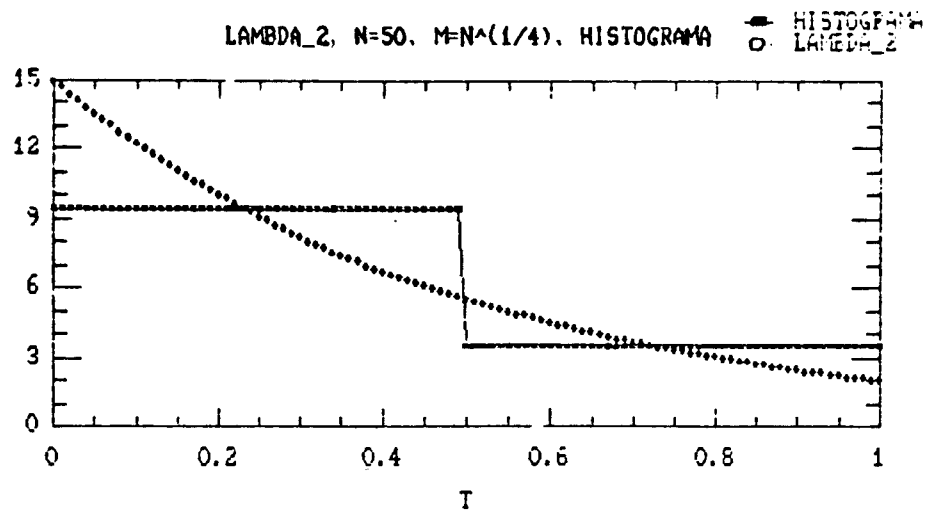
FOURIER
LAMBDA_1



Gráficos para λ_2

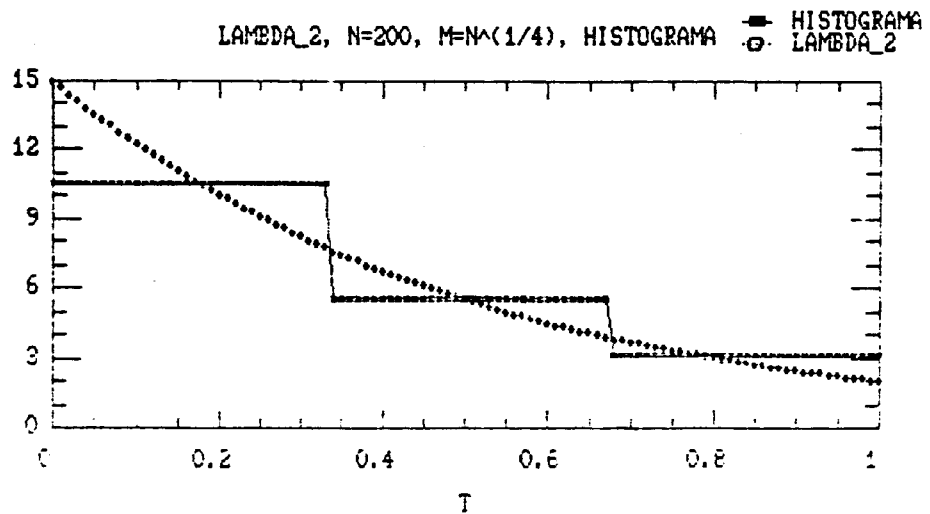
$n = 50$

$\epsilon = 1/4$



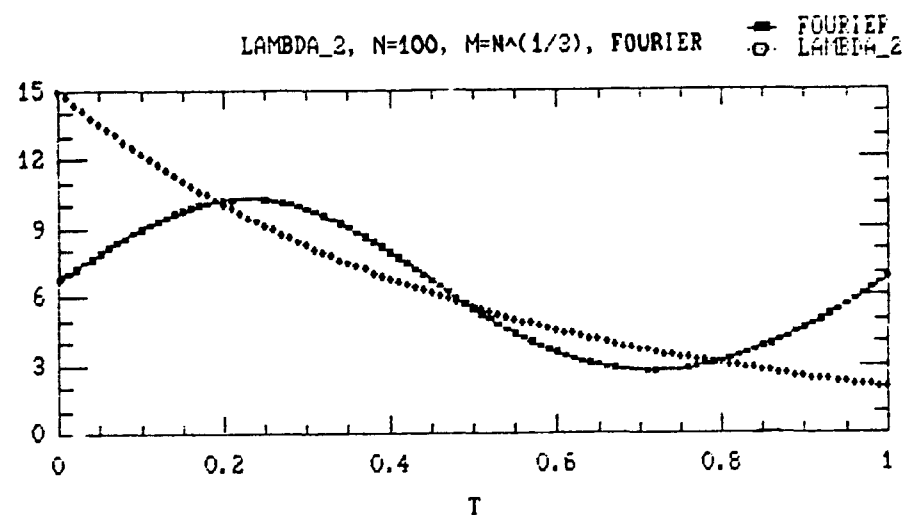
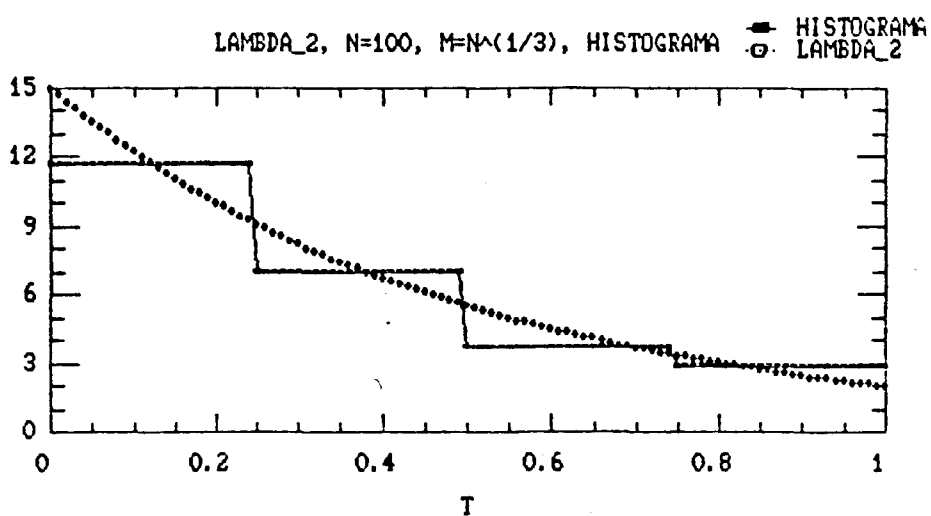
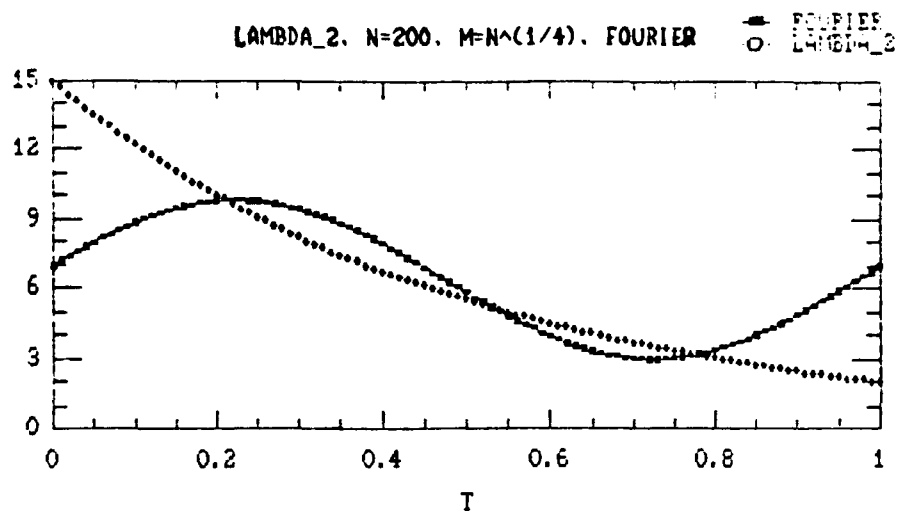
$n = 200$

$\epsilon = 1/4$

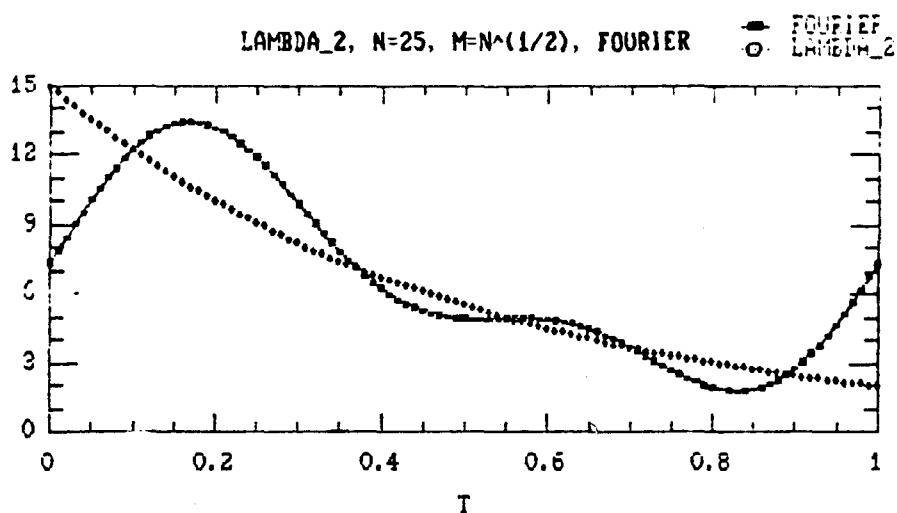
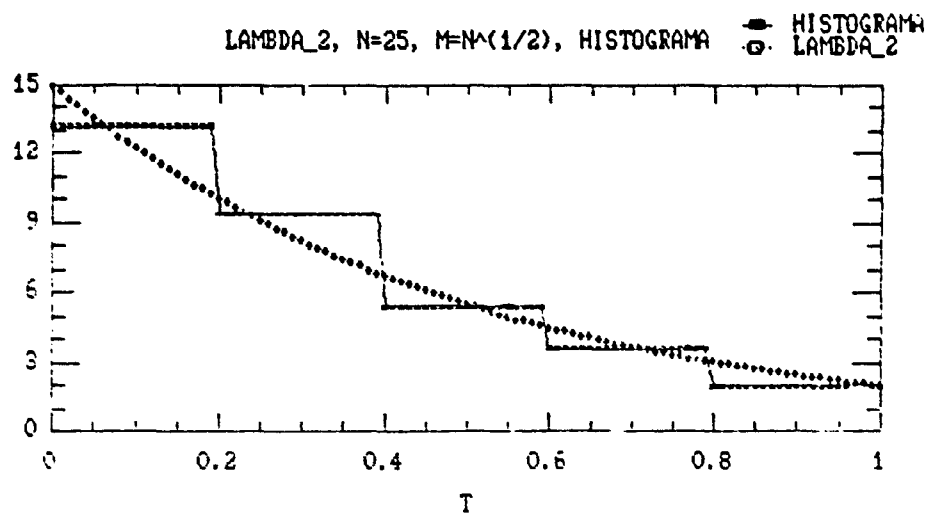


$n = 100$

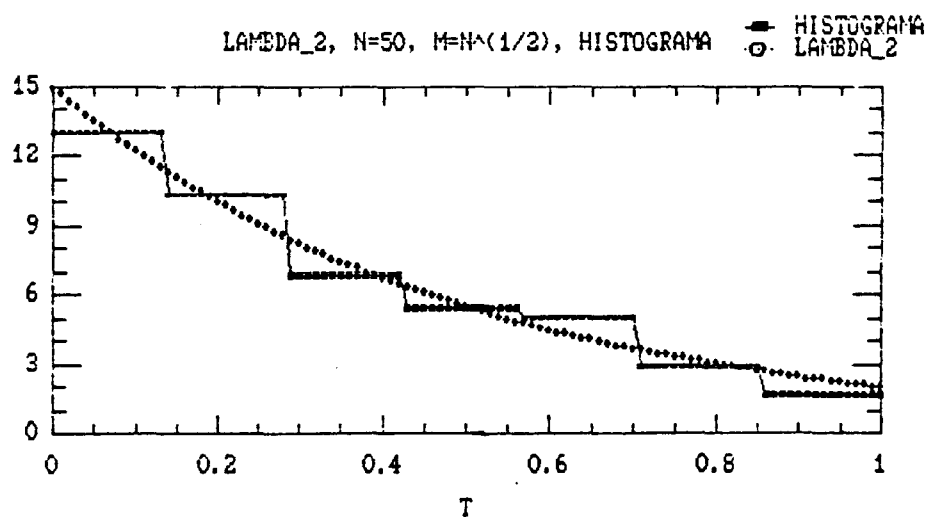
$\varepsilon = 1/3$



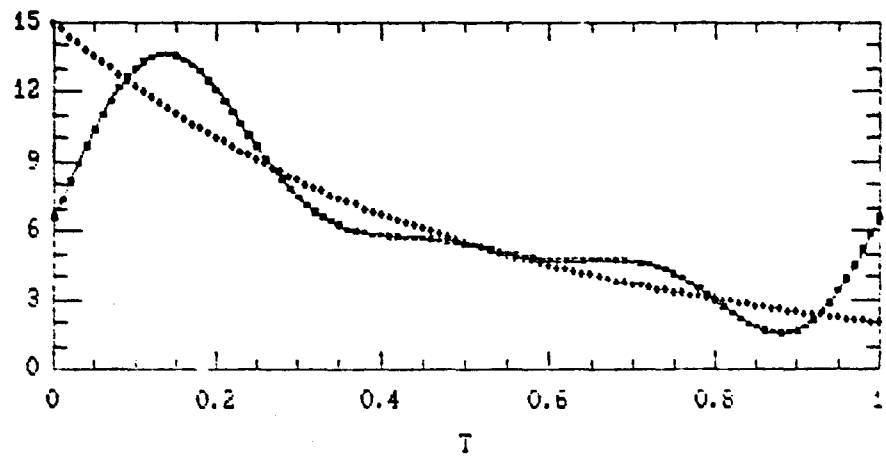
$n = 25$
 $\epsilon = 1/2$



$n = 50$
 $\epsilon = 1/2$



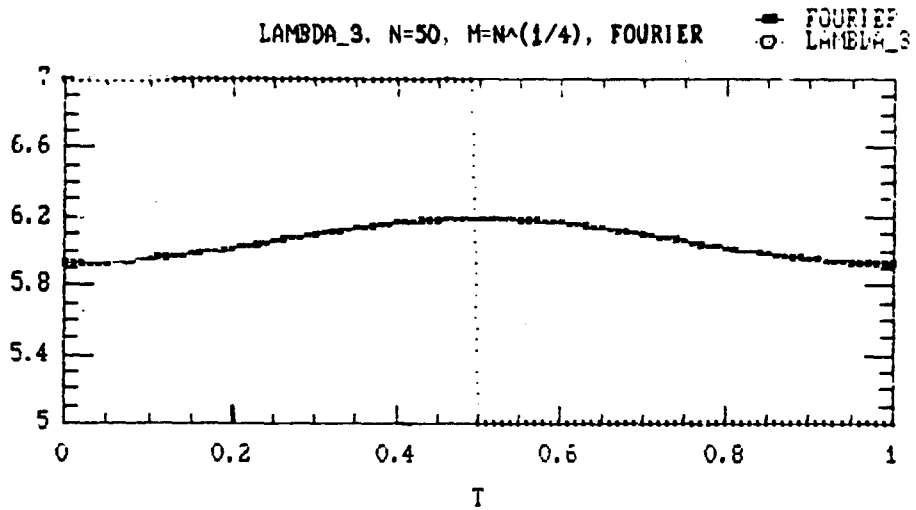
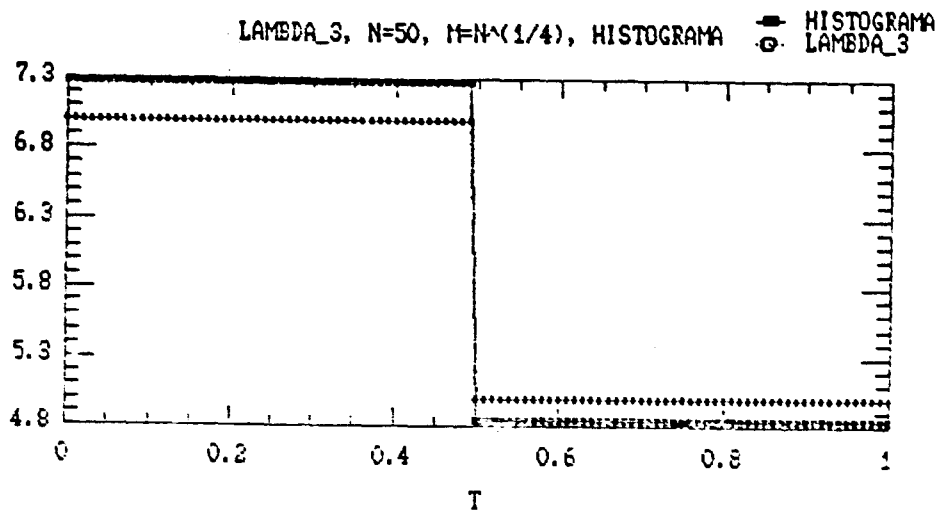
LAMBDA_2, N=50, M=N^(1/2), FOURIER — FOURIER
 ○ LAMBDA_2



Gráficos para λ_3

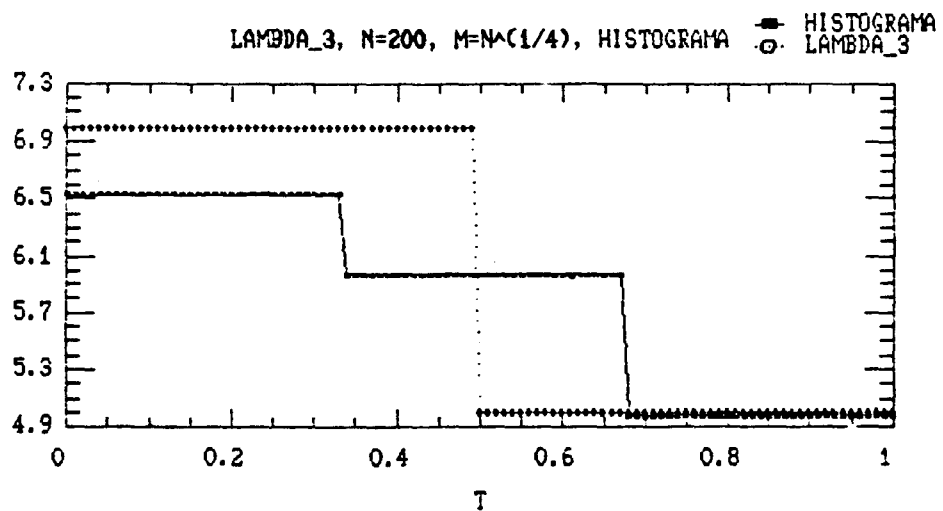
$n = 50$

$\epsilon = 1/4$

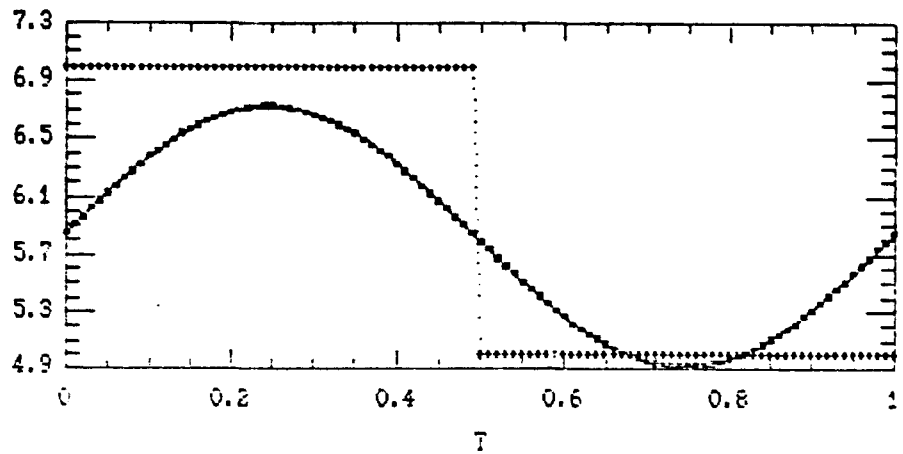


$n = 200$

$\epsilon = 1/4$



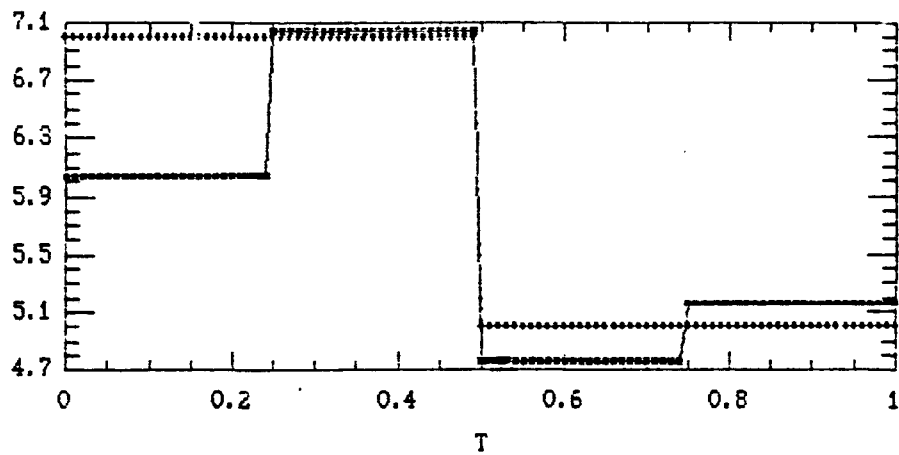
LAMBDA_3, N=200, M=N^(1/4), FOURIER



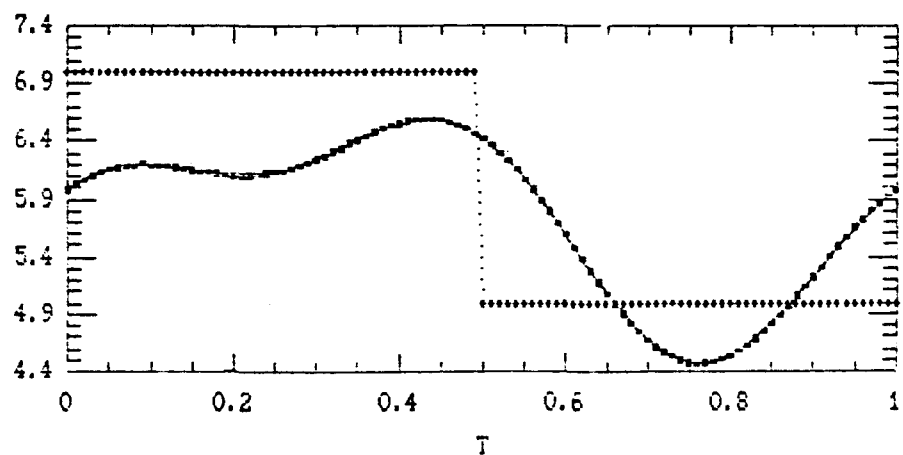
n = 100

e = 1/3

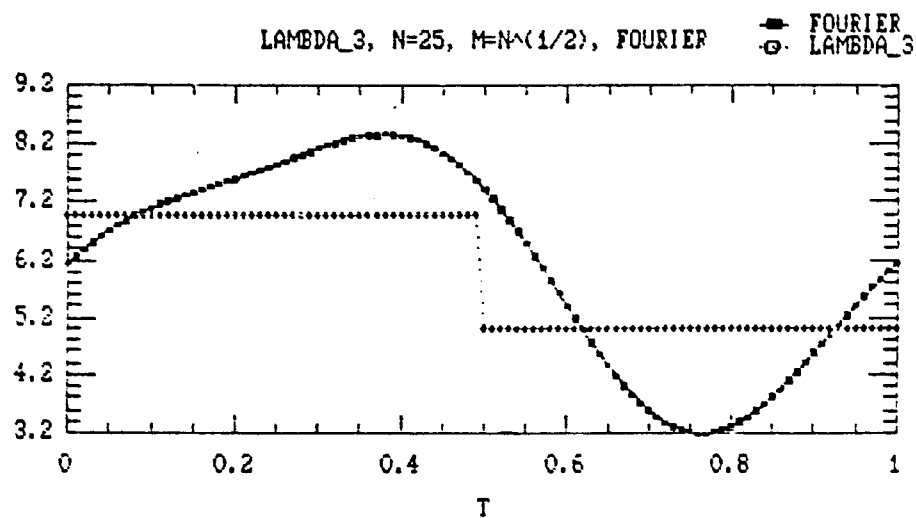
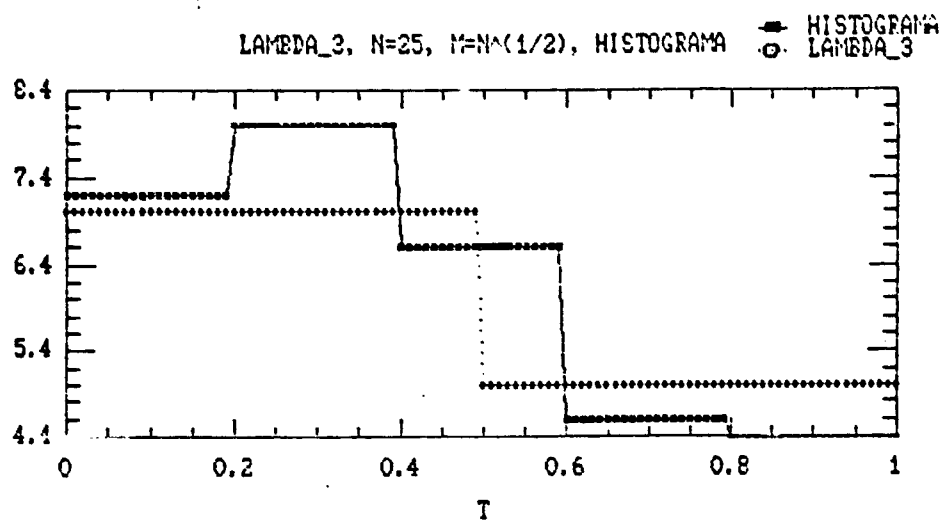
LAMBDA_3, N=100, M=N^(1/3), HISTOGRAMA



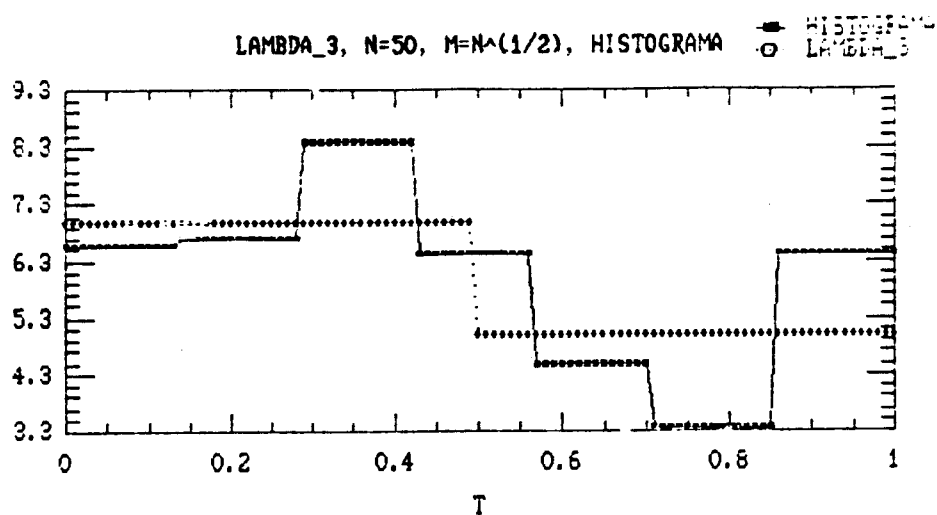
LAMBDA_3, N=100, M=N^(1/3), FOURIER

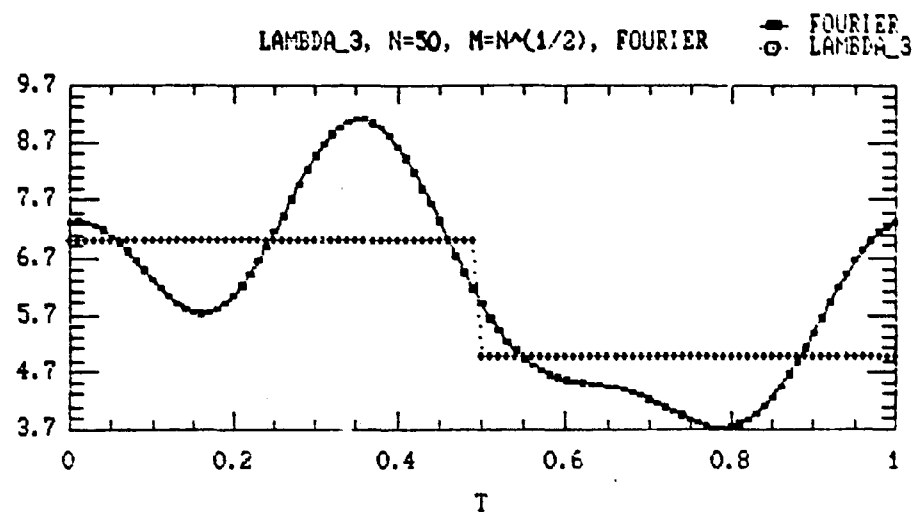


$n = 25$
 $c = 1/2$



$n = 50$
 $c = 1/2$



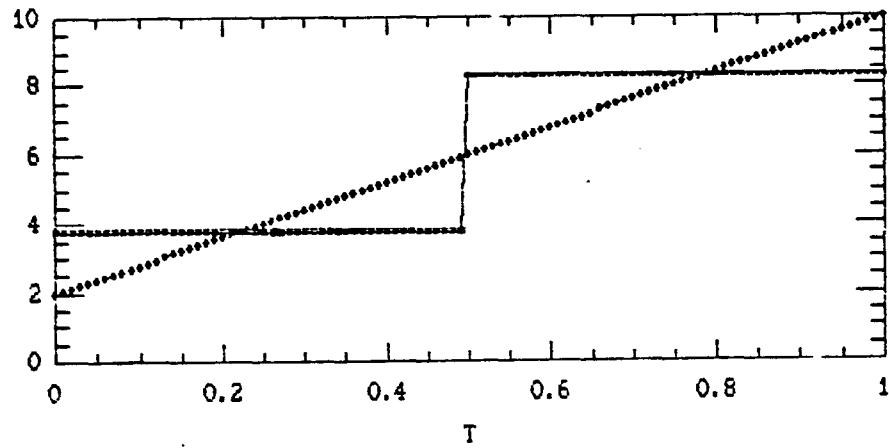


Gráficos para λ_4

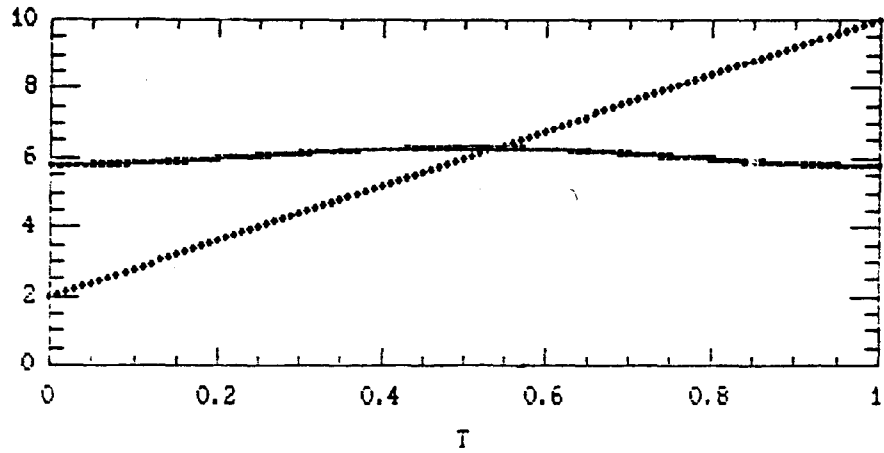
$n = 50$

$\epsilon = 1/4$

LAMBDA_4, N=50, M=N^(1/4), HISTOGRAMA $\blacksquare$ HISTOGRAMA
 $\odot$ LAMBDA_4



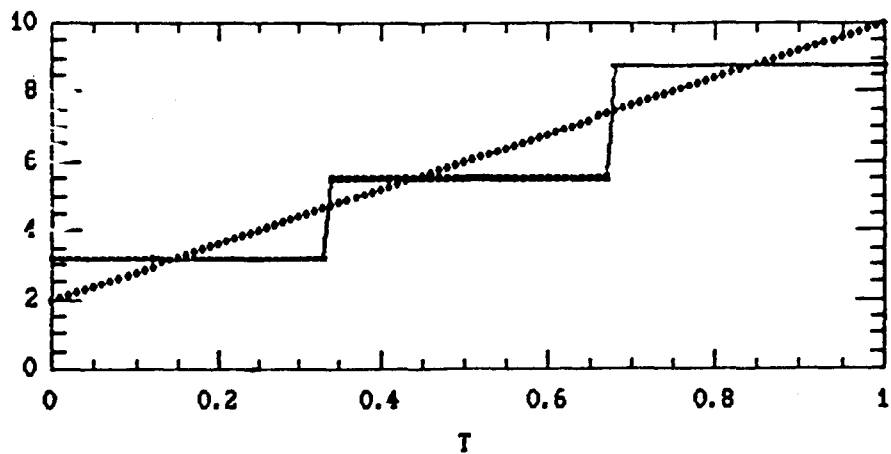
LAMBDA_4, N=50, M=N^(1/4), FOURIER $\blacksquare$ FOURIER
 $\odot$ LAMBDA_4



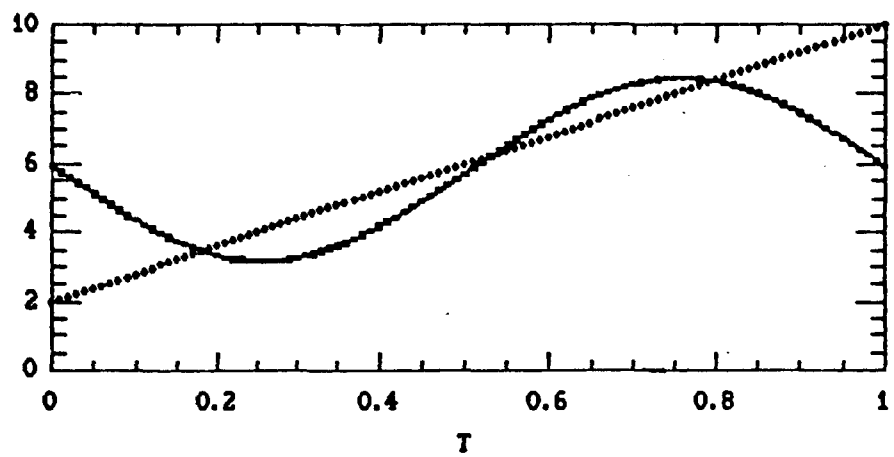
$n = 200$

$\epsilon = 1/4$

LAMBDA_4, N=200, M=N^(1/4), HISTOGRAMA $\blacksquare$ HISTOGRAMA
 $\odot$ LAMBDA_4



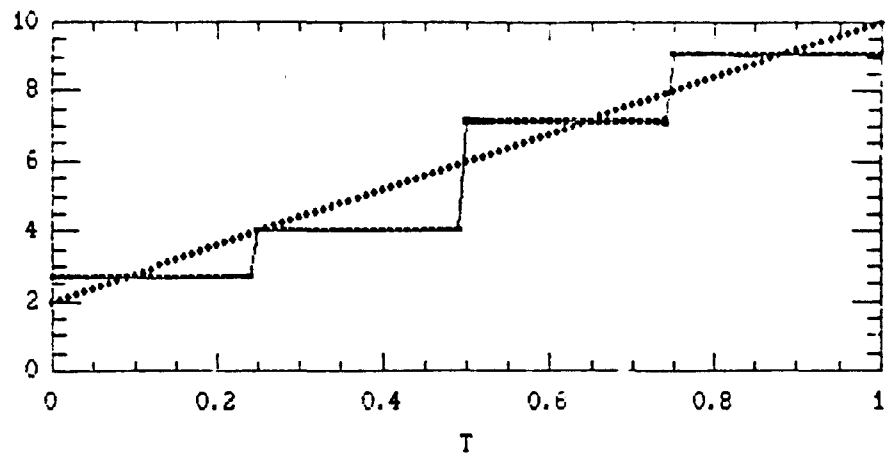
LAMBDA_4, N=200, M=N^(1/4), FOURIER



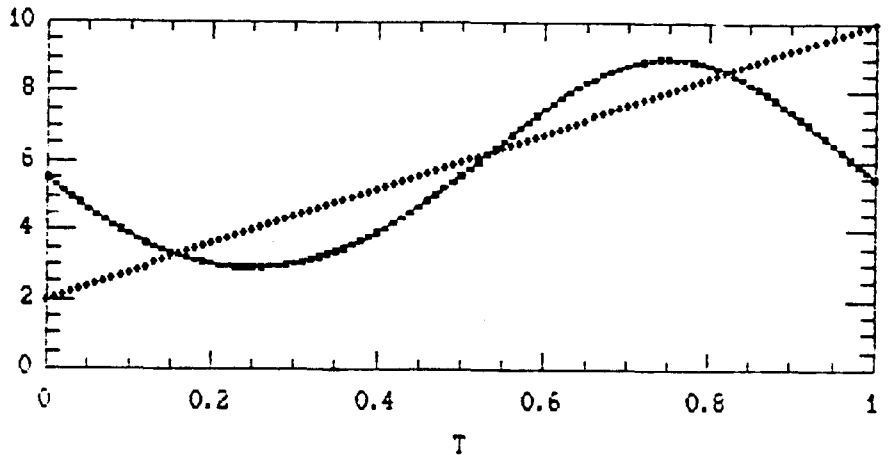
n = 100

c = 1/3

LAMBDA_4, N=100, M=N^(1/3), HISTOGRAM

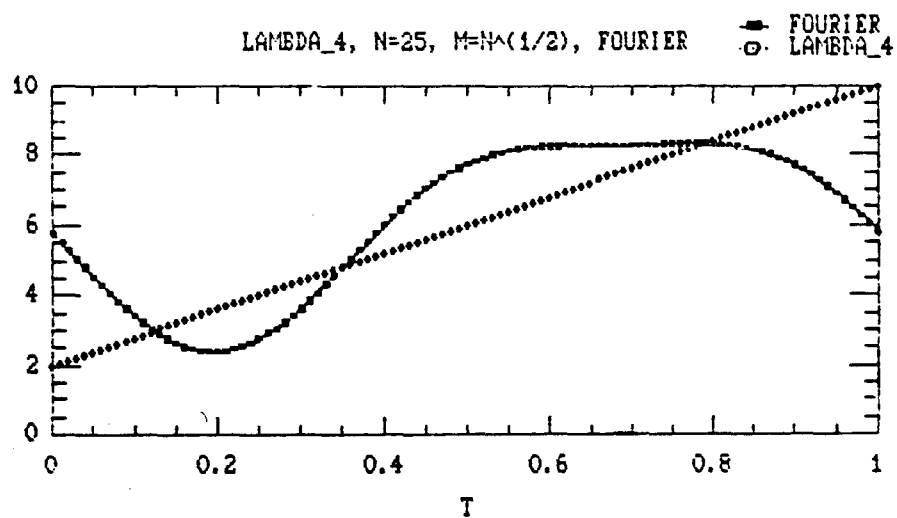
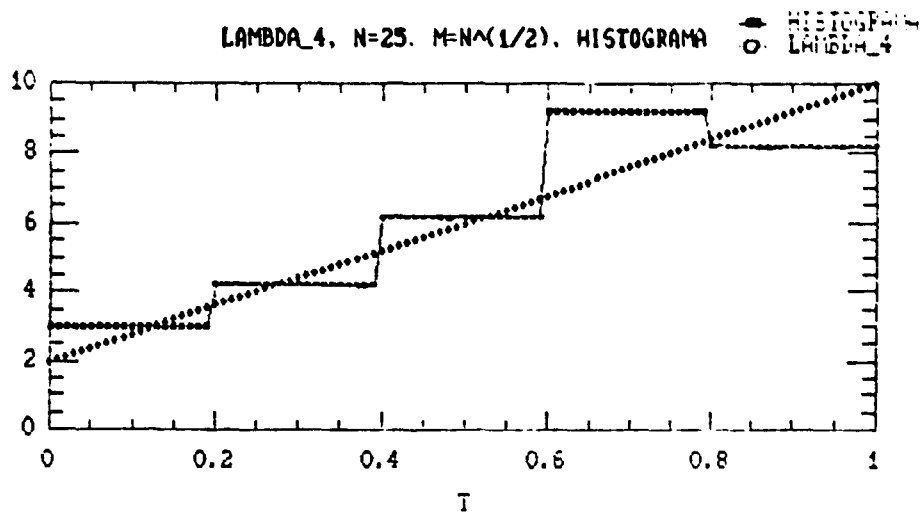


LAMBDA_4, N=100, M=N^(1/3), FOURIER



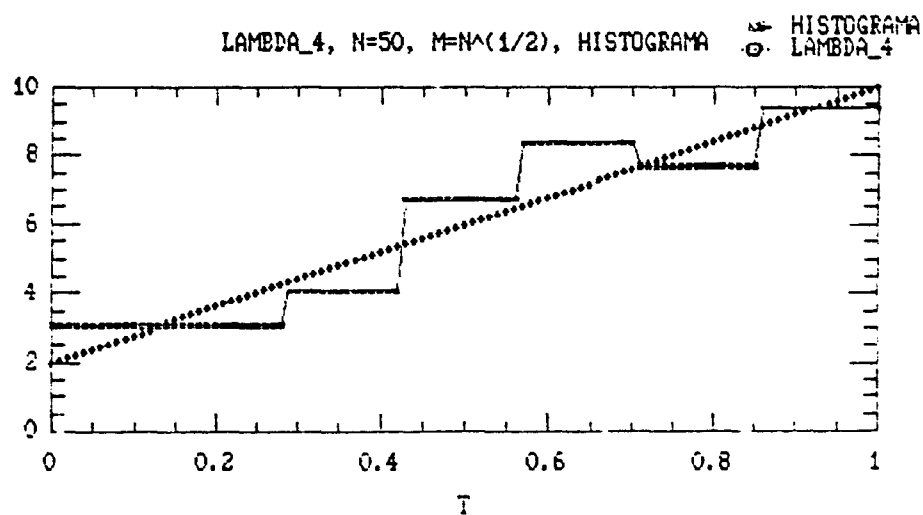
$$n = 25$$

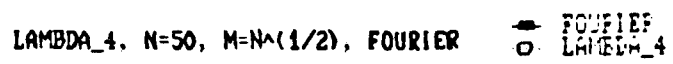
$$\epsilon = 1/2$$



$$n = 50$$

$$\epsilon = 1/2$$

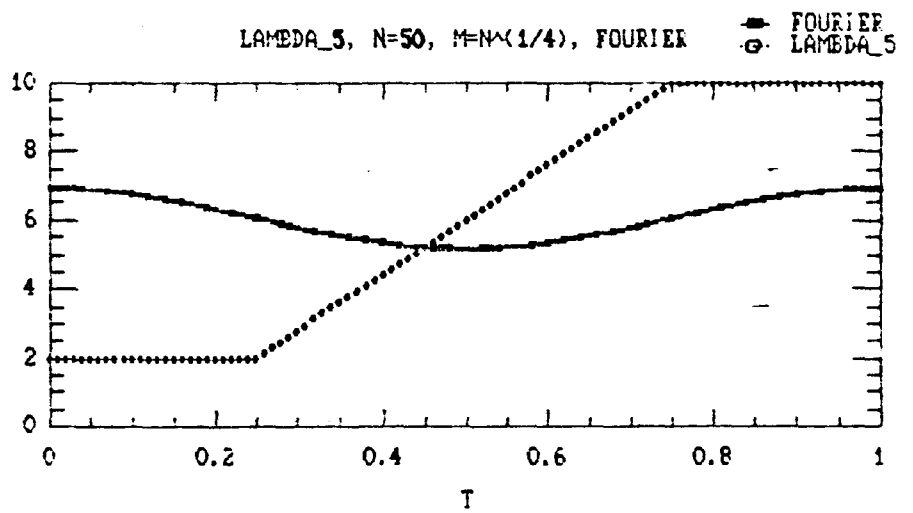
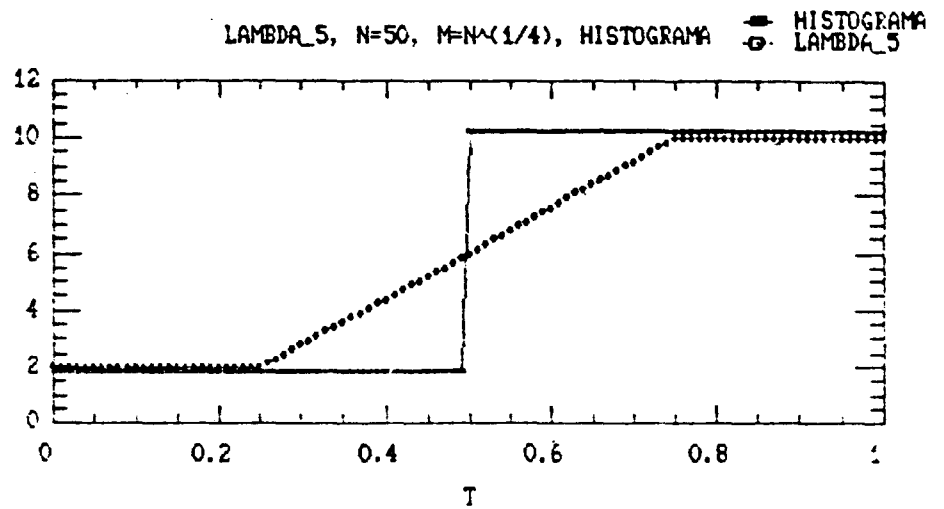




Gráficos para λ_5

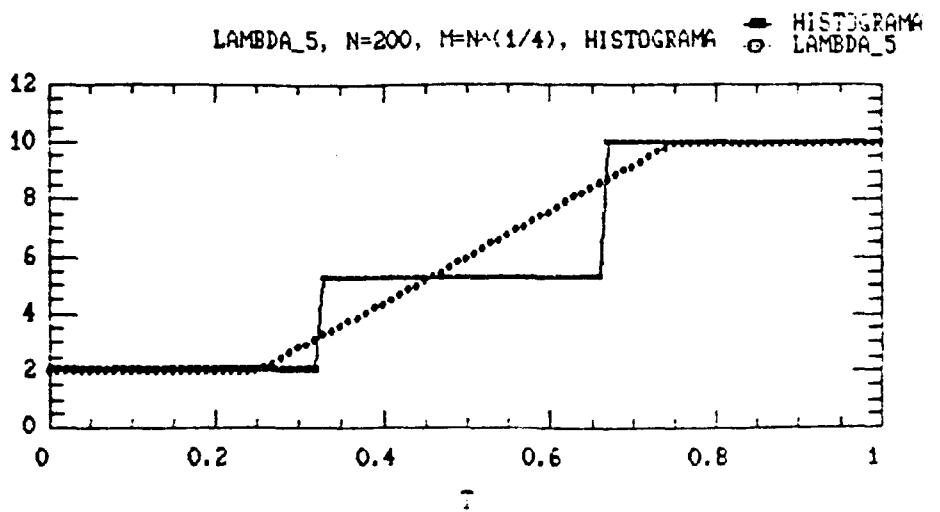
$n = 50$

$\epsilon = 1/4$



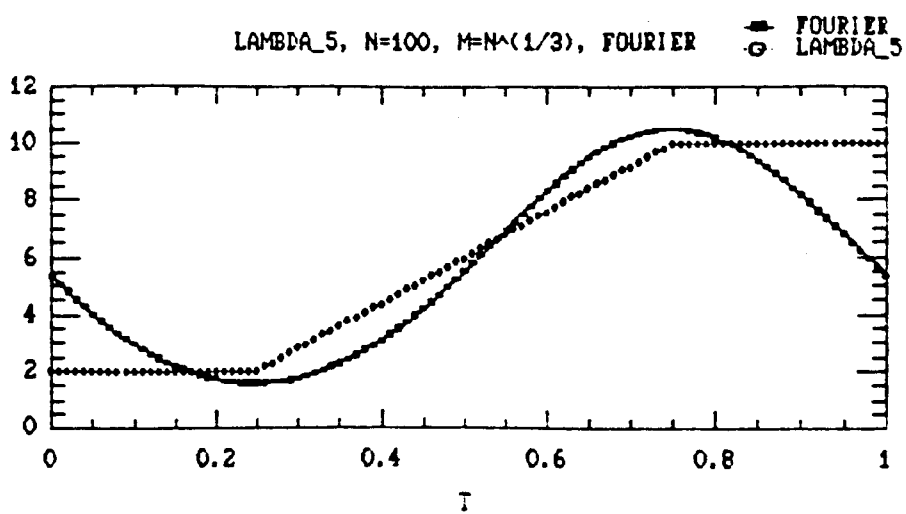
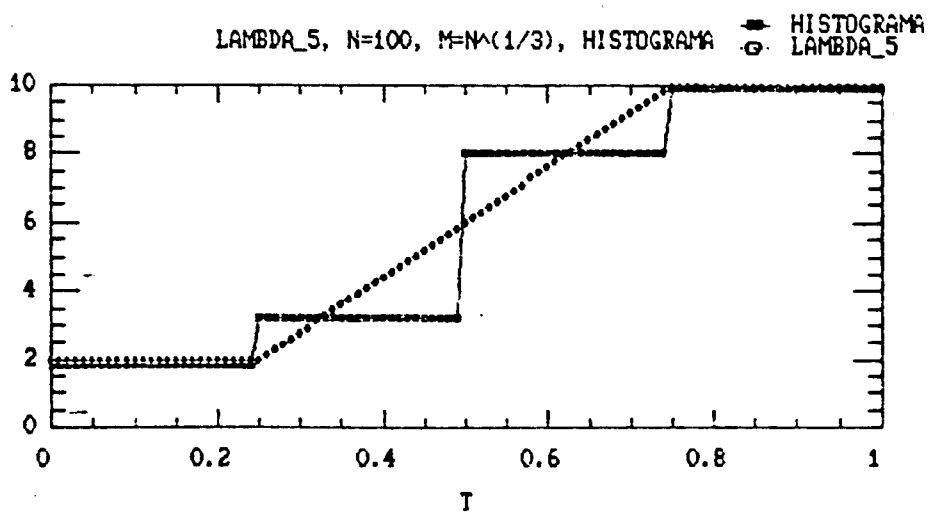
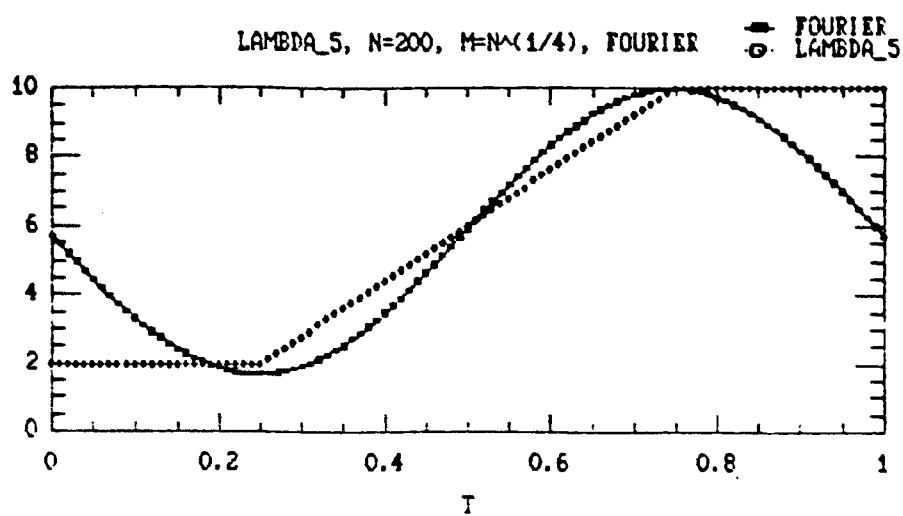
$n = 200$

$\epsilon = 1/4$

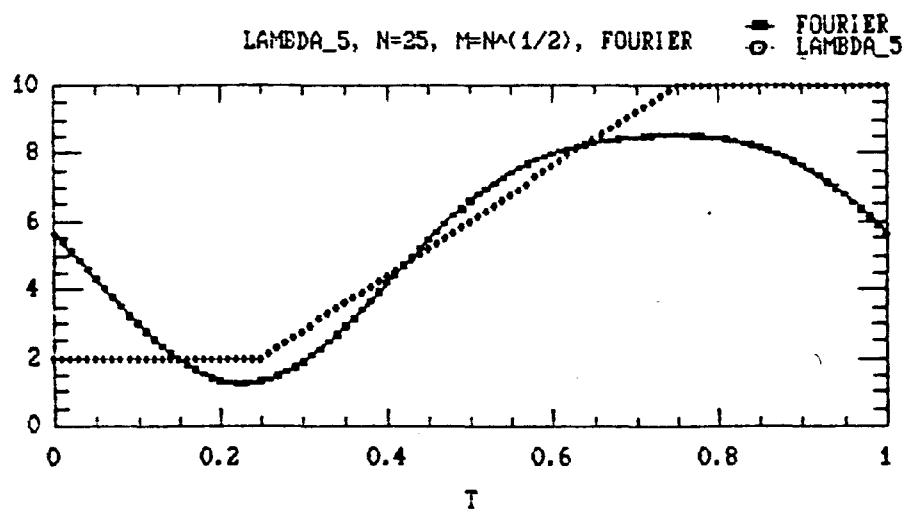
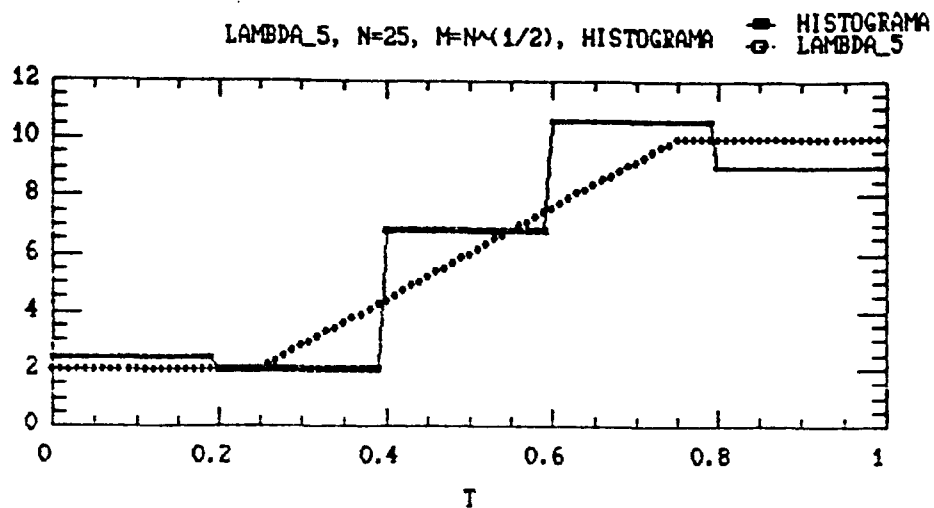


$n = 100$

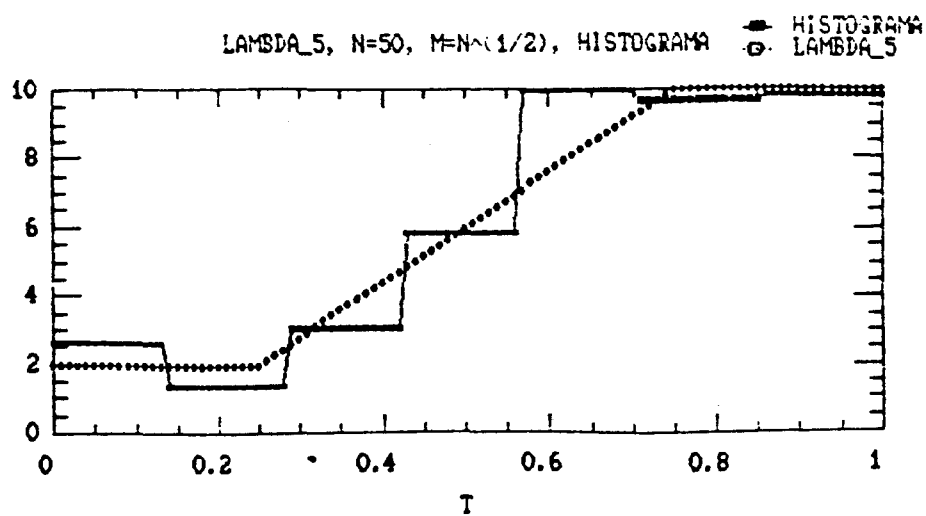
$\epsilon = 1/3$

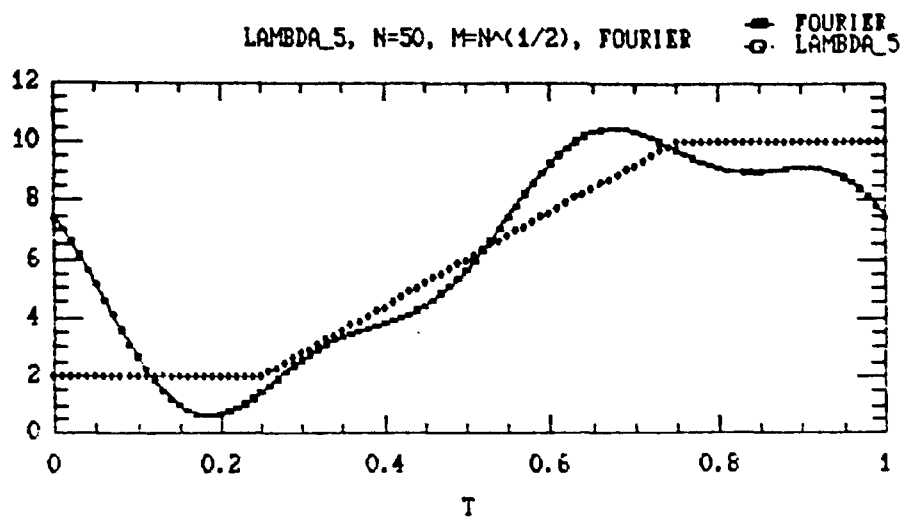


$n = 25$
 $\epsilon = 1/2$



$n = 50$
 $\epsilon = 1/2$

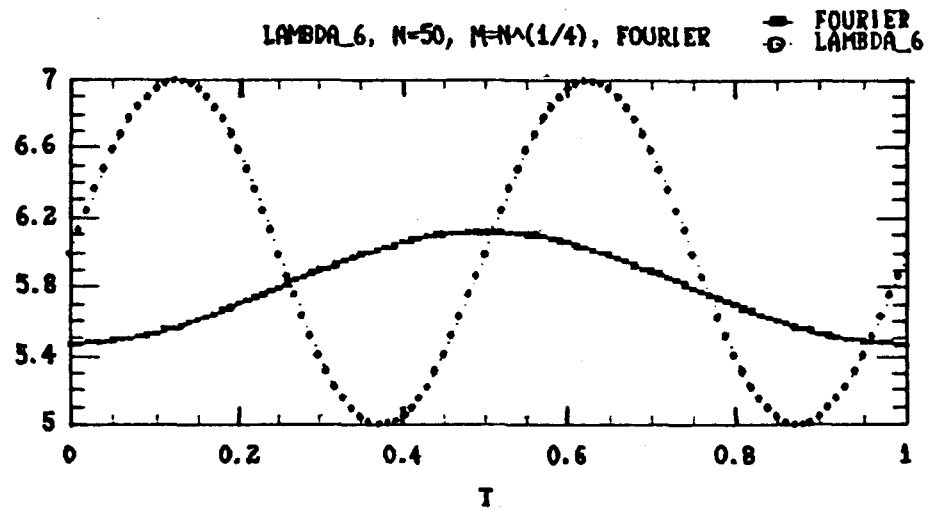
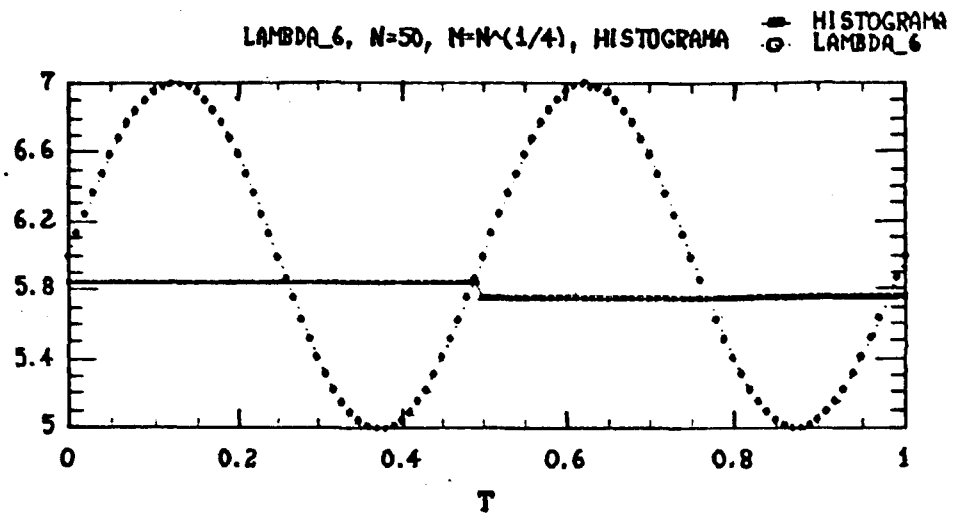




Gráficos para λ_6

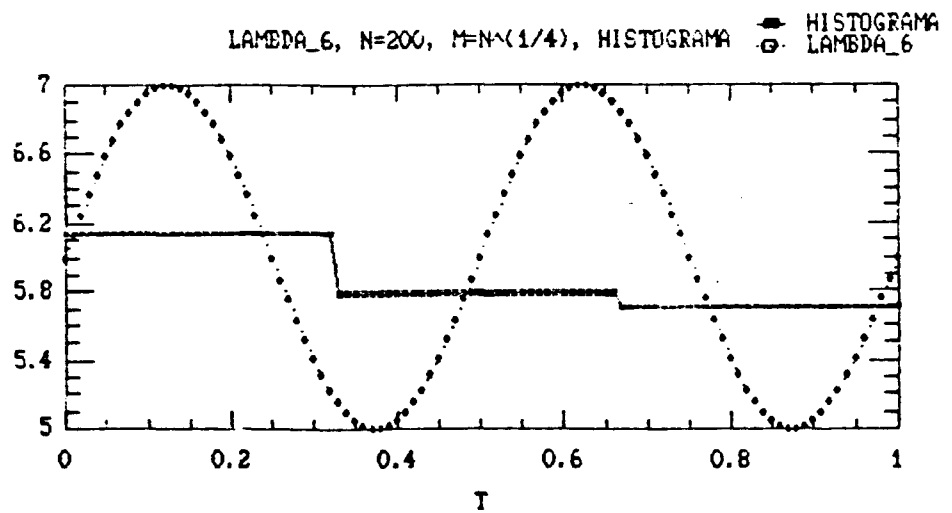
$n = 50$

$\varepsilon = 1/4$

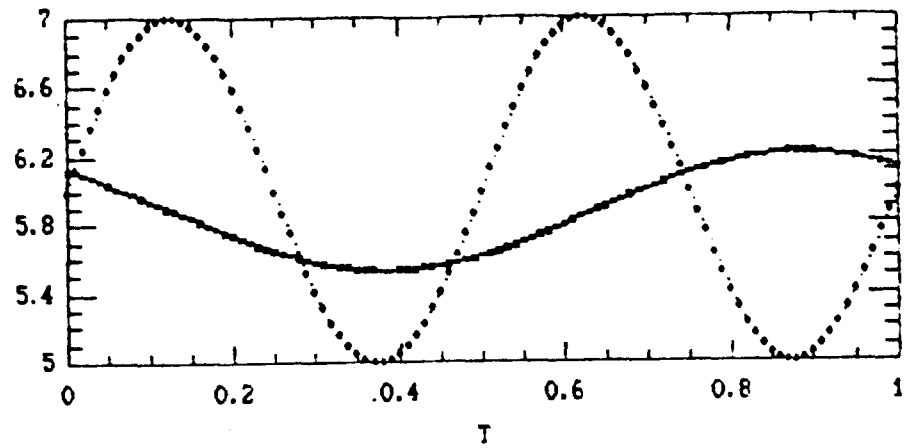


$n = 200$

$\varepsilon = 1/4$



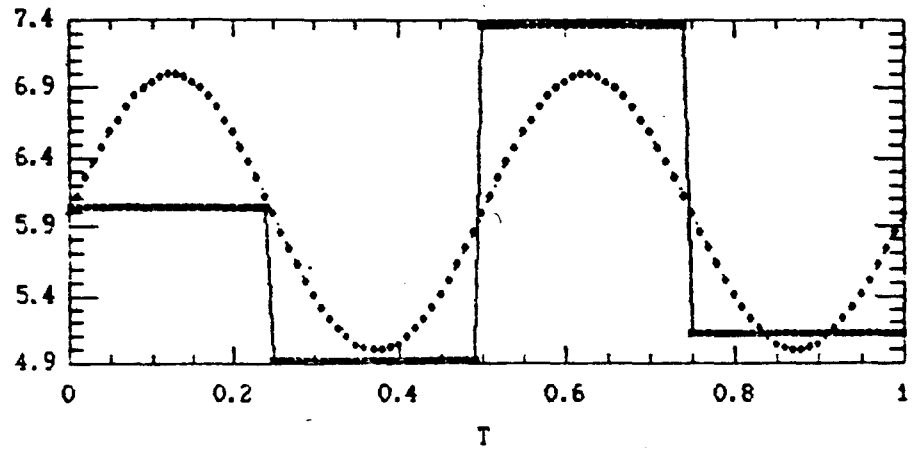
LAMBDA_6, N=200, M=N^(1/4), FOURIER



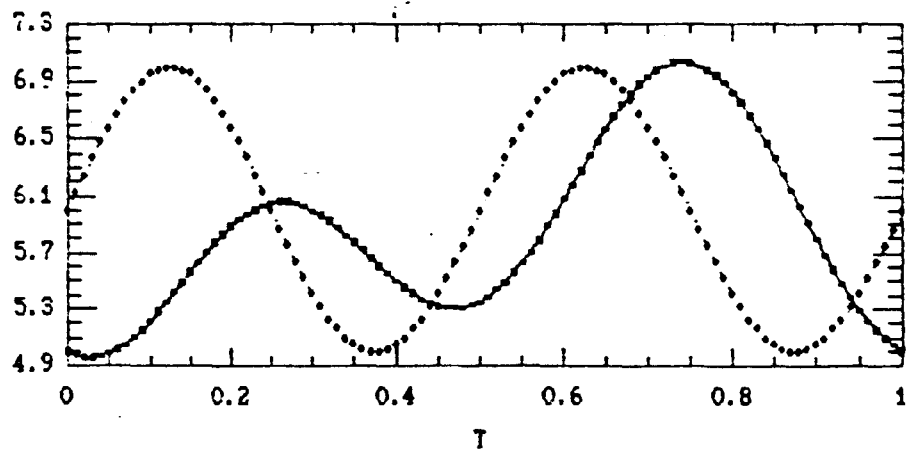
$n = 100$

$\epsilon = 1/3$

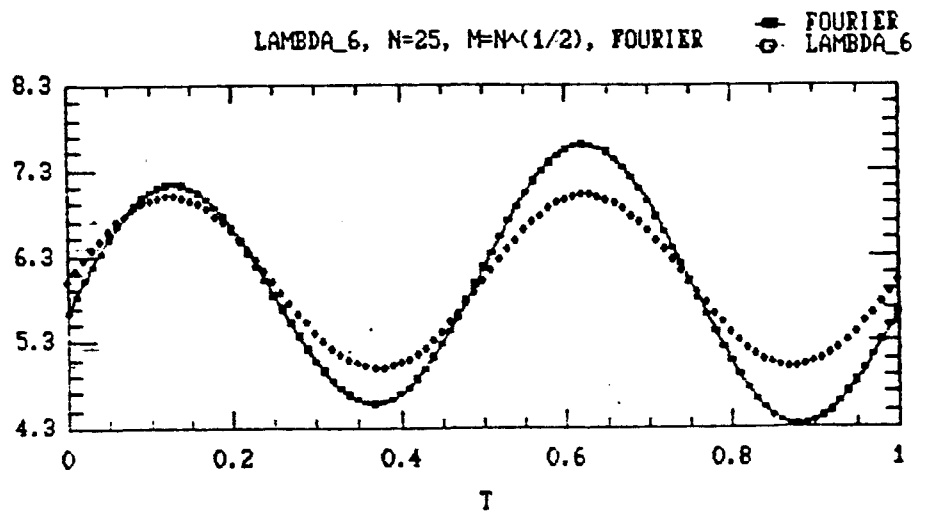
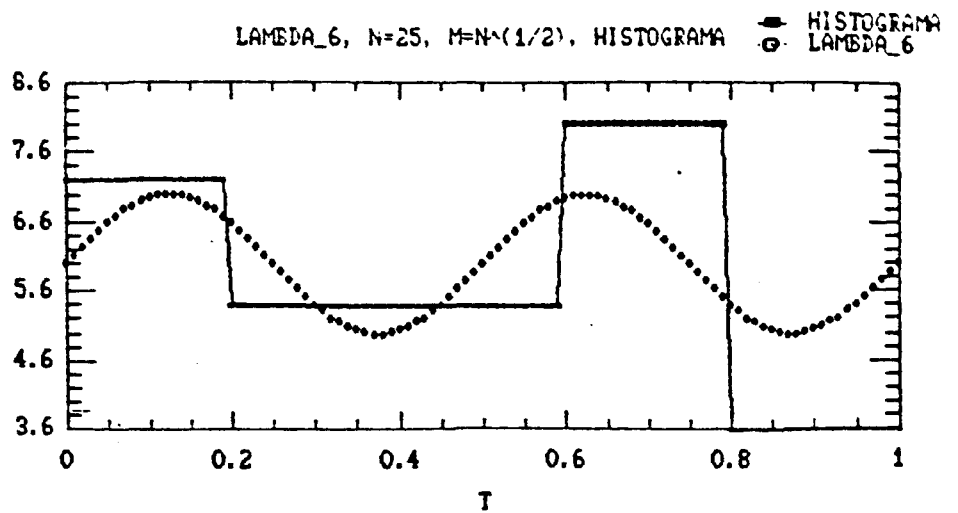
LAMBDA_6, N=100, M=N^(1/3), HISTOGRAM



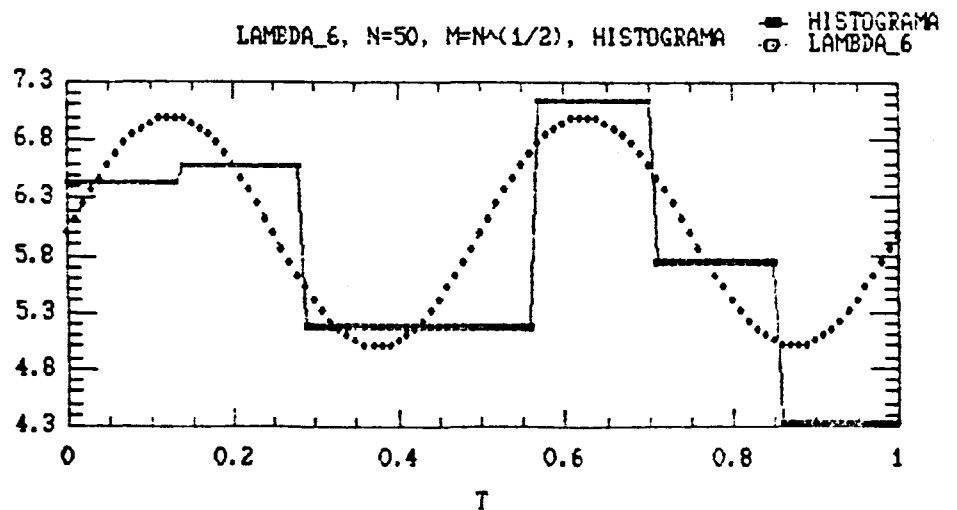
LAMBDA_6, N=100, M=N^(1/3), FOURIER

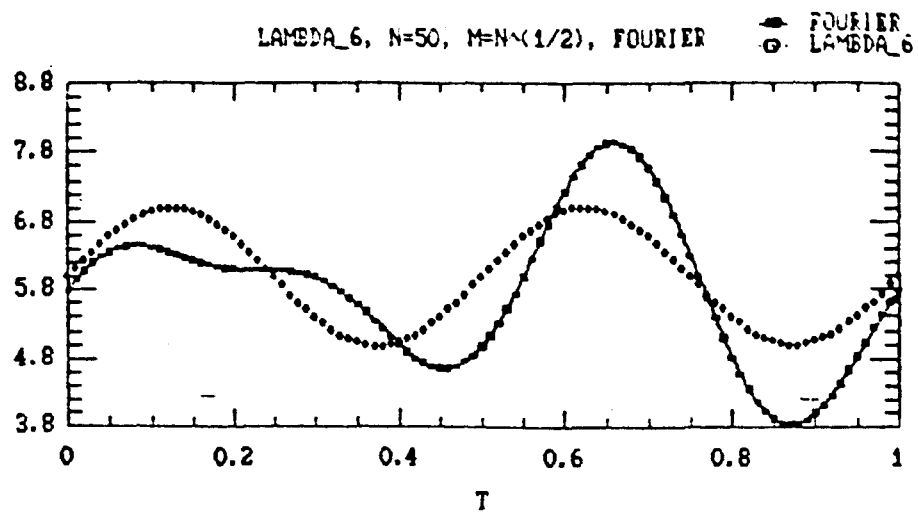


$n = 25$
 $\epsilon = 1/2$



$n = 50$
 $\epsilon = 1/2$





Apêndice III

Programas

Programa para geração da amostra de trajetórias de um processo de Poisson não homogêneo

```
/* É gerado o arquivo SOC "amostra" contendo o número da trajetória e
   os pontos de salto de cada uma delas */

/* cálculo de  $Z(s) = \inf\{ v : \Lambda(v) > s \}$  */
func lambdai()
{
  se (&l <= 1/2) { return(&l/2); }
  se ((&l > 1/2) && (&l <= 7/2)) { return(sqrt(&l/8-3/64)+1/8); }
  se (&l > 7/2) { return((&l+4)/10); }
}

assume semente=339694001;
i=1; s=0; t=1; ns=1; n=10;
enquanto ( i <= n ) {
  u = criarnd(1,1);
  y=-log(1-u);
  s = s+y;
  se ( s >=  $\Lambda(1)$  ) {
    i = i+1; s=0;
  }
  cc {
    se ( ns == 1) {
      amostra = {i p};
    }
    cc {
      amostra = {amostra,i p};
    }
    ns = ns + 1;
  }
}

grave amostra;
```


Programa para cálculo do ponto inicial para o processo iterativo

```
amostra = leia "amostra";
/* o arquivo SOC amostra contém os pontos de cada uma das n
   trajetórias da amostra */
n=10; m=int(n*(1/3)); t=1; pi=3.141592654;
/* a matriz a e o vetor b são construídos tal qual especificado no
   capítulo 4 */
a=(      cos(0*pi*amostra[,2]/t)
      sqrt(2)/t*cos(2*pi*amostra[,2]/t) );
b=(1,0);
l=0;
theta0=vdiag(ident(m));
theta=theta0;
cont1=1;
enquanto (cont1<=3) {
  theta0[1,] = 2 + (cont1 - 1) * 3;
  cont2=1;
  enquanto (cont2<=3) {
    theta0[2,]= -1.0 + cont2 - 1;
    exec "posit.prg";
    se (sinal == -1) {l0=0;}
    cc {l0=soma(log(a*theta0)) + n*t - n*b'*theta0;}
    se (l0>=1) {l=l0; theta=theta0; }
    cont2=cont2+1;
  }
  cont1=cont1+1;
}
imprime l theta;
```

Programa para testar se a função intensidade está contida nos limites do subconjunto do "sieve"

```

/* se para algum dos 'nlamb' pontos do intervalo [0,1] tivermos
    $\lambda(\text{nlamb}) < 1/c_m$  ou  $\lambda(\text{nlamb}) > c_m$ , então o valor da variável
   'sinal' será igual a -1 (+1 caso contrário).
nlamb=100;
sinal=1;
contp=1;
enquanto (contp <= nlamb) {
    lambda =  theta0[1,]
              + theta0[2,]*sqrt(2)/t*cos(2*pi*contp/nlamb) ;
    se ((lambda < 0.001) ::(lambda > 1000)) {
        sinal=-1;
        contp=nlamb+1;
    }
    cc { contp=contp+1; }
}

```

Programa para cálculo dos coeficientes para o "sieve" de histograma

```
/* o vetor 'thetah' contém as estimativas dos coeficientes para o
   "sieve" de histograma calculadas conforme indicado no capítulo 4 */
amostra = leia "amostra";
k = 1; n=10; t=1;
m = int(n*(1/3));
enquanto (k <= m ) {
    se (k == 1)
        { thetah = {1/m 0}; }
    cc {thetah = { thetah,{k/m 0}};}
    k = k+1; }
v = 1;
enquanto (v <= m ) {
    r = 1;
    enquanto ( r <= nlin(amostra) ) {
        se ((amostra[r,2] > ((v-1)/m)) && (amostra[r,2] <= (v/m)))
            { thetah[v,2] = thetah[v,2] + 1; }
        r = r+1;}
    v = v+1;}
imprime thetah $;
thetah[,2] = thetah[,2]*m/(n*t);
imprime thetah $;
```

Programa para cálculo dos coeficientes para o "sieve" de Fourier

```
n=10; m=int(n*(1/3)); t=1; v=1; maximo=20; tol=0.001; iter=0; rp=1;
pi=3.141592654;
/* o arquivo amostra contém os pontos de cada trajetória */
amostra = leia "amostra";
/* a matriz a e o vetor b são construídos conforme especificado no
capítulo 4 */
a={          cos(0*pi*amostra[,2]/t)
  sqrt(2)/t*cos(2*pi*amostra[,2]/t) };
b={1,0};
/* o vetor 'theta' contém o valor inicial para o processo iterativo */
theta={ 5 , -1 };
/* l designa o logaritmo da função de verossimilhança */
/* rp designa a regra de parada (0.001 em nosso caso) */
/* iter é o número máximo de iterações permitido */
/* s é o vetor de primeiras derivadas */
/* h é a matriz Hessiana */
l=soma(log(a*theta)) + n*t - n*b'*theta;
mataux=theta';
tab={iter mataux rp l};
enquanto ((rp > tol) && (iter < maximo)) {
  s=vdiag(ident(m));
  k=1;
  enquanto (k<=m) {
    s[k,]=soma(a[,k]/!(a*theta))-n*b[k,];
    k=k+1;
  }
  h=ident(m);
  i=1;
  enquanto (i<=m) {
    j=i;
    enquanto (j<=m) {
      h[i,j]=-soma((a[,i]*!a[,j])/!((a*theta)#2));
      h[j,i]=h[i,j];
    }
    i=i+1;
  }
  iter=iter+1;
  rp=rp*tol;
}
```

```

        j=j+1;
    }
    i=i+1;
}
iter=iter+1;
v1=v;
enquanto (v1<=20) {
    theta0=theta-((1/2)*v1)*inv(h)*s;
    exec "posit.prg";
    se (sinal == -1){ v1=v1+1; v=v1; }
    cc { l0=soma(log(a*theta0)) + n*t - n*b'*theta0; }
    se (l0 > l) {
        rp=sqrt(squad(theta0-theta))/sqrt(squad(theta));
        theta=theta0;
        l=l0;
        mataux=theta';
        tab={tab,iter mataux rp l};
        v1=2l;
    }
    cc { v1=v1+1; v=v1; }
}
}
se ((sinal == -1) :: (v > 20)) { iter = maximo + 1; }
imprime tab $ ("iter.,theta1,theta2,theta3,r.p.,veros.");

```

REFERÊNCIAS:

- Aalen, O. O. (1978). Nonparametric inference for a family of counting processes, *Ann. Statist.* 6, 701-726.
- Aalen, O. O., and Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations, *Scand. J. Statist.* 5, 141-150.
- Aalen, O. O., Borgan, ø., Keiding, N., and Thormann, J. (1980). Interaction between life history events. Nonparametric analysis for prospective and retrospective data in the presence of censoring, *Scand. J. Statist.* 7, 161-171.
- Altshuler, B. (1970). Theory for the measurement for competing risks in animal experiments, *Math. Biosciences* 6, 1-11.
- Andersen, P. K., and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study, *Ann. Statist.* 10, 1100-1120.
- Andersen, K., and Borgan, ø. (1984). Counting processes models for life history data: a review, 10TH Nordic Conference on Mathematical Statistics, Bolkesjo, Norway.
- Anscombe, F. J. (1953). Sequential estimation, *J. Roy. Statist. Soc. B* 15, 1-29.
- Bartle, R. G. (1966). *The Elements of Integration*, Wiley, New York.
- Basawa, I. V., and Prakasa Rao, B. L. S. (1980). *Statistical Inference for Stochastic Processes*, Academic Press, New York.
- Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

- Borgan, Ø. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data, *Scand. J. Statist.* 11, 1-16.
- Brémaud, P. (1972). A Martingale Approach to Point Processes, Ph. D. Dissertation, Univ. California, Berkeley.
- Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics*, Springer-Verlag, New York.
- Cinlar, E. (1975). *Introduction to Stochastic Processes*, Prentice Hall, Englewood Cliffs, New Jersey.
- Cox, D. R. (1972). Regression models and life tables, *J. Roy. Statist. Soc. B* 34, 187-220.
- Cox, D. R., and Lewis, P. A. W. (1966). *The Statistical Analysis of Series of Events*, Methuen, London.
- Crestana, C. (1991). Um Método para Estimação da Função Intensidade do Processo de Poisson Não Homogêneo, Tese de Mestrado, Unicamp - IMECC.
- Darwin, J. H. (1956). The behaviour of an estimator for a simple birth and death process, *Biometrika* 43, 23-31.
- Epstein, B., and Sobel, M. (1953). Life testing, *J. Amer. Statist. Assoc.* 48, 486-502.
- Feller, W. (1939). Die grundlagen der volterraschen theorie des kampfes ums dasein in wahrscheinlichkeitstheoretischer behandlung, *Acta Blotheoretica* 5, 11-40.
- Fine, N. J. (1949). On the Walsh functions, *Trans. Amer. Math. Soc.* 65, 372-414.

- Fleming, T. R. (1978a). Nonparametric estimation for non-homogeneous Markov processes in the problem of competing risks, *Ann. Statist.* 6
- Fleming, T. R. (1978b). Asymptotic distribution results in competing risks estimations, *Ann. Statist.* 6
- Geman, S. and Hwang, C-R. (1982). Nonparametric maximum likelihood estimation by the method of sieves, *Ann. Statist.* 10, 401-414.
- Gilchrist, R. (1976). Some problems encountered in generating realisations from a useful non-homogeneous Poisson process, *Proceedings of The European Congress of Statisticians*, ed. J. R. Barra et al., Amsterdam, North Holland.
- Grenander, U. (1950). Stochastic processes and statistical inference, *Ark. Mat.* 1, 195-277.
- Grenander, U. (1981). *Abstract Inference*, Wiley, New York.
- Halperin, M. (1952). Maximum likelihood estimation in truncated samples, *Ann. Math. Statist.* 23, 226-238.
- Ibragimov, I. A., and Has'minskii, R. Z. (1981). *Statistical Estimation - Asymptotic Theory*, Springer-Verlag, New York.
- Jacod, J., and Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes*, Springer-Verlag, New York.
- Jenrich, R. I., and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices, *Biometrics* 42, 805-820.
- Kaplan, E. L., and Meyer, P. (1958). Nonparametric estimation from incomplete observations, *J. Amer. Statist. Assoc.* 53, 457-481.

- Kallianpur, G., and Selukar, R. S. (1989). Estimation of Hilbert space valued parameters by the method of sieves, Technical Report no. 278, Center for Stochastic Processes, Univ. North Carolina, Chapel Hill, N.C., USA.
- Karr, A. F. (1986). *Point Processes and Their Statistical Inference*, Marcel Dekker, New York.
- Karr, A. F. (1987). Maximum likelihood estimation in the multiplicative intensity model via sieves, *Ann. Statist.* 15, 473-490.
- Keiding, N. (1975). Maximum likelihood estimation in the birth and death process, *Ann. Statist.* 3, 363-372.
- Kendall, D. G. (1949). Stochastic processes and population growth, *J. Roy. Statist. Soc. B* 11, 230-264.
- Kotz, S., and Johnson, N. L. (1988). *Encyclopedia of Statistical Sciences*, Wiley, New York.
- Kutoyants, Yu. A. (1979). Intensity parameter estimation of an inhomogeneous Poisson process, *Prob. Control Inform. Theory* 8, 137-149.
- Lewis, P. A. W., and Shedler, G. S. (1972). Recent advances in the statistical analysis of univariate point processes, in *Stochastic Point Processes*, ed. P. A. W. Lewis, pp. 1-54, Wiley, New York.
- Lewis, P. A. W., and Shedler, G. S. (1976a). Statistical analysis of non stationary series of events in a data base system, *IBM Journal of Research and Development*, 20, 465-482.
- Lewis, P. A. W., and Shedler, G. S. (1976b). Simulation of

- non-homogeneous Poisson processes with log-linear rate function, *Biometrika* 63, 501-505.
- Lewis, P. A. W., and Shedler, G. S. (1977). Simulation of non-homogeneous Poisson processes with degree-two exponential polynomial rate function, *Opns. Res.* 27, 1026-1040.
- Lewis, P. A. W., and Shedler, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning, *Naval Res. Logistics Quart.* 26, 403-413.
- Liptser, R. S., and Shiriyayev, A. N. (1977). *Statistics of Random Processes*, Vol. I, Springer-Verlag, Berlin.
- Liptser, R. S., and Shiriyayev, A. N. (1978). *Statistics of Random Processes*, Vol. II, Springer-Verlag, Berlin.
- Lima, E. L. (1981). *Curso de Análise*, Vol. II, Projeto Euclides, IMPA, CNPq, Rio de Janeiro.
- MacLean C. J. (1974). Estimation and testing of an exponential polynomial rate function within the nonstationary Poisson process, *Biometrika* 61, 81-85.
- McFadden, J. A. (1965). The entropy of a point process, *SIAM J. Appl. Math.* 13, 988-994.
- Mckeague, I. W. (1986). Estimation for a semimartingale regression model using the method of sieves, *Ann. Statist.* 14, 579-589.
- Moran, P. A. P. (1951). Estimation methods for evolutive processes, *J. Roy. Statist. Soc. B* 13, 141-146.
- Moran, P. A. P. (1953). The estimation of the parameters of a birth and death process, *J. Roy. Statist. Soc. B* 15, 241-245.

- Nelson, W. (1969). Hazard plotting for incomplete failure data, *J. Qual. Tech.* 1, 27-52.
- Nelson, W. (1972). Theory and application of hazard plotting for censored failure data, *Technometrics* 14, 945-966.
- Nguyen, H. T., and Pham, T. D. (1982). Identification of nonstationary diffusion model by the method of sieves, *SIAM J. Control and Optimization* 20, 603-611.
- NTIA (1989). Manual do SOC - Módulo Matricial, EMBRAPA, Campinas.
- Paley, R. E. A. C. (1932). A remarkable series of orthogonal functions, *Proc. London Math. Soc.* 34, 241-279.
- Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces*, Academic Press, New York.
- Parzen, E. (1962). *Stochastic Processes*, Holden-Day, San Francisco.
- Patrow, M. L. (1977). A Comparison of Two Algorithms for The Simulation of Non-homogeneous Poisson Processes with Degree Two Exponential Polynomial Intensity Function, M. S. Thesis, Dept. of Operations Research, Naval Postgraduate School, Monterey, California.
- Ramlau-Hansen, H. (1983a). Smoothing counting processes intensities by means of kernel functions, *Ann. Statist.* 11, 453-466.
- Ramlau-Hansen, H. (1983b). The choice of a kernel function in the graduation of counting process intensities, *Scand. Actuarial J.* 165-182.
- Rényi, A. (1970). *Probability Theory*, North-Holland, Amsterdam.
- Reyment, R. A. (1976). Geomagnetic reversal rate statistically

- appraised, *Tectonophysics* 31, 73-91.
- Royden, H. L. (1968). *Real Analysis*, Macmillan, New York.
- Ross, S. M. (1983). *Stochastic Processes*, Wiley, New York.
- Shiryayev, A. N. (1984). *Probability*, Springer-Verlag, New York.
- Steinijans, V. (1976). A stochastic point process model of the occurrence of major freezes in lake Constance, *Applied Statistics* 25, 58-61.
- Svensson, A. (1990). Asymptotic estimation in counting processes with parametric intensities based on one realization, *Scand. J. Statist.* 17, 23-33.
- Tapia, R. A., and Thompson, J. R. (1978). *Nonparametric Probability Density Estimation*, The Johns Hopkins University Press, Baltimore, Maryland.
- Wald, A. (1949). Note on the consistency of the maximum likelihood estimate, *Ann. Math. Statist.* 20, 595-601.
- Walsh, J. L. (1923). A closed set of normal orthogonal functions, *Amer. J. Math.* 45, 5-24.
- Wolff, R. W. (1965). Problems of statistical inference for birth and death queueing models, *Operations Res.* 13, 343-357.