

**Um Novo Teste para
a Hipótese de Separabilidade
em Processos Pontuais**

Alexandra Maia

alexandra@est.mest.ufmg.br

Dissertação submetida ao departamento de Estatística da

UNIVERSIDADE FEDERAL DE MINAS GERAIS

em 13 de Maio de 2005

em cumprimento às exigências para a obtenção do grau de

MESTRE EM ESTATÍSTICA

Área de Concentração: **Estatística**

Orientador: **Prof. Renato Martins Assunção**

-Este trabalho recebeu apoio financeiro da CAPES: 5/2003-4/2005

Índice

Introdução	6
1 Processo Pontual	8
1.1 Introdução	8
1.2 Preliminares Matemáticas	9
1.3 Processo Estocástico	10
1.4 Processo Pontual: Noções Gerais e Exemplos	11
1.5 Processo de Poisson	14
1.5.1 Processo de Poisson Homogêneo	14
1.5.2 Processo de Poisson Não-Homogêneo	15
1.6 Efeitos de Primeira e Segunda Ordem	19
2 Teste de Escore	24
2.1 Introdução	24
2.2 Teste da Razão de Verossimilhança	24
2.3 Testes Localmente mais Poderosos	25
2.4 Teste de Escore	26
3 Separabilidade	28
3.1 Introdução	28
3.2 Definição e Exemplos	29
3.3 Investigando a Hipótese de Separabilidade	31
3.4 Testes de Separabilidade	32
4 Um Teste de Escore para Separabilidade	33
4.1 Introdução	33

4.2	Definindo a Estatística de Teste	34
4.3	Relação entre as Estatísticas de Teste	36
5	Simulação	40
5.1	Introdução	40
5.2	Realizando os Testes	41
	Conclusões Finais	59

Lista de Figuras

1.1	<i>Localizações de 62 mudas de sequóia em um quadrado de lado 23 metros (Strauss, 1975; Ripley, 1977)</i>	13
1.2	<i>Localizações de 42 células em uma seção de tecido (Ripley, 1977)</i>	14
1.3	<i>Processo de Poisson Homogêneo com média 70</i>	15
1.4	<i>Processo de Poisson Não-Homogêneo com função intensidade $\lambda(\mathbf{x}) = e^{(3+3x_1)} + e^{(3+x_2)}$</i>	16
1.5	<i>Processo de Poisson Não-Homogêneo com função intensidade $\lambda(\mathbf{x}) = e^{(2x_1+4x_2)}$</i>	17
1.6	<i>Processo de Poisson Não-Homogêneo com função intensidade $\lambda(\mathbf{x}) = 40 I_{\{(x_1, x_2) \in [0, 0.5]^2\}} + 480 I_{\{(x_1, x_2) \in [0.75, 1]^2\}}$</i>	18
1.7	<i>Estimação por kernel nas localizações de 42 células em uma seção de tecido com largura de banda 0.05</i>	21
1.8	<i>Estimação por kernel nas localizações de 42 células em uma seção de tecido com largura de banda 0.25</i>	22
1.9	<i>Estimação por kernel nas localizações de 62 mudas de sequóia em um quadrado de lado 23 metros com largura de banda 0.05</i>	23
1.10	<i>Estimação por kernel nas localizações de 62 mudas de sequóia em um quadrado de lado 23 metros com largura de banda 0.25</i>	23
4.1	<i>Modelo exponencial aditivo não-separável. Os gráficos na parte superior são as superfícies da intensidade condicional $\lambda = \lambda(t, \mathbf{x}, m)$ (esquerda) e o produto das intensidades marginais $s = s(t, \mathbf{x}, m)$ (direita). Da esquerda para a direita, na parte inferior $(\lambda - s)^2$, $(\lambda - s)^2/\lambda$, e $(\lambda - s)^2 \lambda$.</i>	39

5.1	<i>Processo de Poisson Homogêneo com 100 pontos no quadrado unitário.</i>	42
5.2	<i>Processo de Poisson Não-Homogêneo que gera 100 pontos uniformemente em $[0, 1] \times [0, 1]$ e 100 pontos em $[0.5, 0.75] \times [0.5, 0.75]$.</i>	42
5.3	<i>Curvas poder para as estatísticas $S_1 - S_5$ e T em 1000 realizações do processo $\lambda(t, \mathbf{x}, m) = 100 + 100\epsilon I_{\{(t,m) \in [0.5,0.75]^2\}}$.</i>	43
5.4	<i>Processo que gera 4 centros de clusters e distribui, ao redor de cada centro de cluster, 49 pontos segundo uma distribuição normal com $1/\sigma = \epsilon = 0.1$.</i>	44
5.5	<i>Processo que gera 4 centros de clusters e distribui, ao redor de cada centro de cluster, 49 pontos segundo uma distribuição normal com $1/\sigma = \epsilon = 16.67$.</i>	44
5.6	<i>Curvas poder para as estatísticas $S_1 - S_5$ e T em 1000 realizações do processo que gera 4 clusters aleatoriamente.</i>	45
5.7	<i>Processo que gera 100 pontos distribuídos uniformemente e 100 pontos distribuídos segundo uma normal bivariada com $1/\sigma = \epsilon = 0.1$.</i>	46
5.8	<i>Processo que gera 100 pontos distribuídos uniformemente e 100 pontos distribuídos segundo uma normal bivariada com $1/\sigma = \epsilon = 16.67$.</i>	46
5.9	<i>Curvas poder para as estatísticas $S_1 - S_5$ e T em 1000 realizações do processo que gera 100 pontos distribuídos uniformemente em $[0, 1]^2$ e 100 pontos distribuídos segundo uma normal bivariada.</i>	47
5.10	<i>Processo de cluster Neyman-Scott com $\alpha = 20$, $k = 3$ e $\epsilon = 0.67$.</i>	48
5.11	<i>Processo de cluster Neyman-Scott com $\alpha = 20$, $k = 3$ e $\epsilon = 20$.</i>	48
5.12	<i>Curvas poder para as estatísticas $S_1 - S_5$ e T em 1000 realizações do processo de cluster Neyman-Scott.</i>	49
5.13	<i>Processo de cluster Thomas com $\alpha = 20$, $\beta = 3$ e $\epsilon = 0.67$.</i>	50
5.14	<i>Processo de cluster Thomas com $\alpha = 20$, $\beta = 3$ e $\epsilon = 25$.</i>	50
5.15	<i>Curvas poder para as estatísticas $S_1 - S_5$ e T em 1000 realizações do processo de cluster Thomas.</i>	51
5.16	<i>Processo de cluster Matern com $\alpha = 20$, $\beta = 3$ e $\epsilon = 0.67$.</i>	52
5.17	<i>Processo de cluster Matern com $\alpha = 20$, $\beta = 3$ e $\epsilon = 20$.</i>	52
5.18	<i>Curvas poder para as estatísticas $S_1 - S_5$ e T em 1000 realizações do processo de cluster Matern.</i>	53

5.19	<i>Processo de inibição MaternI com $\alpha = 70$, e $\epsilon = 0.005$.</i>	54
5.20	<i>Processo de inibição MaternI com $\alpha = 70$, e $\epsilon = 0.07$.</i>	54
5.21	<i>Curvas poder para as estatísticas $S_1 - S_5$ e T em 1000 realizações do processo de inibição MaternI.</i>	55
5.22	<i>Processo de inibição MaternII com $\alpha = 50$, e $\epsilon = 0.005$.</i>	56
5.23	<i>Processo de inibição MaternII com $\alpha = 50$, e $\epsilon = 0.07$.</i>	56
5.24	<i>Processo de inibição SSI com $k = 20$, e $\epsilon = 0.005$.</i>	57
5.25	<i>Processo de inibição SSI com $k = 20$, e $\epsilon = 0.07$.</i>	57

Introdução

Modelos de processos pontuais multi-dimensionais, tais como modelos de processos pontuais espaço-temporais marcados, têm sido crescentemente usados em uma variedade de aplicações para representar observações de eventos discretos tais como terremotos, queimadas, incidência de epidemias, etc (Ripley, 1977; Diggle, 1983; Schoenberg, Brillinger e Guttorp, 2002). Os modelos frequentemente utilizados em tais aplicações, quase invariavelmente têm a intensidade na forma de produto de duas funções, uma envolvendo apenas o espaço e tempo e a outra apenas as marcas. Neste caso, a intensidade é dita *separável* (Cressie, 1993). A hipótese de separabilidade é muito forte: por exemplo, um modelo separável para queimadas afirma que a intensidade de eventos (focos de queimada) com marcas distintas (as marcas podem ser, por exemplo, tamanhos dos focos), por exemplo m_1 e m_2 , se difere apenas por uma constante:

$$\lambda(t, x, m_1) = \underbrace{g(t, x) h(m_1)}_{c_1 g(t, x)} \quad \lambda(t, x, m_2) = \underbrace{g(t, x) h(m_2)}_{c_2 g(t, x)},$$

em que c_1 e c_2 são constantes. Além disso, análises, como estimação de parâmetros, por exemplo, podem ser feitas separadamente, facilitando muito o estudo.

Apesar da importância desta suposição, a separabilidade de tais processos raramente é rigorosamente examinada. Como exceção, Schoenberg (2004) descreveu e comparou testes não-paramétricos para investigar a separabilidade de processos pontuais multi-dimensionais. Mostrou que um destes testes (S_3 , apresentado na Seção 3.4) é mais poderoso para identificar alternativas graduais de separabilidade, e que um teste residual baseado em reescalonamento aleatório do processo é mais poderoso para detectar cluster ou regularidade. Apresentou uma aplicação em dados de queimadas na qual mostrou que a hipótese de separabilidade é invalidada.

Neste trabalho apresentamos um teste não-paramétrico, baseado na estatística de escore, para testar se um processo pontual multi-dimensional é separável. Realizamos simulações para investigar o desempenho de cada estatística (a estatística apresentada neste trabalho e algumas apresentadas em Schoenberg, 2004). Seguindo uma revisão da terminologia de processo pontual no Capítulo 1 e teste de escore no Capítulo 2, a separabilidade é definida mais precisamente em termos da intensidade condicional no Capítulo 3, onde são apresentados exemplos de processos separáveis e não-separáveis e também métodos diretos para testar separabilidade, baseados na comparação das estimativas da intensidade. No Capítulo 4 apresentamos o teste baseado na estatística de escore, proposto neste trabalho, e para o qual esperamos bom desempenho em alternativas de separabilidade, já que é baseado em um teste localmente mais poderoso. Neste capítulo verificamos também a semelhança deste teste com as estatísticas S_3 , S_4 e S_5 , apresentadas em Schoenberg (2004), as quais apresentam bom desempenho diante de alternativas graduais de separabilidade. Devido a essa semelhança, esperamos que o desempenho dessas estatísticas seja parecido. No Capítulo 5 apresentamos as simulações utilizadas para investigar o poder de cada um dos testes sob várias alternativas de separabilidade. Alguns dos processos foram avaliados também por Schoenberg (2004). No Capítulo Conclusões Finais apresentamos as conclusões deste estudo.

Capítulo 1

Processo Pontual

1.1 Introdução

O processo estocástico utilizado para modelar dados na forma de uma coleção enumerável de pontos, distribuídos de maneira aparentemente aleatória sobre algum conjunto, é denominado *Processo Pontual*. O tipo mais simples e conhecido de processo pontual é o Processo de Poisson.

Um *Processo Pontual Espacial* é um processo estocástico que gera um conjunto enumerável de eventos, em uma região de estudo no plano. Há várias extensões para a estrutura básica de Processo Pontual Espacial: Processos Pontuais Multivariados, Processos Espaço-Temporais, Processo Pontual Marcado.

Em geral, o comportamento de fenômenos espaciais é freqüentemente o resultado de uma mistura dos *efeitos de primeira ordem e segunda ordem*. O Método de Estimção por Kernel, utilizado para investigar efeitos de primeira ordem, é empregado para estimar a intensidade de um processo pontual. Neste capítulo, apresentamos algumas definições preliminares de matemática na Seção 1.2 e definições de processo estocástico e processo pontual na Seção 1.3. Exemplos de processos pontuais são apresentados na Seção 1.4, incluindo as definições dos Processos de Poisson Homogêneo e Não-Homogêneo. Definimos, na Seção 1.6 os efeitos de primeira e segunda ordem, bem como o Método de Estimção por Kernel. Sem perda de generalidade, alguns exemplos e definições apresentados neste Capítulo são sobre processo pontual espacial, sendo claro que podem ser estendidos para processos pontuais mais gerais.

1.2 Preliminares Matemáticas

Um conceito básico na teoria de probabilidade e processos estocásticos é o de *espaço de probabilidade*. Um espaço de probabilidade é compreendido por três quantidades: um *espaço amostral* Ω , uma coleção \mathcal{F} de certos subconjuntos de Ω chamada *álgebra de eventos*, e uma *medida de probabilidade* μ . Estas quantidades são definidas como se segue.

Espaço Amostral

Um elemento ω de Ω é um possível resultado de um experimento aleatório. Em situações mais simples Ω consiste de uma coleção enumerável de elementos $\omega_1, \omega_2, \dots$. Exemplos incluem o experimento aleatório de lançamento de um dado, para o qual Ω consiste de seis elementos com ω_i correspondendo à i -ésima face mostrada no lançamento, e o experimento aleatório de lançamento de uma moeda até aparecer cara, para o qual Ω consiste de um número infinito de elementos com ω_i correspondendo a ocorrência da face cara no i -ésimo lançamento. Espaços amostrais não enumeráveis são comuns. Exemplos são o experimento aleatório altura média (em metros) dos alunos de uma sala de aula, para o qual $\Omega = \{\omega : 0 < \omega \leq 2,5\}$, e o experimento aleatório tempo de falha de uma lâmpada incandescente ligada no tempo t_0 , para o qual $\Omega = \{\omega : \omega \geq t_0\}$.

Sigma Álgebra de Eventos

Um subconjunto de Ω é chamado um *evento*. Um exemplo é o subconjunto $A = \{\omega : 1,5 \leq \omega \leq 1,70\}$ do espaço amostral $\Omega = \{\omega \leq 2,5\}$ associado ao experimento aleatório altura média (em metros) dos alunos de uma sala de aula; em palavras, A é o evento: a altura média dos alunos está entre 1,5 e 1,70. A sigma álgebra de eventos \mathcal{F} é uma coleção particular de eventos. Para espaços amostrais enumeráveis, \mathcal{F} pode consistir de todos os subconjuntos possíveis de Ω , mas não é necessário. Para espaços amostrais não enumeráveis, \mathcal{F} não pode consistir de todos os possíveis subconjuntos de Ω , mas apenas de certos subconjuntos chamados *eventos mensuráveis*. As seguintes propriedades são requeridas de \mathcal{F} :

1. \mathcal{F} não é vazio;
2. se um evento A está em \mathcal{F} então o complementar $A^c = \{\omega : \omega \in \Omega, \omega \notin A\}$;
3. se A_1, A_2, \dots são eventos em \mathcal{F} então a união $\bigcup_i A_i = \{\omega : \omega \in A_i, \text{ para algum } i\}$.

Medida de Probabilidade

A medida de probabilidade $\mu(\cdot)$ é uma função que leva cada conjunto mensurável em \mathcal{F} no intervalo unitário $[0, 1]$. As seguintes propriedades são requeridas desta função:

1. $0 \leq \mu(A) \leq 1$ para todo conjunto mensurável $A \in \mathcal{F}$;
2. $\mu(\Omega) = 1$;
3. se A_1, A_2, \dots são conjuntos disjuntos em \mathcal{F} então $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$.

Um evento A é dito ocorrer *com probabilidade um* ou *quase certo* se $\mu(A) = 1$.

Um espaço de probabilidade, denotado por $(\Omega, \mathcal{F}, \mu)$, consiste de um espaço amostral Ω , uma sigma álgebra \mathcal{F} de subconjuntos de Ω e uma medida de probabilidade μ (Snyder e Miller, 1991).

1.3 Processo Estocástico

Denotando por X_n o valor de uma ação numa bolsa de valores na n -ésima unidade de tempo, podemos representar sua evolução por uma família de variáveis aleatórias $\{X_0, X_1, \dots\}$ indexadas pelo parâmetro discreto $n \in \mathbb{Z}_+$. O número de acidentes de carro em uma cidade durante o intervalo de tempo $[0, t]$ fornece uma coleção de variáveis aleatórias $\{X_t : t \geq 0\}$ indexadas pelo parâmetro contínuo t . A velocidade $X_{\mathbf{u}}$ de vento turbulento em um ponto \mathbf{u} em um campo fornece uma família de variáveis aleatórias $\{X_{\mathbf{u}} : \mathbf{u} \in \mathbb{R}^3\}$ indexadas pelo parâmetro espacial multidimensional \mathbf{u} . Mais geralmente temos a seguinte definição.

Definição 1.1. *Dado um conjunto de índices I , um processo estocástico indexado por I é uma coleção de variáveis aleatórias $\{X_\lambda : \lambda \in I\}$ sobre o espaço de probabilidade*

(Ω, \mathcal{F}, P) assumindo valores em um conjunto S . O conjunto S é chamado espaço de estado do processo.

- Nos exemplos acima temos, respectivamente: (i) $I = \mathbb{Z}_+$, $S = \mathbb{R}_+$;
(ii) $I = [0, \infty)$, $S = \mathbb{Z}_+$; (iii) $I = \mathbb{R}^3$, $S = \mathbb{R}^3$ (Bhattacharya e Waymire, 1990).

Processo Pontual

Seja B um conjunto em um espaço n -dimensional e seja \mathcal{A} uma família de subconjuntos de B . Um *processo pontual* é um processo estocástico indexado por conjuntos $A \in \mathcal{A}$ e tendo o conjunto $\{0, 1, \dots, \infty\}$ de inteiros não-negativos como seu espaço de estado. Podemos pensar em "pontos" surgindo em B de forma aleatória e $N(A)$ como a contagem do número de pontos no conjunto A . Desde que $N(A)$ é uma função de contagem há exigências adicionais sobre cada realização. Por exemplo, se A_1 e A_2 são conjuntos mutuamente disjuntos em \mathcal{A} cuja união $A_1 \cup A_2$ está também em \mathcal{A} , então

$$N(A_1 \cup A_2) = N(A_1) + N(A_2),$$

e se o conjunto vazio \emptyset está em \mathcal{A} então $N(\emptyset) = 0$ (Karlin, 1975).

Na próxima Seção apresentamos alguns conceitos e exemplos de processo pontual.

1.4 Processo Pontual: Noções Gerais e Exemplos

Investigações em diversas áreas da ciência conduzem a dados na forma de uma coleção enumerável de pontos distribuídos de maneira aparentemente aleatória sobre algum conjunto. Frequentemente os dados podem conter tempos de ocorrência de algum fenômeno, ou localizações nas quais alguns objetos são observados ou algum fenômeno ocorre. Exemplos simples de tais dados são: tempos de emissão de pulsos em uma fibra nervosa, tempos de chegada de pacientes em uma unidade de tratamento intensivo, localizações de árvores em uma floresta.

O tipo particular de processo estocástico utilizado para modelar tais dados é conhecido como *processo pontual*. Assim, um processo pontual é um modelo matemático para descrever um conjunto enumerável de pontos aleatoriamente distribuídos sobre

algum espaço. O Processo de Poisson fornece o mais simples e conhecido tipo de processo pontual.

A teoria de processo pontual é rica em estrutura matemática, tendo muitas conexões importantes com outras áreas de probabilidade e teoria de processo estocástico. Suas origens (cf. Daley e Vere-jones, 1988) estão nas áreas de tabelas de vida e teoria de renovação, problemas de contagem começando com trabalho de S. D. Poisson e conduzindo a aplicações em física de partículas e processos de população, teoria de filas e engenharia de comunicação (Milne, 2001).

Um *Processo Pontual Espacial* é um processo estocástico que gera um conjunto enumerável de eventos s_i em uma região de estudo A no plano. Usamos o termo "eventos" para designar uma ampla variedade de fenômenos espaciais que podem ser considerados ocorrendo em localizações pontuais (Bailey e Gatrell, 1995). Alguns exemplos de processo pontual espacial incluem localizações de crateras vulcânicas, núcleos de células em uma seção de tecido, endereço de residência (ou local de infecção) de casos de alguma doença, epicentros de terremotos, focos de queimadas, etc.

Há várias extensões para a estrutura básica de Processo Pontual Espacial. *Processos Pontuais Multivariados* são aqueles nos quais os eventos são da forma (s_i, δ_i) onde δ_i registra o tipo do i -ésimo evento dentre dois ou mais tipos qualitativamente diferentes. Um exemplo são as localizações de duas espécies de árvores em uma floresta, onde pode se estar interessado em investigar a relação espacial entre as duas espécies. Neste caso, $\delta_i = 1$ ou 2 e indica a espécie da i -ésima árvore. *Processos Espaço-Temporais* são aqueles nos quais cada evento é da forma (s_i, t_i) onde t_i é o tempo de ocorrência associado ao i -ésimo evento. Exemplos incluem dados de terremotos, onde a localização e data de cada terremoto foi registrada, ou dados epidemiológicos, onde a localização e data do início de cada caso de uma particular doença foi registrado. O caso mais geral é quando cada ponto tem associado a ele um conjunto de variáveis, que podem ser categóricas ou contínuas. Este é chamado *Processo Pontual Marcado*. Um exemplo seria dados de terremotos nos quais a localização, data, magnitude (ou tipo) e valor econômico dos danos de cada terremoto foi medido. Observamos que um processo multivariado pode ser considerado como processo marcado com uma única variável de marca categórica, e um processo espaço-temporal como um processo mar-

cado com uma única variável de marca contínua (Diggle e Rowlingson, 1993).

Um dos interesses em analisar um processo pontual espacial está, basicamente, em investigar se os eventos observados exibem qualquer padrão sistemático, contrário a um padrão aleatório na região de estudo A (Bailey e Gatrell, 1995 - pág 76). Quando a ocorrência de um evento em uma particular localização, aumenta a probabilidade de ocorrências de outros eventos próximos a ela, a configuração resultante exibe um tipo de padrão descrito como *agregado* (ou *cluster*). Um exemplo de um padrão agregado, extraído por Ripley (1977) de Strauss (1975), é apresentado na Figura 1.1: são as localizações de 62 mudas de sequóia em um quadrado cujo lado mede aproximadamente 23 metros. Em contraste, quando é provável que haja uma certa distância entre os eventos, o padrão total será uma distribuição espacial de eventos mais *regular*. Um exemplo de padrão regular é apresentado na Figura 1.2: são as localizações de 42 células em uma seção de tecido (Diggle, 2003 - pág 63).

Se algum padrão sistemático parece estar presente, podemos estar interessados em investigar sobre qual escala isto ocorre e se agregações espaciais particulares, ou clusters, estão associados com a sua proximidade a fontes particulares de algum outro fator (como por exemplo casos de doenças próximos à uma indústria que lança um tipo de resíduo no ar). Em casos onde os eventos são de diferentes tipos ou ocorrem em diferentes pontos no tempo, ou em uma região com uma 'população de risco' heterogênea, várias outras hipóteses se tornam possíveis (Bailey e Gatrell, 1995).

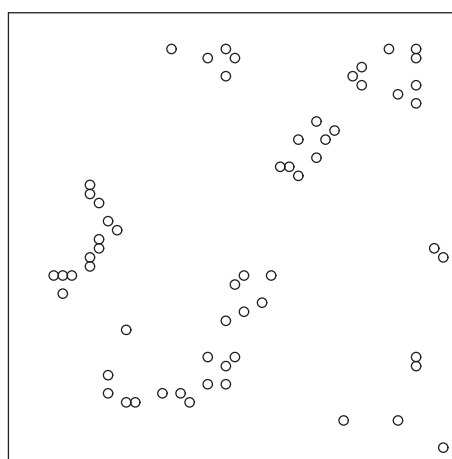


Figura 1.1: *Localizações de 62 mudas de sequóia em um quadrado de lado 23 metros (Strauss, 1975; Ripley, 1977)*

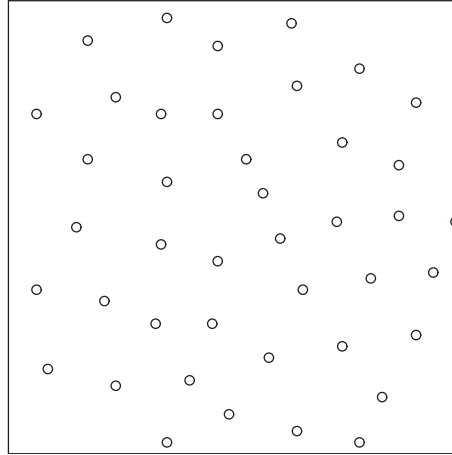


Figura 1.2: *Localizações de 42 células em uma seção de tecido (Ripley, 1977)*

1.5 Processo de Poisson

1.5.1 Processo de Poisson Homogêneo

O Processo de Poisson (homogêneo) é a pedra angular na qual a teoria de processo pontual espacial é construída. Ele representa o mecanismo estocástico mais simples possível para a geração de processo pontual espacial, e em aplicações é usado como um padrão idealizado de CSR (Complete Spatial Randomness), aleatoriedade espacial completa, a qual, se estritamente inatingível na prática, algumas vezes fornece uma descrição aproximada de um padrão observado. O Processo de Poisson está convenientemente definido pelos seguintes postulados:

PP1 - Para algum $\lambda > 0$, e qualquer região plana finita A , $N(A)$ (número de eventos em A), segue uma distribuição de Poisson com média $\lambda|A|$ ($|A|$ corresponde à área de A).

PP2 - Dado $N(A) = n$, os n eventos em A formam uma amostra aleatória independente com distribuição uniforme em A .

Para simular uma realização de um processo de Poisson em A dado um valor fixo de $N(A)$, precisamos gerar eventos independentemente e uniformemente distribuídos

sobre A . Em regiões com formatos inconvenientes, basta simular o processo em uma região maior, em forma de retângulo ou disco, e reter apenas os pontos que estão em A (Diggle, 2003 - págs 47, 48).

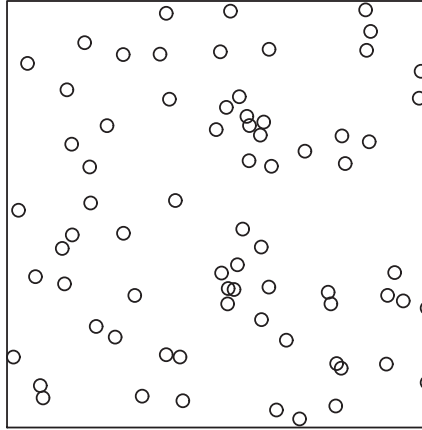


Figura 1.3: *Processo de Poisson Homogêneo com média 70*

Na Figura 1.3 foi simulado um Processo de Poisson Homogêneo com média 70, segundo PP2. Foram gerados 72 pontos uniformemente distribuídos no quadrado unitário.

1.5.2 Processo de Poisson Não-Homogêneo

Uma outra classe de processo pontual é obtida se a média λ do processo de Poisson é substituída por uma função intensidade variando espacialmente, $\lambda(x)$. Isto define a classe de Processo de Poisson Não-Homogêneo com as seguintes propriedades:

PPNH1 - $N(A)$ tem uma distribuição de Poisson com média $\int_A \lambda(x)dx$.

PPNH2 - Dado $N(A) = n$, os n eventos em A formam uma amostra aleatória independente a partir da distribuição em A com função densidade de probabilidade proporcional a $\lambda(x)$.

O método para simular um processo de Poisson não-homogêneo é utilizando PPNH2. Um algoritmo, em sua forma mais simples, consiste em simular um pro-

cesso de Poisson em A com média λ_0 igual ao valor máximo de $\lambda(x)$ em A , e reter um evento em x com probabilidade $\lambda(x)/\lambda_0$ (Diggle, 2003 - págs 67, 68).

Apresentamos, a seguir, alguns exemplos de Processo de Poisson Não-Homogêneo. Tomamos $\mathbf{x} = (x_1, x_2)$. Estes processos foram gerados utilizando o algoritmo acima descrito. A Figura 1.4 mostra o processo com função intensidade $\lambda(\mathbf{x}) = e^{(3+3x_1)} + e^{(3+x_2)}$. Podemos observar que o número de pontos aumenta no sentido de crescimento das coordenadas. A coordenada x_1 contribui razoavelmente para que um ponto seja retido, desta forma, pontos em que a coordenada x_1 apresenta valores mais altos têm maior probabilidade de ser retidos. Este processo gera em média 162 pontos em $[0, 1] \times [0, 1]$ (o valor da integral de $\lambda(\mathbf{x})$ no quadrado unitário).

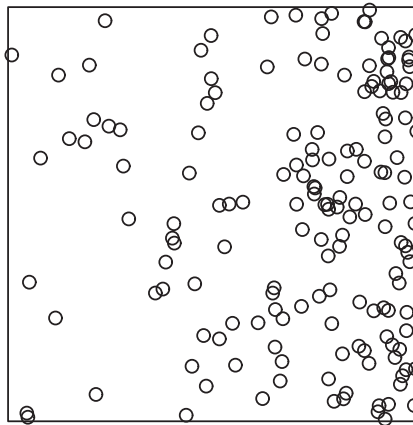


Figura 1.4: *Processo de Poisson Não-Homogêneo com função intensidade $\lambda(\mathbf{x}) = e^{(3+3x_1)} + e^{(3+x_2)}$.*

Outro exemplo é apresentado na Figura 1.5. Este processo, com função intensidade $\lambda(\mathbf{x}) = e^{(2x_1+4x_2)}$, gera, em média 43 pontos. Observamos que há um maior número de pontos onde as coordenadas x_1 e x_2 têm maior valor.

O processo apresentado na Figura 1.6, com função intensidade $\lambda(\mathbf{x}) = 40 I_{\{(x_1,x_2) \in [0,0.5]^2\}} + 480 I_{\{(x_1,x_2) \in [0.75,1]^2\}}$, é muito interessante. Este processo gera, em média, 10 pontos em $[0, 0.5] \times [0, 0.5]$ e 30 pontos em $[0.75, 1] \times [0.75, 1]$. É um processo de Poisson Não-Homogêneo em $[0, 1] \times [0, 1]$, mas é homogêneo em $[0, 0.5] \times [0, 0.5]$ e em $[0.75, 1] \times [0.75, 1]$. Pode-se simular uma realização deste processo através do algoritmo (como apresentado nas Figuras 1.4 e 1.5) ou gerando pontos uniformemente distribuídos em $[0, 0.5] \times [0, 0.5]$ e $[0.75, 1] \times [0.75, 1]$, como na Figura 1.3.

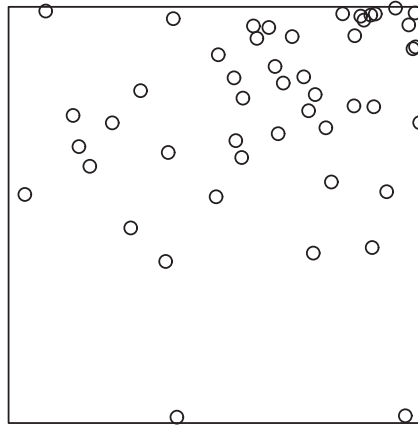


Figura 1.5: *Processo de Poisson Não-Homogêneo com função intensidade $\lambda(\mathbf{x}) = e^{(2x_1+4x_2)}$.*

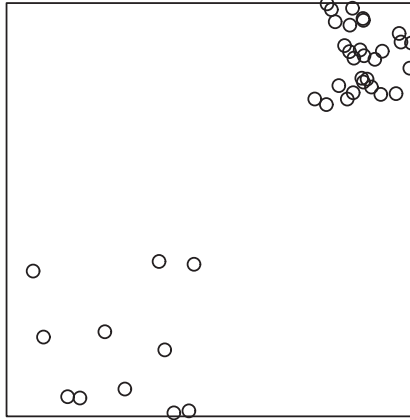


Figura 1.6: *Processo de Poisson Não-Homogêneo com função intensidade $\lambda(\mathbf{x}) = 40 I_{\{(x_1, x_2) \in [0, 0.5]^2\}} + 480 I_{\{(x_1, x_2) \in [0.75, 1]^2\}}$.*

Log-Verossimilhança do Processo de Poisson Não-Homogêneo

Apresentamos nessa Seção, sem perda de generalidade, a verossimilhança para um processo de Poisson dependente do tempo. Para casos mais gerais, como processo espaço-temporal marcado é análogo.

Consideremos um processo de Poisson com função intensidade $\lambda(t)$ observada no intervalo $[0, t_0)$. Sejam os eventos observados no tempo y_1, \dots, y_n . Vamos dividir o intervalo $[0, t_0)$ em um grande número m de subintervalos cada um com comprimento h tal que $mh = t_0$, e vamos denotar estes intervalos por $[a_j, a_j + h)$ para $j = 1, \dots, m$. Então, por propriedades do processo de Poisson, o subintervalo $[a_j, a_j + h)$ contribui para a verossimilhança com um fator $\lambda(a_j)h + o(h) = \lambda(y_i)h + o(h)$ se $a_j \leq y_i < a_j + h$ para algum i , enquanto que se um evento não ocorre em $[a_j, a_j + h)$ a contribuição é $1 - \lambda(a_j)h + o(h)$. Como os números de eventos em intervalos disjuntos são independentes, temos

$$\prod_{i=1}^n \{\lambda(y_i)h + o(h)\} \prod_j^* \{1 - \lambda(a_j)h + o(h)\}, \quad (1.1)$$

onde \prod^* é o produto sobre todo j tal que $[a_j, a_j + h)$ não contém nenhum y_i , $i = 1, \dots, n$. Quando $h \rightarrow 0$ este segundo produto tende a

$$\exp \left\{ - \int_0^{t_0} \lambda(u) du \right\}.$$

Se omitirmos o fator h^n em 1.1, a omissão corresponde à conversão de uma probabilidade em uma densidade de probabilidade, assim podemos tomar

$$l\{\lambda(t); y_1, \dots, y_n\} = \left\{ \prod_{j=1}^n \lambda(y_j) \right\} \exp \left\{ - \int_0^{t_0} \lambda(u) du \right\} \quad (1.2)$$

(Cox e Hinkley, 1974).

1.6 Efeitos de Primeira e Segunda Ordem

Em geral, o comportamento de fenômenos espaciais é freqüentemente o resultado de uma mistura dos *efeitos de primeira ordem* e dos *efeitos de segunda ordem*. Os efeitos de primeira ordem se relacionam com a variação no valor médio do processo no espaço e são descritos em termos da *intensidade*, $\lambda(s)$ do processo a qual é o número médio de eventos por unidade de área no ponto s . Formalmente, é definida como um limite matemático:

$$\lambda(s) = \lim_{|ds| \rightarrow 0} \left\{ \frac{E[N(ds)]}{|ds|} \right\}$$

onde ds é uma pequena região ao redor do ponto s , $|ds|$ é a área de ds e $N(A)$ é o número de eventos ocorrendo na região A . Os efeitos de segunda ordem resultam da estrutura de correlação espacial ou dependência espacial no processo e são formalmente descritos pela *intensidade de segunda ordem*, $\gamma(s_i, s_j)$ do processo que novamente envolve eventos por unidade de área e é formalmente definida como o limite matemático:

$$\gamma(s_i, s_j) = \lim_{|ds_i|, |ds_j| \rightarrow 0} \left\{ \frac{E[N(ds_i)N(ds_j)]}{|ds_i||ds_j|} \right\}$$

com notação similar à usada anteriormente.

Há vários métodos para investigar efeitos de primeira e segunda ordem, mas o que precisamos para testar a hipótese de separabilidade, e que será utilizado durante todo o trabalho, é o Método de Estimação por Kernel, utilizado para investigar efeitos

de primeira ordem (Bailey e Gatrell, 1995 - págs 32 e 77). Embora alguns autores brasileiros usem a expressão "núcleo", preferimos adotar a palavra "kernel", um caso de anglicismo já usual em português.

Estimação por Kernel

A *Estimação por Kernel* foi originalmente desenvolvida para obter uma estimativa suave de uma densidade de probabilidade univariada ou bivariada a partir de uma amostra observada; em outras palavras, um histograma suave. Estimar a intensidade de um processo pontual espacial é como estimar uma densidade de probabilidade bivariada e a estimação de kernel bivariada pode ser facilmente adaptada para fornecer uma estimativa de intensidade. Se s representa uma localização geral em A , e s_1, \dots, s_n são as localizações dos n eventos observados, então a intensidade $\lambda(s)$, em s é estimada por

$$\hat{\lambda}_\tau(s) = \frac{1}{\delta_\tau(s)} \sum_{i=1}^n \frac{1}{\tau^2} k\left(\frac{s - s_i}{\tau}\right)$$

As distâncias $s - s_i$ do ponto avaliado s a cada evento observado s_i , em unidades de τ , são medidas e contribuem para a estimativa de $\lambda(s)$ de acordo com a proximidade de cada evento s_i a s . $k(\cdot)$ é uma função de densidade de probabilidade bivariada escolhida adequadamente, conhecida como *kernel*, a qual é simétrica em relação à origem. O parâmetro $\tau > 0$ é conhecido como largura de banda (*bandwidth*) e determina a quantidade de suavização - essencialmente é o raio de um disco centrado em s dentro do qual os pontos s_i contribuirão significativamente para $\hat{\lambda}_\tau(s)$. O fator

$$\delta_\tau(s) = \int_A \frac{1}{\tau^2} k\left(\frac{s - u}{\tau}\right) du$$

é uma correção de borda - o volume sob o kernel centrado em s que está contido em A .

Para quaisquer kernel e largura de banda escolhidos, os valores de $\hat{\lambda}_\tau(s)$ podem ser examinados nas localizações de uma grade de pontos, adequadamente escolhida sobre A , para fornecer uma indicação visual da variação na intensidade $\lambda(s)$, sobre a região de estudo. Uma típica escolha para $k(\cdot)$ pode ser o kernel quártico

$$k(u) = \begin{cases} \frac{3}{\pi}(1 - u^T u)^2 & \text{se } u^T u \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

Quando realizamos tal estimação de kernel sobre uma grade de localizações em A , obtemos uma função tri-dimensional avaliada em cada ponto s da grade.

Seja qual for a escolha de $k(\cdot)$, o efeito de crescimento da largura de banda τ é aumentar a região ao redor de s dentro da qual os eventos observados influenciam a estimativa em s . Para um valor de τ muito grande, $\hat{\lambda}_\tau(s)$ será mais plana e as características locais serão obscurecidas; se τ é pequeno, então $\hat{\lambda}_\tau(s)$ será uma coleção de picos centrados em s_i .

Podemos observar essas diferenças entre escolhas de larguras de banda nos exemplos apresentados nas Figuras 1.7 - 1.10. Realizamos uma estimativa por kernel, utilizando o kernel quártico em uma grade 100×100 (ou seja, a intensidade é avaliada em 100 pontos na coordenada x e 100 pontos na coordenada y) com larguras de banda 0.05 e 0.25, respectivamente.

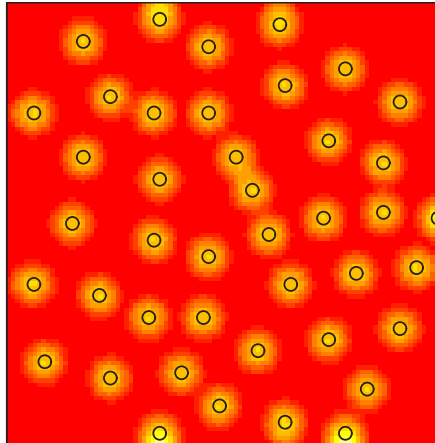


Figura 1.7: *Estimação por kernel nas localizações de 42 células em uma seção de tecido com largura de banda 0.05*

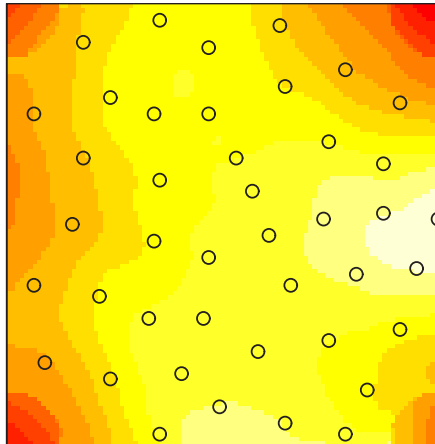


Figura 1.8: *Estimação por kernel nas localizações de 42 células em uma seção de tecido com largura de banda 0.25*

As áreas mais claras, nesses gráficos são as áreas de maior intensidade. Nas figuras 1.7 e 1.8, processo de inibição das localizações das células, observamos grandes diferenças. Quando a largura de banda é menor (0.05) a estimativa de intensidade aparece como picos, com maior valor nas localizações dos eventos, como explicado anteriormente. Quando a largura de banda aumenta (0.25), a estimativa de intensidade perde algumas características locais, obtendo valores muito parecidos nos eventos mais centrais. Este fato ocorre também devido à forma regular com que os eventos estão distribuídos. As diferenças dos valores da intensidade próximos à borda da figura, provavelmente ocorrem devido aos efeitos de borda, apesar dessa estimativa apresentar correção para efeitos de borda.

Nas Figuras 1.9 e 1.10, processo de cluster das localizações de mudas de sequóia, as diferenças também são claras. Quando a largura de banda é menor (0.05) a estimativa de intensidade tem valores diferentes de zero praticamente nas localizações dos eventos, sendo muito próxima de zero onde não há evento. Quando a largura de banda aumenta (0.25), a estimativa de intensidade se torna bem diferente, conseguindo identificar melhor as áreas de maior intensidade (com maior número de eventos). Este fato também ocorre devido à forma com que se distribuem os eventos nesta região: de forma agregada.

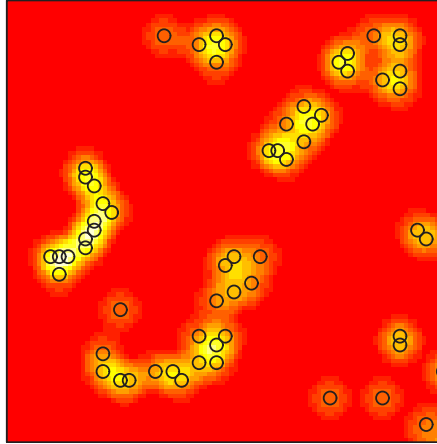


Figura 1.9: *Estimação por kernel nas localizações de 62 mudas de sequóia em um quadrado de lado 23 metros com largura de banda 0.05*

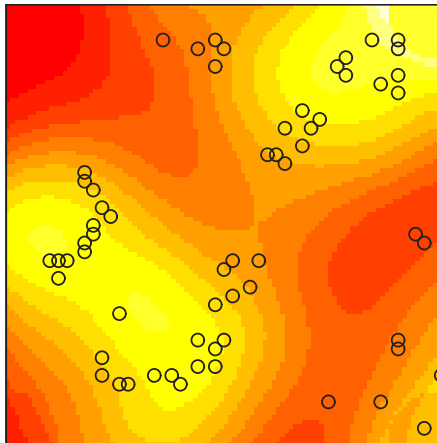


Figura 1.10: *Estimação por kernel nas localizações de 62 mudas de sequóia em um quadrado de lado 23 metros com largura de banda 0.25*

Capítulo 2

Teste de Escore

2.1 Introdução

O Teste da Razão de Verossimilhança, TRV, é muito utilizado, mas há circunstâncias em que outros testes podem ser preferíveis. Um destes testes é o teste de escore, que é localmente mais poderoso. Sua vantagem sobre o TRV está no fato de não requerer o estimador de máxima verossimilhança do parâmetro θ . A definição de testes localmente mais poderosos é apresentada na Seção 2.3. O Teste da Razão de Verossimilhança é apresentado na Seção 2.2 e o teste de escore na Seção 2.4. O teste para separabilidade, apresentado no Capítulo 4, é baseado nesta estatística.

2.2 Teste da Razão de Verossimilhança

Sejam X_1, X_2, \dots, X_n , uma amostra aleatória de uma distribuição com função densidade de probabilidade $f(x; \theta)$, e vamos supor que desejamos testar

$$H_0 : \theta \in \omega \quad vs \quad H_1 : \theta \in \Omega - \omega.$$

Definimos

$$\lambda = \{Max_{\theta \in \omega} L(\theta; x) / Max_{\theta \in \Omega} L(\theta; x)\}$$

e construímos um teste com região crítica da forma $\lambda \leq A$ para alguma constante A . Este é o Teste de Razão de Verossimilhança (TRV).

A forma deste teste é intuitivamente aceitável; temos $0 \leq \lambda \leq 1$, e λ seria próxima de 1 se $\theta \in \omega$, mas seria mais distante de 1 se $\theta \notin \omega$. Sua respeitabilidade é intuiti-

vamente acentuada por sua conexão com o método de estimação de Verossimilhança. Também quando H_0 e H_1 são ambas simples, o TRV se reduzirá ao melhor teste dado pelo Lema de Neyman-Pearson. Assim como a estatística de teste dada pelo Lema de Neyman-Pearson e os estimadores de Máxima Verossimilhança, λ depende de x apenas através da estatística suficiente e minimal para θ . Desta forma não é surpresa que testes baseados em λ tenham boas propriedades. Estas propriedades são geralmente assintóticas. Cox e Hinkley (1974), por exemplo, citam algumas:

1. O TRV é assintoticamente mais poderoso não-viciado (pág.319);
2. O TRV é assintoticamente similar (pág. 323)
3. O TRV é assintoticamente eficiente (pág. 337)

Um inconveniente do TRV está no fato de este requerer o cálculo do estimador de máxima verossimilhança de θ em Ω .

Distribuição da estatística de teste no TRV

Frequentemente λ não tem uma distribuição conhecida. Em tais casos, para n grande, há um resultado aproximado amplamente utilizado: sob H_0 , $-2\ln(\lambda) \sim \chi_d^2$, onde d é a diferença entre a dimensão de H_0 e $H_0 \cup H_1$, e onde a união inclui todos os valores possíveis de $\theta \in \Omega$. Sendo mais específico, se p_0 é o número de parâmetros estimados sob H_0 , menos o número de restrições relacionando estes parâmetros com outros, e p_1 está similarmente definido sob $H_0 \cup H_1$, então $d = p_1 - p_0$.

2.3 Testes Localmente mais Poderosos

Os testes *localmente* mais poderosos (LMP) são testes mais poderosos para alternativas próximas de H_0 . Restringimos nossa atenção para o caso em que H_0 é simples e nos restringimos também a um único parâmetro. Assim, consideramos $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_0 + \delta$ para algum δ pequeno, $\delta > 0$.

O melhor teste de H_0 vs. H_1 tem região crítica da forma

$$\frac{L(\theta_0 + \delta; \mathbf{x})}{L(\theta_0; \mathbf{x})} \geq A$$

ou, em termos da log-verossimilhança,

$$l(\theta_0 + \delta; \mathbf{x}) - l(\theta_0; \mathbf{x}) \geq \ln(A)$$

para uma constante A escolhida apropriadamente. Usando expansão de Taylor, e assumindo algumas condições de regularidade, temos, para δ pequeno,

$$l(\theta_0 + \delta; \mathbf{x}) \approx l(\theta_0; \mathbf{x}) + \delta \left(\frac{\partial l}{\partial \theta} \right)_{\theta=\theta_0}$$

Assim o teste localmente mais poderoso tem região crítica da forma

$$\left(\frac{\partial l}{\partial \theta} \right)_{\theta=\theta_0} \geq B \text{ para alguma constante } B.$$

Este teste é baseado na derivada primeira da log-verossimilhança (o escore) e portanto é um *teste de escore*. Sob certas condições de regularidade, temos

$$E \left[\frac{\partial l}{\partial \theta} \right] = 0 \quad \text{e} \quad \text{Var} \left[\frac{\partial l}{\partial \theta} \right] = I_\theta$$

onde

$$I_\theta = E \left[\left(\frac{\partial \ln[L(\theta; \mathbf{x})]}{\partial \theta} \right)^2 \right] = E \left[\left(\frac{\partial l}{\partial \theta} \right)^2 \right]$$

é conhecida como matriz de informação de Fisher. Além disso, para n grande, $\frac{\partial l}{\partial \theta}$ é aproximadamente normalmente distribuída.

2.4 Teste de Escore

O TRV é muito utilizado, mas há circunstâncias em que outros testes podem ser preferíveis. Um destes testes é o teste de score. Na Seção 2.3 vimos que o teste de score é localmente mais poderoso. A estatística de teste está baseada em $\frac{\partial l}{\partial \theta}$ quando θ é escalar, ou no vetor

$$\mathbf{u}(\boldsymbol{\theta}) = \left(\frac{\partial l}{\partial \theta_1}, \frac{\partial l}{\partial \theta_2}, \dots, \frac{\partial l}{\partial \theta_k} \right)$$

para o vetor $\boldsymbol{\theta}$. Sob certas condições de regularidade, se a hipótese nula é $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, com $H_1 : \boldsymbol{\theta} \in \boldsymbol{\Omega} - \{\boldsymbol{\theta}_0\}$, então a estatística de teste é escrita como $\mathbf{u}^T(\boldsymbol{\theta}_0) I_{\boldsymbol{\theta}_0}^{-1} \mathbf{u}(\boldsymbol{\theta}_0)$, onde $I_{\boldsymbol{\theta}_0}^{-1}$ é a inversa da matriz de informação de Fisher calculada em $\boldsymbol{\theta}_0$. Sob H_0 esta estatística tem assintoticamente uma distribuição χ^2 com k graus de liberdade,

e é assintoticamente equivalente ao correspondente TRV, embora as duas estatísticas sejam diferentes em amostras pequenas (Cox e Hinkley, 1974, Seção 9.3). O teste de escore também é conhecido como *teste de multiplicador de Lagrange*. Sua vantagem sobre o TRV está no fato de não requerer o estimador de máxima verossimilhança de θ (Garthwaite, Jolliffe e Jones, 1995).

Capítulo 3

Separabilidade

3.1 Introdução

Os modelos frequentemente utilizados para representar observações de eventos discretos tais como terremotos, queimadas, incidência de epidemias, etc, quase invariavelmente têm a intensidade na forma de produto de duas funções, uma envolvendo apenas o espaço e o tempo e a outra apenas as marcas. Neste caso, a intensidade é dita *separável*. Apesar da importância desta suposição, a separabilidade de tais processos raramente é rigorosamente examinada. Como exceções, Ogata (1988) usou modelos paramétricos de reescalonamento para observar alternativas de separabilidade no modelo sequencial de epidemias após choques (*epidemic-type aftershock sequence (ETAS) model*) para ocorrências de terremotos. Guttorp (1995) detectou não-separabilidade em dados de precipitação em Middletown, Pennsylvania. Vários autores investigaram mudanças espaço-temporais em distribuições de marcas (Ogata e Kastura, 1993; Johnson, 1992; Keeley, Fotheringham e Morais, 1999) e testes de separabilidade foram construídos para séries temporais e processos espaciais autoregressivos (Shitan e Brockwell, 1995), mas outros exemplos de testes para separabilidade em processos pontuais são evasivos.

Neste capítulo definimos separabilidade, apresentamos estimativas da intensidade (separável ou não) baseadas em estimação por kernel e apresentamos também métodos diretos para testar separabilidade, baseados na comparação das estimativas da intensidade.

3.2 Definição e Exemplos

Durante o período de tempo $[0, t_F]$ eventos aleatórios são observados em uma região finita d -dimensional A . À sua localização no espaço e tempo, os eventos têm associados marcas que, como visto na Seção 1.4, podem registrar qualquer característica de interesse. Para cada ocorrência de eventos como queimadas ou terremotos, por exemplo, pode ser identificado um ponto $(t, \mathbf{x}, m) \in \mathbb{R}^6$, onde t representa o tempo de origem do evento, \mathbf{x} corresponde a localização tri-dimensional e m a medida em valor real de seu tamanho. Sendo N um processo pontual, ou seja, uma coleção aleatória de pontos em um espaço métrico \mathcal{X} , sejam $\mathbf{x}_i = (x_{1i}, \dots, x_{di})$ as localizações, t_i os tempos de referência e m_i as marcas dos n eventos ocorrendo em uma região geográfica específica A , em um período de tempo $[0, t_F]$ e indexados por $i, i = 1, \dots, n$. Vamos assumir que eles são uma realização de um Processo de Poisson em uma região finita $\mathcal{X} = A \times [0, t_F] \times \mathbb{R}$ com função intensidade condicional denotada por $\lambda(t, \mathbf{x}, m | \mathcal{F}_t)$, (como Schoenberg, 2004), a taxa média de ocorrência de eventos por volume espaço-tempo-marca condicional sobre o passado (ver Artigo em Anexo). Denotamos por $E(N(B))$ o número esperado de eventos em um conjunto mensurável B contido em \mathcal{X} .

A log-verossimilhança dos eventos observados pode ser escrita como

$$l = \sum_{i=1}^n \log(\lambda(t_i, \mathbf{x}_i, m_i | \mathcal{F}_t)) - \int_{\mathcal{X}} \lambda(t, \mathbf{x}, m | \mathcal{F}_t) d\mathbf{x} dt dm$$

(Fishman e Snyder, 1976).

Dizemos que um processo pontual espaço-temporal é separável se a intensidade pode ser expressa como

$$\lambda(t, \mathbf{x}, m | \mathcal{F}_t) = \lambda_1(t, \mathbf{x} | \mathcal{F}_t) f(m), \quad (3.1)$$

onde λ_1 é um processo não negativo e f é uma função fixa não-negativa, t é o tempo de ocorrência dos eventos e \mathbf{x} o vetor das coordenadas espaciais e m as marcas. Quando 3.1 ocorre, chamamos o processo separável com respeito a marca m . Se, além disso, a intensidade pode ser mais reduzida para a forma

$$\lambda(t, \mathbf{x}, m | \mathcal{F}_t) = \lambda_1(t | \mathcal{F}_t) f_1(\mathbf{x}) f_2(m),$$

onde λ_1 é novamente um processo não-negativo e cada f_i é uma função fixa não-negativa, então o processo é completamente separável. Um exemplo de separabilidade completa é o processo pontual temporal marginal composto na página 336 de Rathbun (1993), designado para descrever germinação anual de plantas, em que as localizações dos eventos são modeladas como completamente independente dos tempos dos eventos: em tais casos, os parâmetros em cada processo marginal podem ser estimados individualmente.

Na maioria das aplicações a separabilidade completa raramente é assumida. Contudo, separabilidade é quase sempre implicitamente assumida em modelos para processo pontual espaço-temporal-marcado. Para facilitar a notação, vamos omitir a dependência condicional sobre o passado \mathcal{F}_t na intensidade, desta forma, nos referimos à intensidade escrevendo $\lambda(t, \mathbf{x}, m)$.

É importante observar que cluster em processos espaço-temporais marcados é diferente de separabilidade. Um processo pode ser agregado e ainda assim não ser uma violação de 3.1. Isto porque em alguns processos pode ocorrer cluster espaço-tempo e a distribuição das marcas permanecer inalterada. Quando as marcas são agregadas, isto é, quando espaço-tempo e marcas interagem em forma de cluster, então a existência de um ponto em (t, \mathbf{x}, m) pode mudar a distribuição das marcas, de forma que em um futuro próximo a probabilidade de um ponto ter uma marca próxima a m cresce, isto viola a condição de separabilidade. Processos em que t e m interagem em forma de cluster (respectivamente, inibição) também são exemplos de processo não-separáveis. Isto porque para o processo de cluster (inibição) a ocorrência de um ponto com marca m aumenta (diminui) a probabilidade de um ponto, ocorrendo em um futuro próximo, possuir marca próxima a m . Assim a distribuição das marcas em qualquer tempo t depende dos pontos anteriores, em violação de 3.1.

O Processo de Poisson Homogêneo, como aquele apresentado no Capítulo 1, Figura 1.3, é um processo separável. Neste exemplo a função intensidade pode ser escrita como $\lambda(t, \mathbf{x}, m) = 70 I_{\{(t,m) \in [0,1]^2\}} = 70 I_{\{t \in [0,1]\}} I_{\{m \in [0,1]\}}$. Outro exemplo é o processo com função intensidade $\lambda(t, \mathbf{x}, m) = e^{(2t+4m)} = e^{2t} e^{4m}$, também apresentado no Capítulo 1. A Figura 1.5 apresenta uma realização desse processo, sendo $x_1 = t$ e $x_2 = m$. O processo com função intensidade $\lambda(t, \mathbf{x}, m) = e^{(3+3t)} + e^{3+m}$ é um processo não-separável, desde que para $t \approx 0$ é mais provável que um ponto tenha

valor grande de m , enquanto que para $t \approx 1$ a distribuição de m fica próxima da uniforme (ver Figura 1.4, considerando $x_1 = t$ e $x_2 = m$). Outro processo não-separável é apresentado na Figura 1.6. Nesse processo com função intensidade $\lambda(t, \mathbf{x}, m) = 40 I_{\{(t,m) \in [0,0.5]^2\}} + 480 I_{\{(t,m) \in [0.75,1]^2\}}$, temos que para $t \approx 0$ não existem marcas maiores que 0.5 e para $t \approx 1$ as marcas possuem valores grandes ($m \approx 1$) e não existem marcas menores que 0.75.

3.3 Investigando a Hipótese de Separabilidade

Uma forma não-paramétrica para investigar a hipótese de separabilidade, é obter uma estimativa não-paramétrica (por exemplo, kernel) da intensidade espaço-tempo λ_1 , denotada por $\tilde{\lambda}_1(t, \mathbf{x})$ e outra $\tilde{f}(m)$ da intensidade das marcas f (Vere Jones, 1992; Brillinger, 1998). Isto é,

$$\tilde{\lambda}_1(t, \mathbf{x}) = \int_{\mathcal{X}} k_{d+1}(t - u, \mathbf{x} - \mathbf{y}) dN(u, \mathbf{y})$$

e

$$\tilde{f}(m) = \int_{\mathcal{X}} k_1(m - m^*) dN(m^*)$$

onde k_{d+1} e k_1 são densidades de kernel $d + 1$ -dimensional e unidimensional, respectivamente. A seguir, encontramos uma estimativa da intensidade condicional espaço-tempo-marca λ , não-paramétrica $\hat{\lambda}$ de $\lambda(t, \mathbf{x}, m)$, por exemplo estimativa por kernel

$$\hat{\lambda}(t, \mathbf{x}, m) = \int_{\mathcal{X}} k_{d+2}(t - u, \mathbf{x} - \mathbf{y}, m - m^*) dN(u, \mathbf{y}, m^*)$$

onde k_{d+2} é uma densidade de kernel $d+2$ -dimensional. Podemos, então comparar as estimativas de intensidade condicional resultantes $\hat{\lambda}(t, \mathbf{x}, m)$ e $\tilde{\lambda}(t, \mathbf{x}, m) = \tilde{\lambda}_1(t, \mathbf{x})\tilde{f}(m)/N(\mathcal{X})$. Assim, $\tilde{\lambda}$ é uma estimativa de intensidade condicional não-paramétrica separável, enquanto que $\hat{\lambda}$ pode não ser separável.

Muito tem sido escrito sobre selecionar densidades de kernel e larguras de banda e correção para efeito de borda. O interesse neste trabalho não é a construção de estimativas de intensidade não-paramétricas adequadas, mas como testar a separabilidade depois de tais estimativas serem obtidas. Portanto, não vamos lidar com o problema de efeito de borda.

3.4 Testes de Separabilidade

Sob a hipótese de separabilidade, as duas estimativas de intensidade $\widehat{\lambda}$ e $\widetilde{\lambda}$ seriam similares. Uma forma de comparar as duas estimativas de intensidade $\widehat{\lambda}$ e $\widetilde{\lambda}$, é encontrar o mínimo e o máximo (padronizado) da diferença absoluta entre as duas, isto é,

$$S_1 = \sup \left[|\widehat{\lambda}(t, \mathbf{x}, m) - \widetilde{\lambda}(t, \mathbf{x}, m)| / \sqrt{\widetilde{\lambda}(t, \mathbf{x}, m)}; (t, \mathbf{x}, m) \in \mathcal{X} \right]$$

$$S_2 = \inf \left[|\widehat{\lambda}(t, \mathbf{x}, m) - \widetilde{\lambda}(t, \mathbf{x}, m)| / \sqrt{\widetilde{\lambda}(t, \mathbf{x}, m)}; (t, \mathbf{x}, m) \in \mathcal{X} \right]$$

Outras opções são a estatística Cramer-von Mises-type,

$$S_3 = \int_{\mathcal{X}} \left[\widehat{\lambda}(t, \mathbf{x}, m) - \widetilde{\lambda}(t, \mathbf{x}, m) \right]^2 dt d\mathbf{x} dm$$

a diferença das log-verossimilhanças sob os dois modelos $\widehat{\lambda}$ e $\widetilde{\lambda}$

$$S_4 = \int_{\mathcal{X}} \left[\log \widehat{\lambda}(t, \mathbf{x}, m) - \log \widetilde{\lambda}(t, \mathbf{x}, m) \right] N(dt, d\mathbf{x}, dm) \\ - \int_{\mathcal{X}} \left[\widehat{\lambda}(t, \mathbf{x}, m) - \widetilde{\lambda}(t, \mathbf{x}, m) \right] d\mathbf{x} dt dm$$

e a quinta estatística de teste seria

$$S_5 = \sum_i \left(\widehat{\lambda}(t_i, \mathbf{x}_i, m_i) - \widetilde{\lambda}(t_i, \mathbf{x}_i, m_i) \right)^2$$

Valores muito grandes de qualquer destas estatísticas indicam evidências contra a hipótese de separabilidade (Schoenberg, 2004).

Capítulo 4

Um Teste de Escore para Separabilidade

4.1 Introdução

Schoenberg (2004) utilizou as estatísticas apresentadas no capítulo anterior para várias alternativas de separabilidade. São estatísticas que comparam as estimativas de intensidade separável e não-separável. No entanto, não há qualquer fundamento ou base teórica que explique o sucesso ou não de cada uma delas.

Anderson e Titterington (1997) investigaram o poder de estatísticas tais como a distância ao quadrado integrada e o máximo absoluto da diferença entre duas estimativas por kernel para identificar cluster na diferença entre dois conjuntos de dados de processos pontuais espaciais. Zimmerman (1993) estudou a diferença ao quadrado integrada entre a função de distribuição empírica de um processo pontual e de um processo de Poisson correspondente como um teste para cluster espacial, e comparou variações nesta estatística com medidas padrão de cluster incluindo estatísticas de distâncias entre vizinhos e as funções K e L de Ripley (1977).

Neste capítulo apresentamos o teste, baseado na estatística de escore, para testar a hipótese de separabilidade em processos pontuais espaço-temporais marcados. Para tal, introduzimos um método semi-paramétrico para modelar hipóteses alternativas de separabilidade, o qual permite obter a estatística de teste de escore. Apresentamos também semelhanças entre essa estatística e as apresentadas no capítulo anterior.

4.2 Definindo a Estatística de Teste

Seja

$$\mu = E(N(\mathcal{X})) = \int_{\mathcal{X}} \lambda(t, \mathbf{x} m) dt d\mathbf{x} dm$$

o número esperado de eventos em \mathcal{X} . Definimos também as funções baseadas nas intensidades marginais

$$\lambda_{ST}(t, \mathbf{x}) = \mu^{-1} \int_{\mathbb{R}} \lambda(t, \mathbf{x} m) dm$$

e

$$\lambda_M(m) = \mu^{-1} \int_{[0, t_F]} \int_A \lambda(t, \mathbf{x}, m) d\mathbf{x} dt .$$

Note que

$$1 = \int_{[0, t_F]} \int_A \lambda_{ST}(t, \mathbf{x}) d\mathbf{x} dt = \int_{\mathbb{R}} \lambda_M(m) dm$$

Sob a hipótese de separabilidade, temos

$$\lambda(t, \mathbf{x}, m) = \mu \lambda_{ST}(t, \mathbf{x}) \lambda_M(m) . \quad (4.1)$$

Se o processo for não-separável, existe uma constante ϵ e uma função $g(t, \mathbf{x}, m)$ tal que $\lambda(t, \mathbf{x})$ pode ser escrita como

$$\lambda(t, \mathbf{x}, m) = \mu \lambda_{ST}(t, \mathbf{x}) \lambda_M(m) (1 + \epsilon g(t, \mathbf{x}, m)) . \quad (4.2)$$

Sob a hipótese de separabilidade, temos $\epsilon g(t, \mathbf{x}, m) \equiv 0$. Queremos contrastar a hipótese nula de separabilidade (4.1) com a hipótese (4.2) onde espaço, tempo e marcas interagem. De particular interesse são os casos em que não há alternativas óbvias para a hipótese nula. Neste caso, vamos considerar ϵ suficientemente pequeno tal que (4.2) pode ser visto como uma alternativa local de (4.1).

Vamos assumir inicialmente que $\lambda_{ST}(t, \mathbf{x})$, $\lambda_M(m)$ e $g(t, \mathbf{x}, m)$ são funções conhecidas. Então, sob a hipótese alternativa, a log-verossimilhança do parâmetro ϵ é dada por

$$\begin{aligned} l(\epsilon) &= \sum_i (\log \mu \lambda_{ST}(t_i, \mathbf{x}_i) \lambda_M(m_i) + \log (1 + \epsilon g(t_i, \mathbf{x}_i, m_i))) \\ &\quad - \mu \int_{\mathcal{X}} \lambda_{ST}(t, \mathbf{x}) \lambda_M(m) d\mathbf{x} dt dm - \epsilon \mu \int_{\mathcal{X}} \lambda_{ST}(t, \mathbf{x}) \lambda_M(m) g(t, \mathbf{x}, m) d\mathbf{x} dt dm \end{aligned}$$

e portanto a estatística de escore é dada por

$$\frac{\partial l}{\partial \epsilon} = \sum_{i=1}^n \frac{g(t_i, \mathbf{x}_i, m_i)}{1 + \epsilon g(t_i, \mathbf{x}_i, m_i)} - \mu \int_{\mathcal{X}} \lambda_{ST}(t, \mathbf{x}) \lambda_M(m) g(t, \mathbf{x}, m) d\mathbf{x} dt dm \quad (4.3)$$

Quando avaliada em $\epsilon = 0$, (4.3) resulta na estatística de teste de escore

$$T = \frac{\partial l}{\partial \epsilon} \Big|_{\epsilon=0} = \sum_{i=1}^n g(t_i, \mathbf{x}_i, m_i) - \mu \int_{\mathcal{X}} \lambda_{ST}(t, \mathbf{x}) \lambda_M(m) g(t, \mathbf{x}, m) d\mathbf{x} dt dm . \quad (4.4)$$

Um teste baseado em T é o teste localmente mais poderoso no sentido que maximiza a derivada da função poder em $\epsilon = 0$ (Cox e Hinkley, 1974, pág 113). Contudo, um teste baseado em T não é viável pois, usualmente, as funções λ_{ST} , λ_M e g não são conhecidas. Uma forma de investigar a validade da hipótese nula é baseada em estimativas não-paramétricas (ex. kernel, spline, wavelet) das intensidades envolvidas. Sejam $\tilde{\lambda}_{ST}(t, \mathbf{x})$ e $\tilde{\lambda}_M(m)$ estimativas das intensidades espaço-tempo e de marcas, respectivamente, tais como, por exemplo, estimativas por kernel

$$\tilde{\lambda}_M(m) = \frac{1}{\tilde{\mu}} \int_{\mathcal{X}} k_M(m - m^*) dN(t, \mathbf{x}, m^*)$$

e

$$\tilde{\lambda}_{ST}(t, \mathbf{x}) = \frac{1}{\tilde{\mu}} \int_{\mathcal{X}} k_{ST}(t - u, \mathbf{x} - \mathbf{y}) dN(u, \mathbf{y}, m)$$

onde k_M e k_{ST} são as densidades de kernel das marcas e espaço-temporal, respectivamente e $\tilde{\mu} = N(\mathcal{X})$. Sob a hipótese nula (4.1), uma estimativa de $\lambda(t, \mathbf{x}, m)$ é dada pelo produto

$$\tilde{\lambda}(t, \mathbf{x}, m) = \tilde{\mu} \tilde{\lambda}_{ST}(t, \mathbf{x}) \tilde{\lambda}_M(m) \quad (4.5)$$

A seguir, obtemos uma estimativa da intensidade espaço-tempo-marcas sob o modelo alternativo (4.2), tal como estimativa por kernel:

$$\hat{\lambda}(t, \mathbf{x}, m) = \int_{\mathcal{X}} k(t - u, \mathbf{x} - \mathbf{y}, m - m^*) dN(u, \mathbf{y}, m^*) .$$

Como $g(t, \mathbf{x}, m)$ não é conhecida, notamos que

$$\epsilon g(t, \mathbf{x}, m) = \lambda(t, \mathbf{x}, m) / (\mu \lambda_{ST}(t, \mathbf{x}) \lambda_M(m)) - 1.$$

Portanto, sob o modelo de hipótese alternativa, uma estimativa de $\epsilon g(t, \mathbf{x}, m)$ é dada por $\hat{\lambda}(t, \mathbf{x}, m) / \tilde{\lambda}(t, \mathbf{x}, m) - 1$. Substituímos g (a menos da constante de proporcionalidade ϵ) em (4.4) por sua estimativa empírica para obter uma nova estatística

de teste não-paramétrica

$$\begin{aligned} T &= \sum_{i=1}^n \left(\frac{\widehat{\lambda}(t_i, \mathbf{x}_i, m_i)}{\widetilde{\lambda}(t_i, \mathbf{x}_i, m_i)} - 1 \right) - \int_{\mathcal{X}} \widetilde{\lambda}(t, \mathbf{x}, m) \left(\frac{\widehat{\lambda}(t, \mathbf{x}, m)}{\widetilde{\lambda}(t, \mathbf{x}, m)} - 1 \right) d\mathbf{x} dt dm \\ &= \sum_{i=1}^n \frac{\widehat{\lambda}}{\widetilde{\lambda}}(t_i, \mathbf{x}_i, m_i) - \int_{\mathcal{X}} \left[\widehat{\lambda}(t, \mathbf{x}, m) - \widetilde{\lambda}(t, \mathbf{x}, m) \right] d\mathbf{x} dt dm - n. \end{aligned}$$

T é, então, uma nova proposta para testar separabilidade em processos pontuais.

4.3 Relação entre as Estatísticas de Teste

Observamos que há uma semelhança entre as estatísticas S_4 e T . Realizando uma expansão de primeira ordem da função $\log(1+x)$, em série de Taylor em torno de 0, temos:

$$\log(1+x) = \log(1) + x + o(h)$$

Fazendo $1+x = \frac{\widehat{\lambda}}{\widetilde{\lambda}}$, a expressão acima fica:

$$\log\left(\frac{\widehat{\lambda}}{\widetilde{\lambda}}\right) = \left(\frac{\widehat{\lambda}}{\widetilde{\lambda}} - 1\right) + o(h)$$

Substituindo esta relação em S_4 , temos:

$$\begin{aligned} S_4 &\approx \sum_{i=1}^n \left(\frac{\widehat{\lambda}}{\widetilde{\lambda}} - 1 \right) - \int_{\mathcal{X}} \left[\widehat{\lambda}(t, \mathbf{x}, m) - \widetilde{\lambda}(t, \mathbf{x}, m) \right] d\mathbf{x} dt dm \\ &= \sum_{i=1}^n \frac{\widehat{\lambda}}{\widetilde{\lambda}}(t_i, \mathbf{x}_i, m_i) - \int_{\mathcal{X}} \left[\widehat{\lambda}(t, \mathbf{x}, m) - \widetilde{\lambda}(t, \mathbf{x}, m) \right] d\mathbf{x} dt dm - n. \end{aligned}$$

Obtemos, então, a estatística T (proposta neste trabalho) como uma expansão da estatística S_4 . Desta forma, sob separabilidade, as estatísticas S_4 e T seriam similares e sob alternativas de separabilidade, esperamos que o desempenho destas seja bem parecido, como poder, por exemplo.

Vamos fazer uma expansão de segunda ordem em série de Taylor em torno de 0 da função $\log(1+x)$:

$$\log(1+x) = \log(1) + x - \frac{x^2}{2} + o(h)$$

Fazendo $1 + x = \frac{\tilde{\lambda}}{\hat{\lambda}}$, a expressão acima fica:

$$\log\left(\frac{\tilde{\lambda}}{\hat{\lambda}}\right) = \left(\frac{\tilde{\lambda} - \hat{\lambda}}{\hat{\lambda}}\right) - \frac{1}{2}\left(\frac{\tilde{\lambda} - \hat{\lambda}}{\hat{\lambda}}\right)^2 + o(h) \quad (4.6)$$

Agora, pelo Teorema de Campbell (Stoyan e Stoyan, 1994), para uma função bem comportada f e uma função intensidade λ , temos

$$E\left[\sum_i f(t_i, \mathbf{x}_i, m_i)\right] = \int_{\mathcal{X}} f(t, \mathbf{x}, m) \lambda(t, \mathbf{x}, m) d\mathbf{x} dt dm \quad (4.7)$$

Dessa forma, tomando $f = \log(\hat{\lambda}/\tilde{\lambda}) = -\log(\tilde{\lambda}/\hat{\lambda})$ temos uma aproximação para o termo $\sum_i \log(\hat{\lambda}/\tilde{\lambda})(t_i, \mathbf{x}_i, m_i)$ em S_4 , e então, utilizando 4.6 e 4.7, temos:

$$\begin{aligned} S_4 &\approx \int_{\mathcal{X}} \left(\log\frac{\hat{\lambda}}{\tilde{\lambda}}\right) \lambda(t, \mathbf{x}, m) - \int_{\mathcal{X}} \left[\hat{\lambda}(t, \mathbf{x}, m) - \tilde{\lambda}(t, \mathbf{x}, m)\right] d\mathbf{x} dt dm \\ \Rightarrow -S_4 &\approx \int_{\mathcal{X}} \left(\log\frac{\tilde{\lambda}}{\hat{\lambda}}\right) \lambda(t, \mathbf{x}, m) - \int_{\mathcal{X}} \left[\tilde{\lambda}(t, \mathbf{x}, m) - \hat{\lambda}(t, \mathbf{x}, m)\right] d\mathbf{x} dt dm \\ &\approx \int_{\mathcal{X}} \left(\frac{\tilde{\lambda} - \hat{\lambda}}{\hat{\lambda}}\right) \lambda(t, \mathbf{x}, m) d\mathbf{x} dt dm - \frac{1}{2} \int_{\mathcal{X}} \left(\frac{\tilde{\lambda} - \hat{\lambda}}{\hat{\lambda}}\right)^2 \lambda(t, \mathbf{x}, m) d\mathbf{x} dt dm \\ &\quad - \int_{\mathcal{X}} \left[\tilde{\lambda}(t, \mathbf{x}, m) - \hat{\lambda}(t, \mathbf{x}, m)\right] d\mathbf{x} dt dm \end{aligned} \quad (4.8)$$

Como $\hat{\lambda}$ é uma estimativa não-paramétrica de λ , o primeiro e o terceiro termos de 4.8 são aproximadamente iguais, portanto se cancelam e o segundo termo, pelo mesmo motivo, pode ser simplificado de forma que obtemos:

$$S_4 \approx \frac{1}{2} \int_{\mathcal{X}} (\hat{\lambda} - \tilde{\lambda})^2 \frac{1}{\hat{\lambda}} d\mathbf{x} dt dm \quad (4.9)$$

O lado direito de 4.9 é uma estatística S_3 mas ponderando mais fortemente as regiões com baixa intensidade (pois λ está no denominador). Dessa forma temos, então

$$T \approx S_4 \approx \frac{1}{2} \int_{\mathcal{X}} (\hat{\lambda} - \tilde{\lambda})^2 \frac{1}{\tilde{\lambda}} \quad (4.10)$$

Utilizando novamente 4.7 agora na estatística S_5 , temos

$$S_5 = \frac{1}{n} \sum_i (\hat{\lambda}_i - \tilde{\lambda}_i)^2 \approx \frac{1}{n} \int_{\mathcal{X}} (\hat{\lambda} - \tilde{\lambda})^2 \lambda \quad (4.11)$$

As expressões 4.10 e 4.11 fornecem uma relação entre as estatísticas T , S_4 e S_5 e a estatística S_3 . Através dessas relações podemos esperar semelhanças quanto ao desempenho dessas estatísticas.

Vamos observar as diferenças entre as estatísticas tomando um processo não-separável considerado em Schoenberg (2004). Neste exemplo temos $\mathcal{X} = [0, 1]^4$ e a função intensidade dada por

$$\lambda(t, \mathbf{x}, m | \mathcal{F}_t) = \lambda(t, \mathbf{x}, m) = e^{3(1+t)} + e^{3+m} \quad (4.12)$$

Temos que $\mu = \int_{\mathcal{X}} \lambda(t, \mathbf{x}, m) = e^3(e^3/3 + e - 4/3)$,

$$\lambda_{ST}(t, \mathbf{x} | \mathcal{F}_t) = \lambda_{ST}(t, \mathbf{x}) = \frac{1}{\mu} \int_0^1 \lambda(t, \mathbf{x}, m) dm = \frac{1}{\mu} (e^{3(1+t)} + e^3(e - 1))$$

e

$$\lambda_M(m) = \frac{1}{\mu} \int_{[0,1]^3} \lambda(t, \mathbf{x}, m) d\mathbf{x} dt = \frac{1}{\mu} \left(e^{3+m} + \frac{e^3}{3}(e^3 - 1) \right).$$

Na parte superior da Figura 4.1 apresentamos a superfície $\lambda(t, \mathbf{x}, m)$ no lado esquerdo e $s(t, \mathbf{x}, m) = \mu \lambda_{ST}(\mathbf{x}, t) \lambda_M(m)$ no lado direito. O eixo z é o mesmo em ambos os gráficos. O primeiro gráfico à esquerda, na parte inferior da figura, mostra a superfície $(\lambda(\mathbf{x}, t, m) - s(\mathbf{x}, t, m))^2$ e o volume sob esta superfície é estimado por S_3 . As regiões que contribuem mais para a integral S_3 são aquelas em que marcas e tempo são próximos de seus valores extremos. As contribuições são consideravelmente maiores quando as marcas (e a intensidade condicional) são próximas a seus valores máximos. O gráfico na parte central mostra a superfície $(\lambda(\mathbf{x}, t, m) - s(\mathbf{x}, t, m))^2 / \lambda(\mathbf{x}, t, m)$, associada a S_4 e T . As regiões em que marcas ou tempo são extremos contribuem mais para o volume. Em contraste com o gráfico anterior, esta integral é substancialmente afetada pelas regiões com baixa intensidade. O gráfico à direita mostra a superfície $(\lambda(\mathbf{x}, t, m) - s(\mathbf{x}, t, m))^2 \lambda(\mathbf{x}, t, m)$, associada a S_5 . É similar ao primeiro mas quase não dá peso nas regiões com baixa intensidade, onde as marcas têm valores baixos. Assim, parece que as diferenças entre as estatísticas de teste ocorrem devido à forma com que avaliam as marcas e o tempo em regiões de valores extremos.

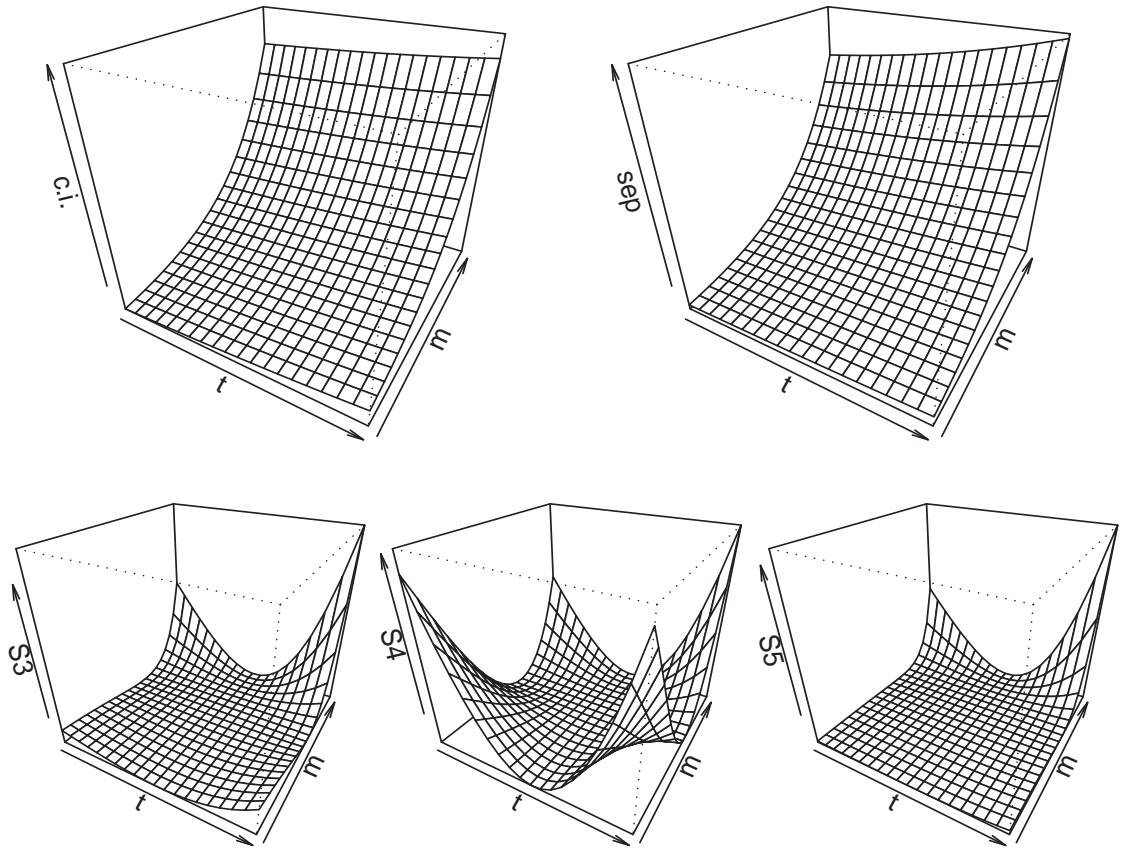


Figura 4.1: *Modelo exponencial aditivo não-separável.* Os gráficos na parte superior são as superfícies da intensidade condicional $\lambda = \lambda(t, \mathbf{x}, m)$ (esquerda) e o produto das intensidades marginais $s = s(t, \mathbf{x}, m)$ (direita). Da esquerda para a direita, na parte inferior $(\lambda - s)^2$, $(\lambda - s)^2/\lambda$, e $(\lambda - s)^2 \lambda$.

Capítulo 5

Simulação

5.1 Introdução

Schoenberg (2004) avaliou o desempenho das estatísticas apresentadas na Seção 3.4 sob várias alternativas de separabilidade, utilizando procedimento de Monte Carlo. Neste trabalho, adotamos um procedimento de Monte Carlo mais simples e ligeiramente diferente daquele adotado em Schoenberg (2004).

Os processos foram simulados no programa R, utilizando os pacotes de estatística espacial `splancs` e `spatstat`. Diggle e Rowlingson (1993) e Diggle e Rowlingson (1996) auxiliaram na elaboração dos programas para simulação. Para a intensidade, utilizamos estimativas de kernel, mais especificamente o kernel quártico e, como Schoenberg (2004), utilizamos larguras de banda de 0.05 em cada coordenada. Desta forma, poderíamos comparar os resultados obtidos neste trabalho com aqueles apresentados em Schoenberg (2004). Também esclarecemos que todas as estimativas têm correção para efeito de borda, realizada pelas funções de estatística espacial no programa R.

A seguir explicamos como as simulações são realizadas e apresentamos vários processos não separáveis para os quais as estatísticas são avaliadas. Muitos dos processos apresentados também foram utilizados por Schoenberg (2004).

5.2 Realizando os Testes

O poder do teste T , baseado na estatística de escore, em comparação com os testes da Seção 3.4, pode ser investigado sob várias alternativas para (4.1). Por exemplo, uma alternativa seria que a interação entre t e m fosse aditiva em vez de multiplicativa, isto é, $\lambda(t, \mathbf{x}, m) = \lambda_1(t, \mathbf{x}) + f(m)$. Para determinar qual estatística de teste parece ser mais sensível a um tipo específico de alternativa, geramos 1000 processos não separáveis (que dependem de ϵ) e, em cada processo, obtemos o valor das estatísticas utilizando as estimativas de kernel separável $\tilde{\lambda}$ e não separável $\hat{\lambda}$, realizamos 99 permutações aleatórias das marcas dos eventos originais, para cada um dos 1000 processos e, em cada permutação, obtemos o valor das estatísticas. O p -valor exato do teste de permutação é a proporção destes processos simulados (mais o processo realmente observado) cujo valor da estatística de teste é maior ou igual ao valor da mesma no processo observado. Assim, em cada um dos 1000 processos simulados, avaliamos se a hipótese de separabilidade é rejeitada. Para $\epsilon = 0$ os processos são separáveis e à medida que ϵ cresce estes se afastam cada vez mais da hipótese de separabilidade. Desta forma (variando ϵ), construímos a curva poder para cada uma das 6 estatísticas.

Investigamos o desempenho das estatísticas sob várias alternativas de separabilidade. Em todos os processos simulados tomamos como domínio o quadrado unitário, ou seja, o conjunto $\{(t, m) : (t, m) \in [0, 1] \times [0, 1]\}$.

Um exemplo é o processo de Poisson com intensidade não-separável dada por $\lambda(t, \mathbf{x}, m) = 100 + 100\epsilon I_{\{(t, m) \in [0.5, 0.75]^2\}}$, $\epsilon > 0$. Este processo gera 100 pontos uniformemente em $[0, 1]^2$ e $\lfloor 100\epsilon/16 \rfloor$ pontos em $[0.5, 0.75]^2$. Este processo é claramente não-separável, desde que para $t \approx 0$ a distribuição de m é uniforme e para $t \in [0.5, 0.75]^2$ a distribuição de m já não é mais uniforme.

Observando as figuras a seguir, podemos entender melhor as diferenças que ocorrem nesse processo à medida que ϵ cresce. Apresentamos duas figuras: na Figura 5.1, com $\epsilon = 0$, temos um processo separável (100 pontos distribuídos uniformemente em $[0, 1] \times [0, 1]$) e na Figura 5.2 temos um processo não separável com 100 pontos distribuídos uniformemente em $[0, 1] \times [0, 1]$ e 100 pontos distribuídos uniformemente em $[0.5, 0.75] \times [0.5, 0.75]$.

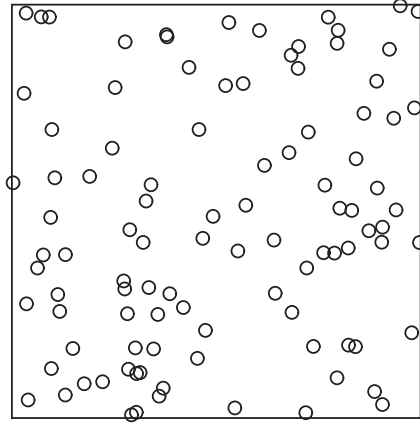


Figura 5.1: *Processo de Poisson Homogêneo com 100 pontos no quadrado unitário.*

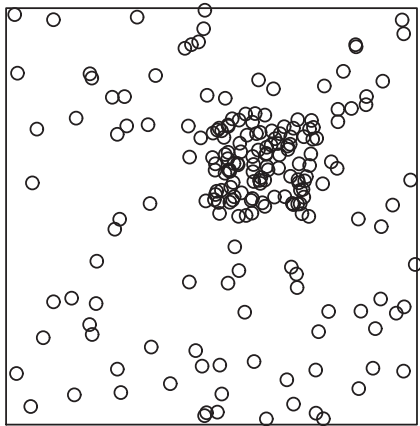


Figura 5.2: *Processo de Poisson Não-Homogêneo que gera 100 pontos uniformemente em $[0, 1] \times [0, 1]$ e 100 pontos em $[0.5, 0.75] \times [0.5, 0.75]$.*

Avaliamos, então o poder das estatísticas de teste sob essa alternativa de separabilidade, variando ϵ entre 0 e 16 (pois a intensidade do processo tem a forma $\lambda(t, \mathbf{x}, m) = 100 + 100\epsilon I_{\{(t,m) \in [0.5,0.75]^2\}}$) com passo 2 e obtemos a curva poder para cada uma das estatísticas.

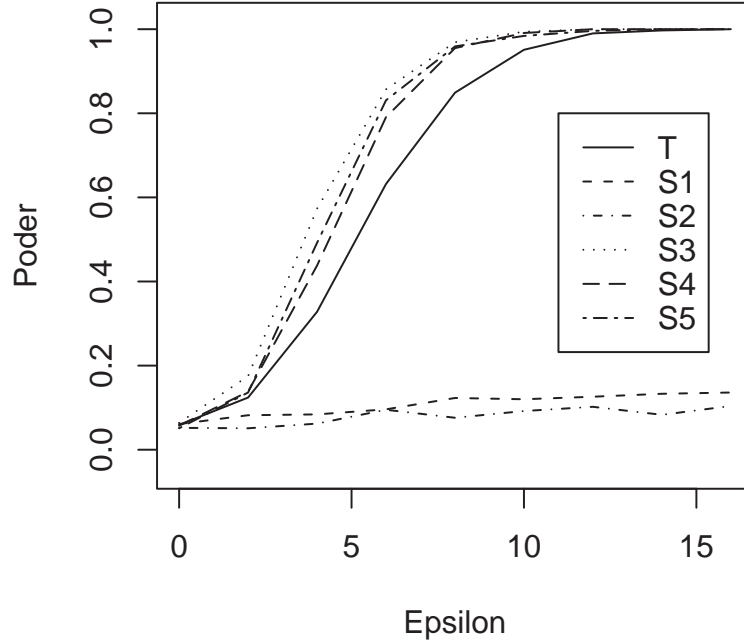


Figura 5.3: *Curvas poder para as estatísticas $S_1 - S_5$ e T em 1000 realizações do processo $\lambda(t, \mathbf{x}, m) = 100 + 100\epsilon I_{\{(t,m) \in [0.5,0.75]^2\}}$.*

Observamos que as estatísticas S_3 , S_4 , e S_5 e T são mais poderosas para identificar esta alternativa de separabilidade e se comportam de maneira semelhante, sendo as estatísticas S_3 , S_4 , S_5 muito próximas, seguidas por T , e S_3 ligeiramente melhor. S_1 e S_2 não conseguem identificar esse tipo de alternativa, apresentando valores de poder extremamente baixos.

Outros exemplos de processos com modelo de intensidade não-separável são:

(a) processo que gera 4 centros de clusters aleatoriamente em $[0, 1]^2$ e distribui, ao redor de cada centro de cluster, 49 pontos em que cada componente segue uma distribuição normal com média no centro de cluster e desvio padrão $1/\epsilon$. À medida que ϵ cresce, clusters mais agregados vão surgindo e, portanto, mais distante da hipótese de separabilidade o processo se torna. Novamente, para $\epsilon \approx 0$ o processo

é aproximadamente separável e, à medida que ϵ cresce, se afasta dessa hipótese. As figuras seguintes apresentam este processo com o menor valor de ϵ avaliado (0.1), ou seja, mais próximo da hipótese de separabilidade, e com o maior valor de ϵ avaliado (16.67), mais distante da separabilidade.

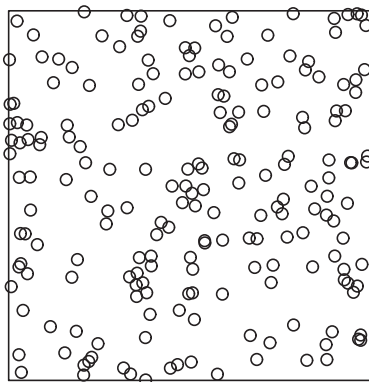


Figura 5.4: *Processo que gera 4 centros de clusters e distribui, ao redor de cada centro de cluster, 49 pontos segundo uma distribuição normal com $1/\sigma = \epsilon = 0.1$.*

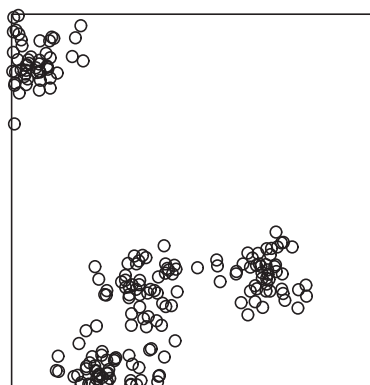


Figura 5.5: *Processo que gera 4 centros de clusters e distribui, ao redor de cada centro de cluster, 49 pontos segundo uma distribuição normal com $1/\sigma = \epsilon = 16.67$.*

Assim, avaliamos o desempenho dessas estatísticas sob os seguintes valores de ϵ : $\{0.1, 1, 2, 5, 8.33, 10, 14.29, 16.67\}$, os quais produziram os seguintes valores de desvio padrão para a distribuição normal bivariada: $\{10.00, 1.00, 0.50, 0.20, 0.12, 0.10, 0.07, 0.06\}$.

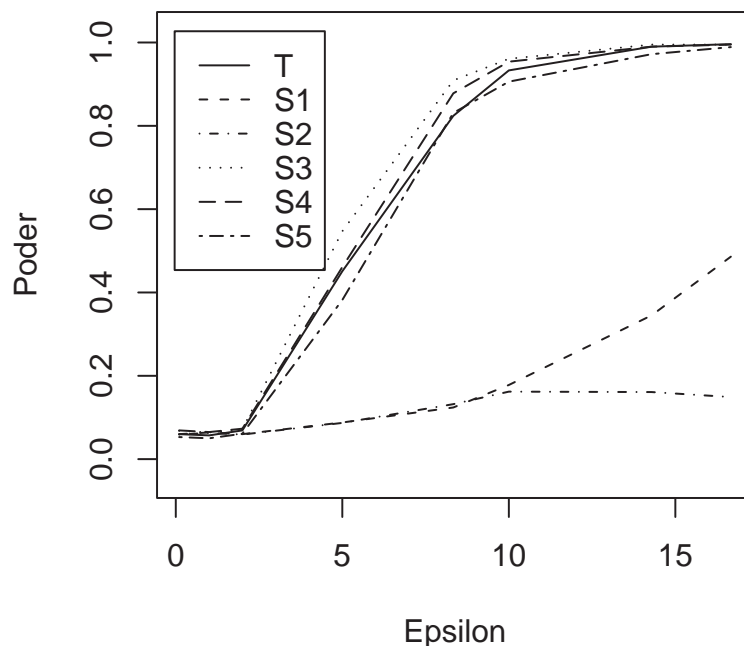


Figura 5.6: *Curvas poder para as estatísticas $S_1 - S_5$ e T em 1000 realizações do processo que gera 4 clusters aleatoriamente.*

Novamente as estatísticas S_3 , S_4 , S_5 e T se destacam como mais poderosas para identificar esta alternativa de separabilidade e se comportam de maneira ainda mais semelhante, sendo a estatística T ligeiramente superada por S_3 e S_4 , e S_5 ligeiramente superada por T . S_1 e S_2 novamente falham nesta identificação.

(b) processo que gera 100 pontos distribuídos uniformemente em $[0, 1]^2$ e 100 pontos distribuídos segundo uma normal bivariada em que cada componente tem média $\mu = 0.5$ e desvio padrão $\sigma = 1/\epsilon$. Para $\epsilon \approx 0$ temos 200 pontos distribuídos uniformemente em $[0, 1]^2$ mas, à medida que ϵ cresce, surge um cluster de tamanho 100, afastando desta forma, da hipótese de separabilidade. Apresentamos as figuras seguintes com duas realizações deste processo. Na primeira, mais próxima da separabilidade, geramos o processo para $\epsilon = 0.1$ e na segunda, mais afastada dessa hipótese, geramos o processo para $\epsilon = 16.67$.

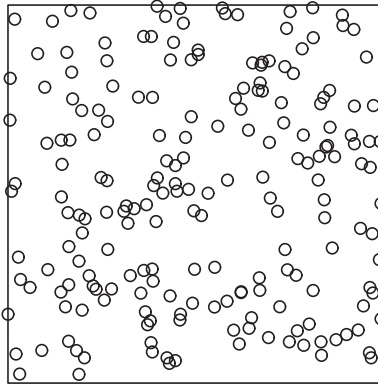


Figura 5.7: *Processo que gera 100 pontos distribuídos uniformemente e 100 pontos distribuídos segundo uma normal bivariada com $1/\sigma = \epsilon = 0.1$.*

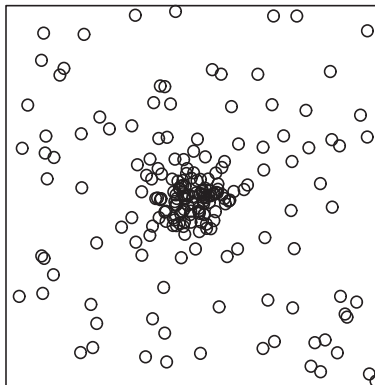


Figura 5.8: *Processo que gera 100 pontos distribuídos uniformemente e 100 pontos distribuídos segundo uma normal bivariada com $1/\sigma = \epsilon = 16.67$.*

Avaliamos o desempenho dessas estatísticas sob os seguintes valores de ϵ : $\{0.1, 1, 2, 5, 8.33, 10, 14.29, 16.67\}$, os quais produziram os seguintes valores de desvio padrão para a distribuição normal bivariada: $\{10.00, 1.00, 0.50, 0.20, 0.12, 0.10, 0.07, 0.06\}$.

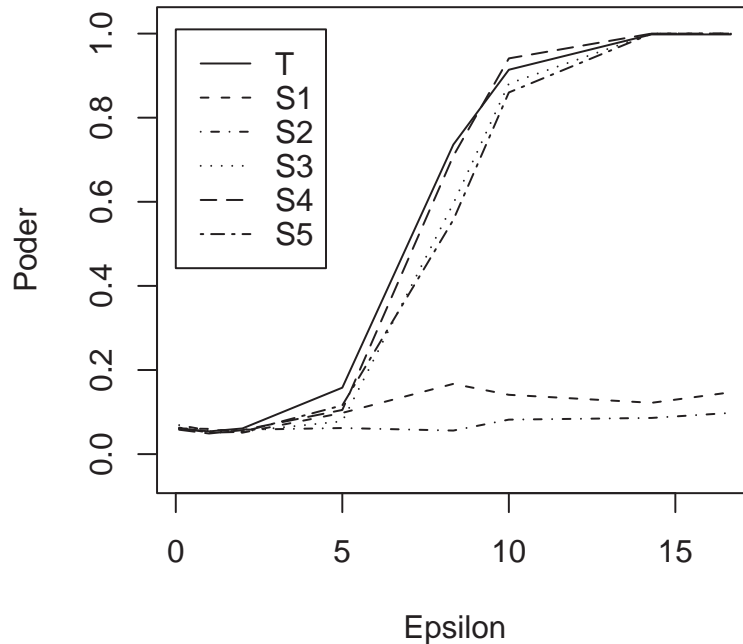


Figura 5.9: *Curvas poder para as estatísticas $S_1 - S_5$ e T em 1000 realizações do processo que gera 100 pontos distribuídos uniformemente em $[0, 1]^2$ e 100 pontos distribuídos segundo uma normal bivariada.*

Novamente as estatísticas S_3, S_4, S_5 e T têm comportamentos muito semelhantes. Neste processo observamos que a estatística T supera ligeiramente as três anteriores. Essas 4 estatísticas apresentam desempenho muito bom, conseguindo identificar essa alternativa de separabilidade, enquanto que S_1 e S_2 novamente falham nesta identificação.

Também avaliamos o poder destas estatísticas em outros processos, tais como os processos de cluster Neyman-Scott, Thomas e Matern e os processos de inibição MaternI, MaternII e SSI (Simple Sequential Inhibition), utilizados também em Schoenberg (2004). Descrevemos brevemente cada um destes processos, sendo encontrados mais detalhes na Seção 8.5 de Cressie (1993). O processo de cluster Neyman-Scott gera centros de cluster segundo um processo de Poisson com média α e distribui

um número fixo k de pontos em um círculo de raio $1/\epsilon$ ao redor de cada centro de cluster, uniformemente e independentemente. Nesta simulação foram utilizados $k = 3$ e $\alpha = 20$. Apresentamos duas figuras em que o Processo de Neyman-Scott foi gerado utilizando $\epsilon = 0.67$ e $\epsilon = 20$, respectivamente. Na primeira o processo é aproximadamente separável e na segunda se afasta claramente dessa hipótese.

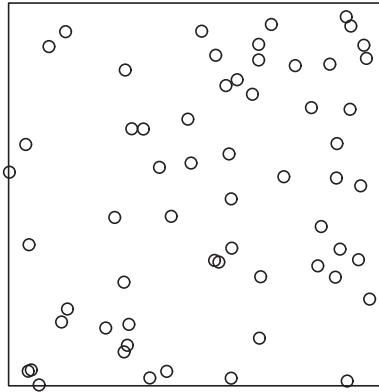


Figura 5.10: *Processo de cluster Neyman-Scott com $\alpha = 20$, $k = 3$ e $\epsilon = 0.67$.*

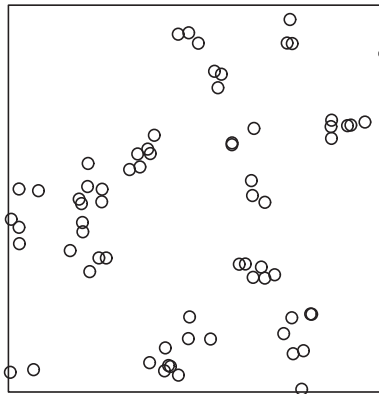


Figura 5.11: *Processo de cluster Neyman-Scott com $\alpha = 20$, $k = 3$ e $\epsilon = 20$.*

Avaliamos o desempenho das estatísticas para essa alternativa de separabilidade variando o raio dos clusters gerados. Os valores de ϵ utilizados foram $\{0.67, 1, 2, 5, 6.67, 10, 14.29, 20\}$ os quais produziram os seguintes valores para raio dos clusters: $\{1.5, 1, 0.50, 0.20, 0.15, 0.10, 0.07, 0.05\}$. Realizando as simulações para esse processo, obtemos a curva poder para cada estatística.

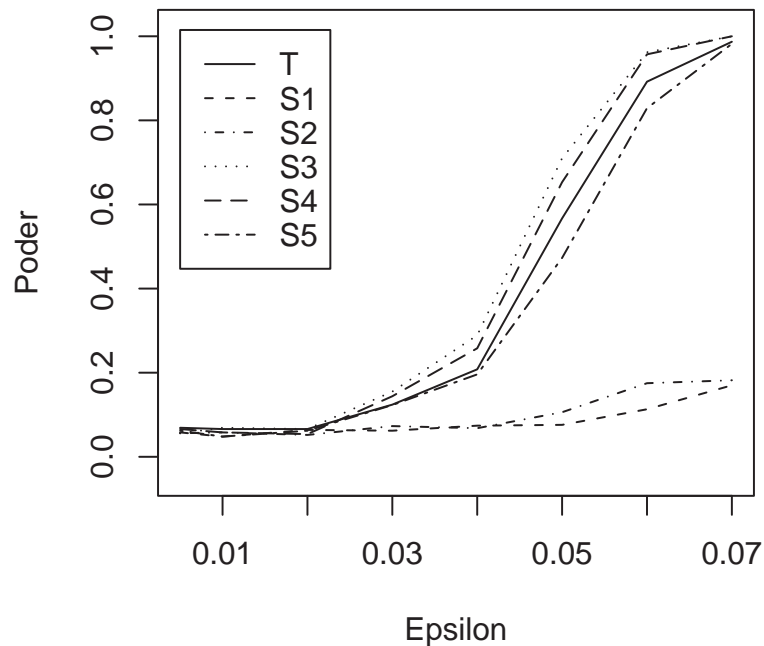


Figura 5.12: *Curvas poder para as estatísticas $S_1 - S_5$ e T em 1000 realizações do processo de cluster Neyman-Scott.*

O processo de cluster Thomas gera centros de cluster segundo um processo de Poisson com média α e o número de pontos em cada cluster segue uma distribuição de Poisson com média β . Os pontos são distribuídos segundo uma distribuição normal com média em cada centro de cluster e desvio padrão $\sigma = 1/\epsilon$. Nesta simulação foram utilizados $\alpha = 20$ e $\beta = 3$. Nas figuras seguintes geramos dois processos de cluster Thomas para os valores de ϵ : 0.67 e 25, os quais produziram valores de desvio padrão para a distribuição normal $\sigma = 1.5$ e $\sigma = 0.04$, respectivamente. O primeiro processo é aproximadamente separável, enquanto o segundo claramente se afasta dessa hipótese.

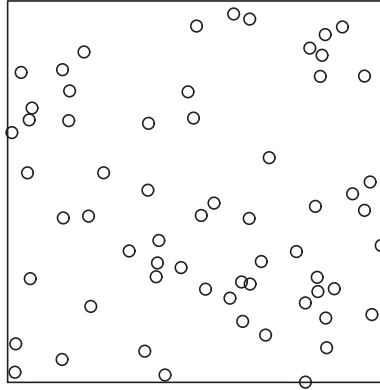


Figura 5.13: *Processo de cluster Thomas com $\alpha = 20$, $\beta = 3$ e $\epsilon = 0.67$.*

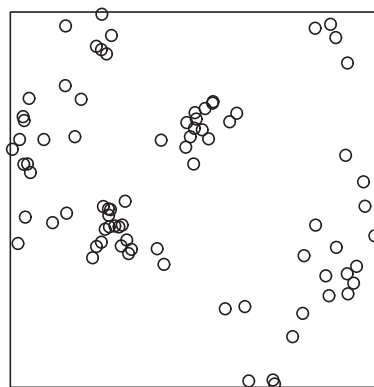


Figura 5.14: *Processo de cluster Thomas com $\alpha = 20$, $\beta = 3$ e $\epsilon = 25$.*

Avaliamos o desempenho das estatísticas para os seguintes valores de ϵ : $\{0.67, 1, 2, 5, 6.67, 10, 14.29, 25\}$, os quais produziram os valores de desvio padrão para a distribuição normal: $\{1.5, 1, 0.50, 0.20, 0.15, 0.10, 0.07, 0.04\}$. Desta forma obtemos a curva poder para cada estatística, apresentada abaixo:

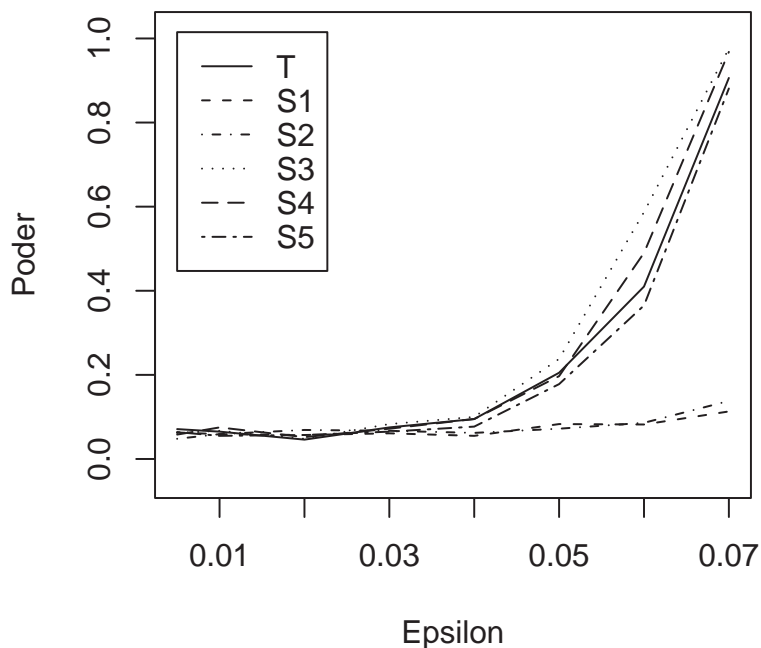


Figura 5.15: *Curvas poder para as estatísticas $S_1 - S_5$ e T em 1000 realizações do processo de cluster Thomas.*

O processo de cluster Matern é semelhante ao processo de cluster Thomas, com a diferença que no processo Matern, os pontos são distribuídos uniformemente e independentemente em um círculo de raio $1/\epsilon$ ao redor de cada centro de cluster. Também foram utilizados $\alpha = 20$ e $\beta = 3$. Geramos dois processos de cluster Matern, os quais são apresentados abaixo. No primeiro, aproximadamente separável, utilizamos $\epsilon = 0.67$, produzindo como raio do círculo o valor 1.5. No segundo, processo não separável, utilizamos $\epsilon = 20$, produzindo como raio do círculo o valor 0.05.

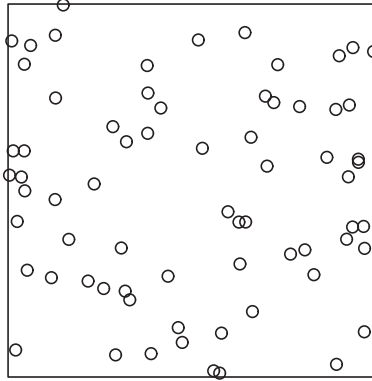


Figura 5.16: *Processo de cluster Matern com $\alpha = 20$, $\beta = 3$ e $\epsilon = 0.67$.*

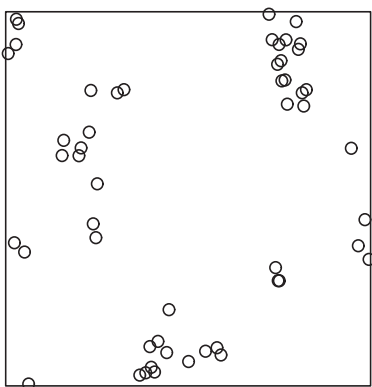


Figura 5.17: *Processo de cluster Matern com $\alpha = 20$, $\beta = 3$ e $\epsilon = 20$.*

Avaliamos o desempenho das estatísticas quanto a seu poder variando ϵ nos valores $\{0.67, 1, 2, 5, 6.67, 10, 14.29, 20\}$, os quais produziram os seguintes valores para raio do círculo $\{1.5, 1, 0.50, 0.20, 0.15, 0.10, 0.07, 0.05\}$.

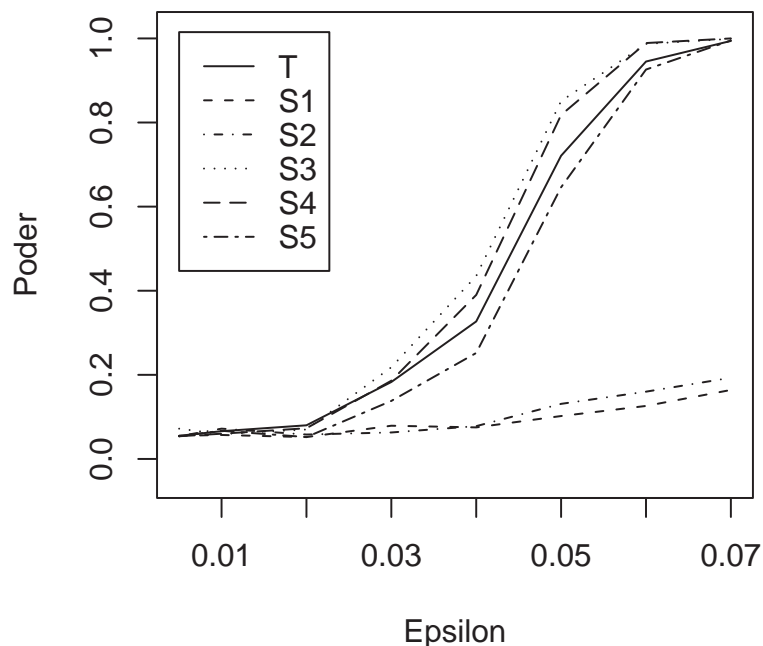


Figura 5.18: *Curvas poder para as estatísticas $S_1 - S_5$ e T em 1000 realizações do processo de cluster Matern.*

Nestes três processos de cluster o comportamento das 6 estatísticas quase não se altera. A estatística S_3 supera ligeiramente S_4 , T e S_5 , sendo que essas 4 estatísticas permanecem muito próximas. As 4 estatísticas apresentam bom desempenho na identificação dessas três alternativas de separabilidade. S_1 e S_2 novamente falham nesta identificação, apresentando poder inexpressivo.

O processo de inibição MaternI gera pontos segundo um processo de Poisson com média α e deleta todos os pares de pontos que estão a distâncias menores que ϵ unidades. Os pontos remanescentes formam o processo de inibição MaternI. Nesta simulação foi utilizado $\alpha = 70$. Apresentamos duas figuras em que o processo de inibição MaternI foi gerado para os valores de distância de inibição $\epsilon = 0.005$ e

$\epsilon = 0.07$. O primeiro valor gera um processo aproximadamente separável enquanto que o segundo não.

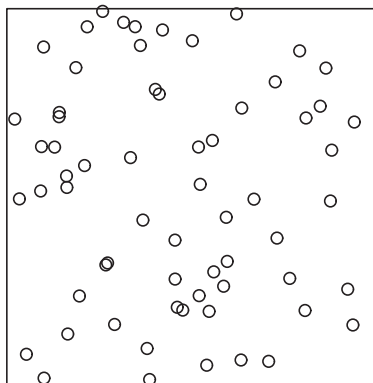


Figura 5.19: *Processo de inibição MaternI com $\alpha = 70$, e $\epsilon = 0.005$.*

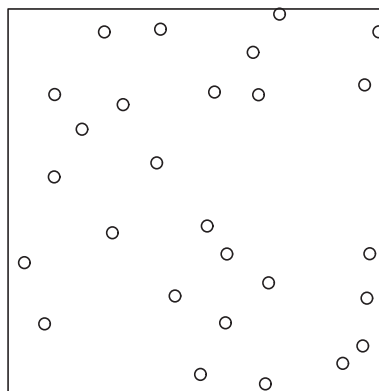


Figura 5.20: *Processo de inibição MaternI com $\alpha = 70$, e $\epsilon = 0.07$.*

Realizamos as simulações para a construção da curva poder tomando ϵ no seguinte conjunto de valores: $\{0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07\}$. Cada valor deste conjunto corresponde à menor distância existente entre os eventos nesse processo.

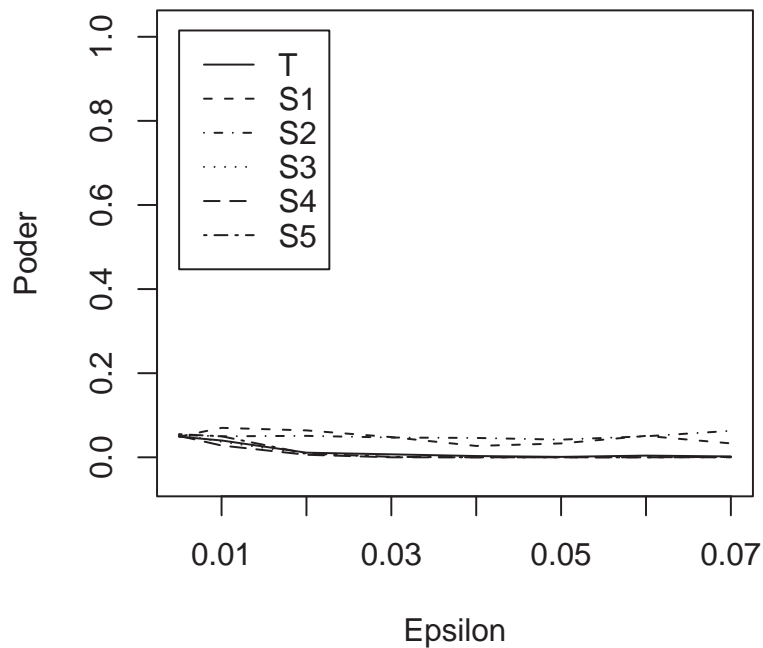


Figura 5.21: *Curvas poder para as estatísticas $S_1 - S_5$ e T em 1000 realizações do processo de inibição MaternI.*

O processo de inibição MaternII é semelhante ao processo anterior com a diferença que os pontos são indexados aleatoriamente e um ponto é deletado se ele está a uma distância menor que ϵ unidades dos outros pontos com índices inferiores. Na simulação foi utilizado $\alpha = 50$. Apresentamos duas figuras nas quais geramos o processo de inibição MaternII. Em uma utilizamos uma pequena distância de inibição entre os eventos $\epsilon = 0.005$ (processo aproximadamente separável), e em outra utilizamos uma distância maior 0.07 (processo não separável).

Realizamos as simulações desse processo variando ϵ no conjunto de valores $\{0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07\}$.

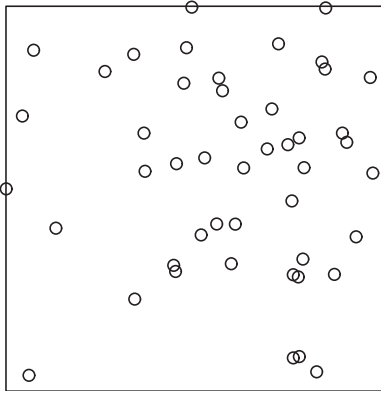


Figura 5.22: *Processo de inibição MaternII com $\alpha = 50$, e $\epsilon = 0.005$.*

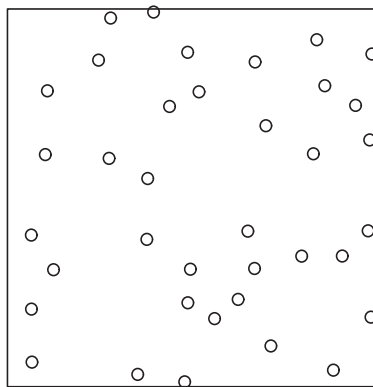


Figura 5.23: *Processo de inibição MaternII com $\alpha = 50$, e $\epsilon = 0.07$.*

No processo de inibição SSI (Simple Sequential Inhibition), os pontos são adicionados um a um. Cada novo ponto é gerado uniformemente e independentemente dos pontos anteriores. Se o novo ponto está a uma distância inferior a ϵ unidades de outro ponto existente é rejeitado, e um novo ponto é gerado. Este processo termina quando o número desejado de pontos é alcançado ou quando a configuração dos pontos não pode ser alterada. Na simulação utilizamos $k = 20$. Apresentamos a seguir duas figuras em que o processo de inibição SSI foi gerado. Na primeira, com processo aproximadamente separável, utilizamos $\epsilon = 0.005$ e na segunda, processo não separável, utilizamos $\epsilon = 0.07$.

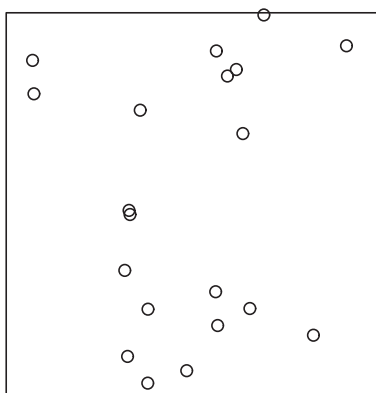


Figura 5.24: *Processo de inibição SSI com $k = 20$, e $\epsilon = 0.005$.*

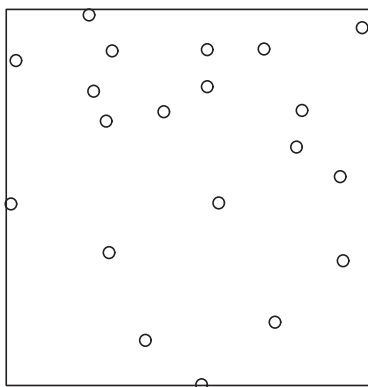


Figura 5.25: *Processo de inibição SSI com $k = 20$, e $\epsilon = 0.07$.*

Avaliamos o desempenho de cada estatística de teste para identificar o processo de inibição SSI como alternativa de separabilidade, variando ϵ no seguinte conjunto de valores, para a distância de inibição: $\{0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07\}$.

Para os três processos de inibição as 6 estatísticas têm comportamento muito semelhante, desta forma não apresentamos os gráficos das curvas poder para os processos MaternII e SSI. Nenhuma estatística consegue identificar este tipo de alternativa de separabilidade, sendo todas as curvas bem próximas de 0. Estas são as circunstâncias em que o processo de reescalonamento aleatório do processo é indicado por Schoenberg (2004).

Conclusões Finais

Apresentamos o teste baseado na estatística de escore como proposta para testar separabilidade em processos pontuais multi-dimensionais. Através das relações obtidas entre as estatísticas, no Capítulo 4, a semelhança entre essas quanto a seu desempenho foi confirmada nas simulações realizadas.

Comparações entre os testes apresentados por Schoenberg (2004) e o teste baseado na estatística de escore e proposto neste trabalho, foram realizadas para várias alternativas de separabilidade. Para alternativas graduais, as estatísticas S_3 , S_4 , S_5 e T apresentaram bom desempenho e pouca diferença entre as curvas poder. A estatística S_3 apresentou maior poder em relação às anteriores, apesar da pequena diferença entre essas, como mostrado em Schoenberg (2004). S_1 e S_2 não conseguiram identificar esse tipo de alternativa, se destacando em relação ao baixo poder.

Para alternativas como cluster, novamente S_3 , S_4 , S_5 e T apresentaram bom desempenho e muita semelhança entre as curvas poder. A estatística S_3 apresenta maior poder em relação às anteriores, apesar da pequena diferença entre essas, como mostrado em Schoenberg (2004). S_1 e S_2 novamente não conseguiram identificar esse tipo de alternativa.

Para alternativas de separabilidade em forma de processos de inibição, todas as 6 estatísticas apresentaram pouco poder, falhando nesta identificação. Como Schoenberg (2004) propôs, para tais casos, o método de reescalonamento aleatório do processo seria o mais indicado para a identificação deste tipo de alternativa.

Diante dos resultados, apesar da estatística proposta neste trabalho ser capaz de identificar, com sucesso, alternativas graduais de separabilidade e mesmo alternativas em forma de cluster, em processos pontuais multi-dimensionais, a estatística S_3 apresentou melhor desempenho na maioria dos processos avaliados.

Schoenberg (2004) apresenta uma aplicação em dados reais, onde o estudo indica que a hipótese de separabilidade é invalidada em virtude de ser uma hipótese forte para os dados de queimadas em Los Angeles. Como este, vários outros modelos em processos pontuais multi-dimensionais utilizados em estudos por exemplo de terremotos, assumem essa hipótese. Desta forma, em casos onde há suspeita de alternativas graduais seria razoável a utilização da estatística S_3 para essa identificação, mas se o pesquisador suspeita da presença de inibição ou cluster, o mais indicado seria o processo de reescalonamento aleatório do processo, descrito em Schoenberg (2004).

Referências Bibliográficas

- [1] Anderson, N. H. and Titterington, D. M. (1997) Some methods for investigating spatial clustering, with epidemiological applications. *JRSS-A*, 160, 7-105.
- [2] Bailey, T. C. and Gatrell, A. C. (1995) *Interactive Spatial Data Analysis*, 2nd ed. Longman Scientific and Technical, N. York.
- [3] Bhattacharya, R. N. and Waymire, E. C. (1990) *Stochastic Processes with Applications*. New York: Wiley.
- [4] Brillinger, D.R. (1998) Some wavelet analyses of point process data. In *31st Asilomar Conference on Signals, Systems and Computers*, 1087-1091. IEEE Computer Society. California: Los Alamitos.
- [5] Cox, D.R. and Hinkley, D.V. (1974) *Theoretical Statistics*. Chapman and Hall: London.
- [6] Cressie, N.A. (1993) *Statistics for spatial data, Revised edition*. New York: Wiley.
- [7] Daley, D.J. and Vere-jones, D.(1988) *An Introduction to the Theory of Point Processes*. Springer-Verlag, New York.
- [8] Diggle, P.J. (1983) *Statistical Analysis of Spatial Point Patterns*. Academic Press: London.
- [9] Diggle, P.J. and Rowlingson, B.S. (1993)- *Splancs: Spatial Point Pattern Analysis Code in S-Plus*. Lancaster University, Lancaster, U.K.
- [10] Diggle, P.J. and Rowlingson, B.S. (1996)- *Splancs Supplement- Spatial and Spatial-Temporal Analysis*. Lancaster University, Lancaster, U.K.

- [11] Diggle, P.J. (2003) *Statistical Analysis of Spatial Point Patterns*, 2nd ed. Oxford University Press Inc. N. York.
- [12] Fishman, P. M. and Snyder, D. L. (1976) The statistical analysis of space-time point processes. *IEEE Transactions on Information Theory* IT-22, 257-274.
- [13] Garthwaite, P. H., Jolliffe, I. T., and Jones, B. (1995) *Statistical Inference*. Prentice Hall. U.K.
- [14] Guttorp, P. (1995) *Stochastic Modeling of Scientific Data*. London: Chapman and Hall.
- [15] Johnson, E. (1992) *Fire and Vegetation Dynamics: Studies from the North American Boreal Forest*. Cambridge: Cambridge University Press.
- [16] Karlin, S. (1975) *A First Course in Stochastic Processes*. Academic Press: New York.
- [17] Keeley, J., Fotheringham, C., and Morais, M. (1999) Reexamining fire suppression impacts on brushland fire regimes. *Science*, 284, 1829-1832.
- [18] Milne, R. K. (2001) Point Processes and Some Related Processes. *Handbook of Statistics*, 19, 599-641.
- [19] Ogata, Y. (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of American Statistical Association*, 83, 9-27.
- [20] Ogata, Y. and Katsura, K. (1993) Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogs. *Geophysical Journal International*, 113, 727-738.
- [21] Rathbun, S. L. (1993) Modeling marked spatio-temporal point patterns. *Bulletin of International Statistical Institute*, 55, 379-396.
- [22] Ripley, B. D. (1977) Modelling spatial patterns. *JRSS-B*, 39, 172-192.
- [23] Schoenberg, F.P., Brillinger, D.R. and Guttorp, P.M. (2002) In *Encyclopedia of Environmetrics*, El-Shaarawi and W. Piegorisch(eds), Volume 3, 1573-1577. New York: Wiley.

- [24] Schoenberg, F.P. (2004) Testing separability in multi-dimensional point processes. *Biometrics*, 60, 471-481.
- [25] Shitan, M. and Brockwell, P.J. (1995) An asymptotic test for separability of a spatial autoregressive model. *Communications Statistics Theory Mathematics*, 24, 2027-2040.
- [26] Snyder, D. L. and Miller, M. I. (1991) *Random Point Processes in Time and Space*. Springer-Verlag, New York.
- [27] Stoyan, D. and Stoyan, H. (1994) *Fractals, Random Shapes and Point Fields*. Wiley, Chichester.
- [28] Strauss, D.J. (1975) A model for clustering. *Biometrika*, 63, 467-475.
- [29] Vere-Jones, D. (1992) Statistical Methods for the Description and Display of Earthquake Catalogs. In *Statistics in Environmental and Earth Sciences*, A. Walden and P. Guttorp (eds), 220-236. London: Edward Arnold.
- [30] Zimmerman, D. L. (1993) A bivariate Cram'er-von Mises type of test for spatial randomness. In *Applied Statistics*, 42, 43-54.