

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

**Inferência Bayesiana Exata para Processos de Cox**  
*Level-set*

**Bárbara da Costa Campos Dias**

**Belo Horizonte**

**2019**

**Bárbara da Costa Campos Dias**

**Inferência Bayesiana Exata para Processos de Cox**  
*Level-set*

Tese de doutorado apresentada ao Programa de Pós-graduação em Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito para obtenção do título de Doutora em Estatística.

**Orientador:** Prof. Doutor Flávio Bamberra Gonçalves

**Belo Horizonte**

**2019**

© 2019, Bárbara da Costa Campos Dias.  
Todos os direitos reservados

**Ficha catalográfica elaborada pela bibliotecária Belkiz Inez  
Rezende Costa CRB 6 1510**

Dias, Bárbara da Costa Campos.

D541i Inferência bayesiana exata para processos de Cox  
level-set / Bárbara da Costa Campos Dias — Belo  
Horizonte, 2019.  
70 f.: il.; 29 cm.

Tese (doutorado) - Universidade Federal de  
Minas Gerais – Departamento de Estatística.

Orientador: Flávio Bamberra Gonçalves.

1. Estatística. - Teses. 2. Teoria bayesiana de decisão  
estatística. 3. Markov, Processos de – Teses. 4  
Processos gaussianos – Teses. Orientador. II. Título.  
CDU 519.2(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

UFMG

## ATA DA DEFESA DE TESE DA ALUNA BÁRBARA DA COSTA CAMPOS DIAS

Realizou-se, no dia 03 de dezembro de 2019, às 16:00 horas, 2025 ICEX, da Universidade Federal de Minas Gerais, a 57ª defesa de tese, intitulada *Inferência Bayesiana Exata para Processos de Cox Level-set*, apresentada por BÁRBARA DA COSTA CAMPOS DIAS, número de registro 2015716798, graduada no curso de ESTATÍSTICA, como requisito parcial para a obtenção do grau de Doutor em ESTATÍSTICA, à seguinte Comissão Examinadora: Prof(a). Flávio Bamberga Gonçalves - Orientador (DEST/UFMG), Prof(a). Dani Gamerman (DEST/UFMG), Prof(a). Marcos Oliveira Prates (DEST/UFMG), Prof(a). Rafael Izbicki (UFSCar), Prof(a). Daiane Aparecida Zuanetti (UFSCar).

A Comissão considerou a tese:

Aprovada

Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

Belo Horizonte, 03 de dezembro de 2019.

Prof(a). Flávio Bamberga Gonçalves (Doutor)

Prof(a). Dani Gamerman (Doutor)

Prof(a). Marcos Oliveira Prates (Doutor)

Prof(a). Rafael Izbicki (Doutor)

Prof(a). Daiane Aparecida Zuanetti (Doutor)

AOS MEUS PAIS, À MINHA IRMÃ E  
AO MEU AMOR FELIPE.

# Agradecimentos

Primeiramente sou grata à Deus, minha fé foi essencial para seguir em frente.

À minha família amada, por não medir esforços pra me ver bem e por vibrarem junto comigo essa conquista.

Ao meu marido Felipe, por tanto companheirismo e paciência durante esses anos, tudo ficou mais fácil com você ao meu lado.

Agradeço ao meu orientador Flávio Bambirra Gonçalves, por me ensinar tanto, por tanta generosidade em compartilhar todo o seu imenso conhecimento.

Aos amigos que fiz na UFMG, obrigada por tornar essa caminhada mais leve, por tantas experiências e conhecimentos compartilhados. Em especial meus agradecimentos à Juliana, Fernanda, Ana Gabriela, Gabi, toda a turma que ingressou comigo no mestrado e aos amigos do LESTE.

Agradeço aos excelentes professores e aos funcionários do departamento de Estatística da UFMG.

À Capes, pelo auxílio financeiro.

Por fim, agradeço à todos que de alguma forma fizeram parte dessa importante jornada.

# Resumo

Esta tese propõe uma família de processos de Cox multidimensionais com função intensidade constante por partes e uma metodologia Bayesiana exata para se fazer inferência estatística nessa família. A família proposta é baseada no modelo Bayesiano Level-set proposto por Dunlop et al. [2016]. A motivação para se propor esses processos é o fato de processos de Cox com intensidade constante por partes serem adequados para modelar diversos fenômenos e, por serem mais simples, proporcionarem uma menor variabilidade quando comparado a modelos em que a intensidade varia continuamente no espaço. Uma função de ajuste de nível, determinada por um processo Gaussiano latente, determina de forma flexível as regiões do espaço que apresentam intensidades constantes. Apesar da não tratabilidade da função de verossimilhança e da dimensionalidade infinita do problema, a metodologia de inferência proposta é baseada em um algoritmo MCMC (Markov chain Monte Carlo) que converge para a distribuição a posteriori exata da função intensidade e outras componentes latentes do modelo. Isso significa que não existe erro de discretização envolvido, como nas metodologias existentes na literatura para abordar o mesmo problema. O algoritmo MCMC utiliza técnicas recentes de simulação estocástica, como *Pseudo-Marginal Metropolis* e *Poisson estimator*. Finalmente, a metodologia proposta é utilizada em exemplos simulados e reais para demonstrar sua eficiência e aplicabilidade.

**Palavras-chave:** Processos de Cox multidimensionais, inferência Bayesiana Exata, processo Gaussiano, Pseudo-Marginal Metropolis.

# Abstract

This thesis proposes a novel family of multidimensional Cox processes with piece-wise constant intensity function and an exact Bayesian approach to perform statistical inference in this family. This family is based on the Bayesian Level-set model proposed by Dunlop et al. [2016] and is motivated by the fact that such processes may be efficient to model a variety of point process phenomena. Furthermore, due to its simpler form when compared to continuously varying intensity functions, it is expected to provide more precise results. A level set function depends on a latent Gaussian process to flexibly determine the regions of the space with constant intensities. Despite the intractability of the likelihood function and infinite dimensionality of the parameter space, the proposed methodology does not resort to discrete approximations of the space (unlike competing methodologies in the literature) and Monte Carlo is the only source of inaccuracy. This arises from an MCMC algorithm that converges to the exact posterior distribution of all the unknown quantities in the model. The MCMC algorithm relies on recent stochastic simulation techniques, such as Pseudo-Marginal Metropolis and Poisson estimator. Finally simulated and real examples are presented to demonstrate the efficiency and applicability of the proposed methodology.

**Keywords:** Multidimensional Cox processes, exact Bayesian approach, Gaussian process, Pseudo-Marginal Metropolis.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>9</b>
<b>2</b>	<b>Revisão da teoria</b>	<b>14</b>
2.1	Processo de Poisson . . . . .	14
2.2	Processo de Cox . . . . .	15
2.3	Processo Gaussiano . . . . .	15
2.4	Algoritmo Metropolis–Hastings . . . . .	16
2.5	Verossimilhança intratável e amostragem retrospectiva . . . . .	17
2.5.1	Algoritmo <i>Pseudo-Marginal Metropolis-Hastings</i> . . . . .	18
2.5.2	Algoritmo de Barker . . . . .	19
<b>3</b>	<b>Modelo Bayesiano para processos de Cox Level-Set espaciais</b>	<b>22</b>
3.1	Inferência para o modelo proposto . . . . .	24
3.2	<i>Pseudo-Marginal</i> para o modelo proposto e o <i>Poisson Estimator</i> . . . . .	26
3.2.1	Relação do <i>Poisson Estimator</i> com a constante $\delta$ . . . . .	27

<b>4</b>	<b>MCMC</b>	<b>30</b>
4.1	Amostragem do $N^*$ . . . . .	31
4.2	Amostragem dos $\lambda$ 's via <i>Pseudo-Marginal</i> . . . . .	32
4.3	Amostragem dos $\lambda$ 's via Barker . . . . .	33
4.4	Amostragem do processo Gaussiano . . . . .	34
4.5	Amostragem do $c$ . . . . .	35
4.6	Aspectos computacionais . . . . .	36
<b>5</b>	<b>Estudos de simulação</b>	<b>40</b>
5.1	Estudo de simulação 1 . . . . .	41
5.2	Estudo de simulação 2 . . . . .	43
5.3	Estudo de simulação 3 . . . . .	45
5.4	Estudo de simulação 4 . . . . .	46
<b>6</b>	<b>Aplicação</b>	<b>48</b>
6.1	Aplicação 1 . . . . .	48
6.1.1	Comparação da qualidade da predição entre a metodologia proposta e um modelo em que a FI varia continuamente . . . . .	50
6.2	Aplicação 2 . . . . .	52
6.3	Aplicação 3 . . . . .	54
<b>7</b>	<b>Considerações Finais</b>	<b>57</b>

# Capítulo 1

## Introdução

Processos pontuais modelam padrões de pontos distribuídos aleatoriamente em algum espaço. Na maioria dos casos cada ponto representa o tempo e/ou a localização de algum evento de interesse, por exemplo, as localizações de crimes ou ocorrências de uma certa doença em um determinado local. O espaço pode ser bastante geral (reta real, plano cartesiano ou espaços mais abstratos).

O processo pontual mais utilizado na literatura estatística é o Processo de Poisson (PP), que é definido de tal modo que, condicionado em uma função intensidade (FI), o número de pontos em qualquer região segue uma distribuição de Poisson e a distribuição de pontos em regiões disjuntas são independentes.

Uma extensão interessante do Processo de Poisson é o processo de Cox, que é um processo de Poisson não-homogêneo onde a função intensidade é modelada estocasticamente. Processos de Cox também são chamados de processos de Poisson duplamente estocástico. Existe uma variedade de trabalhos que exploram este tipo de modelo, por exemplo Møller et al. [1998] propuseram o popular processo de Cox log-Gaussiano, no qual um processo Gaussiano é usado para modelar o logaritmo da função intensidade (ver também Diggle [2013]). Porém, fazer inferência estatística neste tipo de processo é um problema complexo devido à dimensionalidade infinita do PP e da FI e à intratabilidade analítica da função de verossimilhança.

Por este motivo já existem diversos trabalhos na literatura com o intuito de resolver este tipo problema e o uso de aproximações baseadas em discretizações é um dos métodos mais utilizados (ver, por exemplo, Møller et al. [1998], Brix and Diggle [2001]). No entanto, o uso de aproximações discretas apresentam algumas desvantagens e devem ser utilizados com cuidado (Simpson et al. [2016]). Os principais problemas envolvidos neste tipo de abordagem são a produção de resultados tendenciosos e a difícil quantificação do erro envolvido. Além disso, o nível de discretização necessário para obter bons resultados é desconhecido e pode envolver um custo computacional impraticável. Por fim, aproximações discretas podem ocasionar uma descaracterização do modelo, comprometendo as propriedades desejadas do modelo exato.

Estas limitações motivaram o desenvolvimento de soluções exatas de inferência em processos de Cox, onde a exatidão refere-se ao fato de que os métodos estão livres de erro de discretização e a única fonte de imprecisão é o erro de Monte Carlo. Por exemplo, Gonçalves and Gamerman [2018] propuseram uma metodologia exata de inferência em processos de Cox espaço-temporal, onde a função intensidade depende de um processo Gaussiano multivariado, a metodologia é baseada na construção de um MCMC que amostra da distribuição *a posteriori* conjunta dos parâmetros e de variáveis latentes introduzidas no modelo. Outros exemplos de metodologias exatas em classes de processos de Cox podem ser encontrados em Gonçalves et al. [2017b] e Rao and Teh [2013]. O primeiro propõem uma função intensidade que depende de um processo de difusão e o segundo propõem uma função intensidade que é uma Cadeia de Markov em tempo contínuo.

O objetivo deste trabalho é propor uma família de processos de Cox multidimensionais com função intensidade constante por partes e uma metodologia Bayesiana exata para se fazer inferência estatística nessa família. Essa estrutura é obtida através de processos Gaussianos em uma estrutura conhecida como *Bayesian level-set models* (Dunlop et al. [2016]) que, através de uma função de ajuste de nível, determina de forma flexível interfaces entre regiões geométricas. A motivação para esta proposta é que, na prática, podem existir problemas onde a função intensidade é constante em pedaços no espaço, porém, com essas partes definidas

de forma irregular. Além disso, essa estrutura é mais simples que modelos nos quais a FI varia continuamente no espaço e, portanto, deveriam fornecer inferências mais precisas.

Hildeman et al. [2018] também considera um modelo de Cox com FI constante por partes mas, apesar de usar a ideia dos modelos *level set*, considera na verdade a versão discretizada deste modelo, baseada em uma discretização do espaço em um grade regular. Desta forma, as observações são resumidas no número de pontos em cada sub-região da grade. Apesar de mostrarem que a distribuição *a posteriori* referente à versão discretizada do modelo converge para a distribuição *a posteriori* referente ao modelo contínuo em variação total, os autores não apresentam cotas superiores para o erro de aproximação, mas apenas uma cota inferior -  $\mathcal{O}(h^{1/4})$ , para a taxa de convergência da distância de variação total onde  $h$  representa o comprimento do lado das sub-regiões quadradas.

## Motivações práticas

Vários problemas reais motivam a construção da metodologia proposta neste trabalho. Por exemplo, observando estimações de funções intensidades contínuas em aplicações realizadas por outros trabalhos, podemos observar um comportamento que nos leva a acreditar que a função intensidade possa ser satisfatoriamente modelada de forma a ser constante por partes. Neste caso, espera-se que, por ser mais simples, a metodologia proposta obtenha estimações mais precisas. Por exemplo, Gonçalves and Gamerman [2018] analisaram um conjunto de dados (disponível no pacote R *spatstat*) referente à localização de árvores da espécie *White oak* em Lansing Woods, EUA. Para a estimação da função intensidade utilizaram uma metodologia Bayesiana para realização da inferência em processos de Cox onde a função intensidade depende de um processo Gaussiano multivariado. A Figura 1.1 mostra a estimação da função, indicando que um modelo com intensidade constantes em 2 ou 3 partes poderia fornecer um bom ajuste dos dados.

Um outro exemplo, considerado em Frade [2014], consiste em árvores da espécie *Vochysia Tucanorum* localizadas na estação ecológica de Assis no município de Assis, SP. Frade [2014]

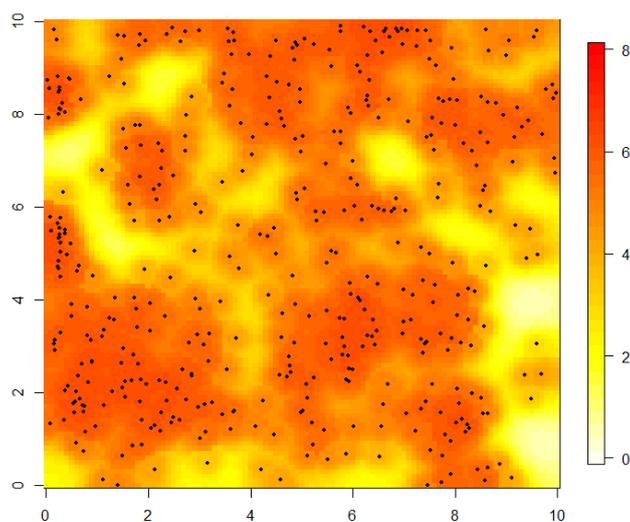


Figura 1.1: Mapa da média *a posteriori* da função intensidade na aplicação realizada por Gonçalves and Gamerman [2018] (*White oak* in Lansing Woods, EUA).

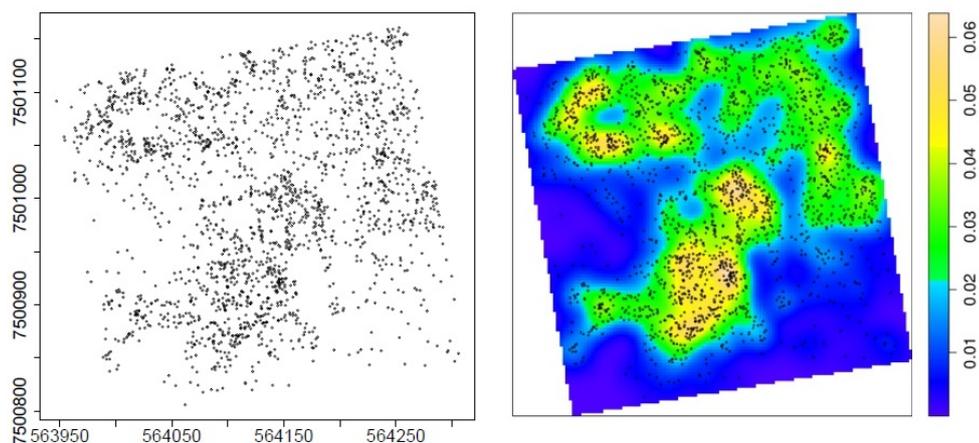


Figura 1.2: Distribuição de árvores da espécie **Vochysia Tucanorum** por meio de sua-  
vização de Kernel na aplicação realizada por Frade [2014].

analisa os dados utilizando estimação via Kernel para gerar estimativas apresentada na Figura 1.2, que fornece indícios de que a função intensidade pode ser constante em 3 ou 4 partes. Silva [2012] estudou a correlação entre casos de óbitos por neoplasias (câncer) relacionados à radiação eletromagnética e as localizações de fontes de campos eletromagnéticos notificados

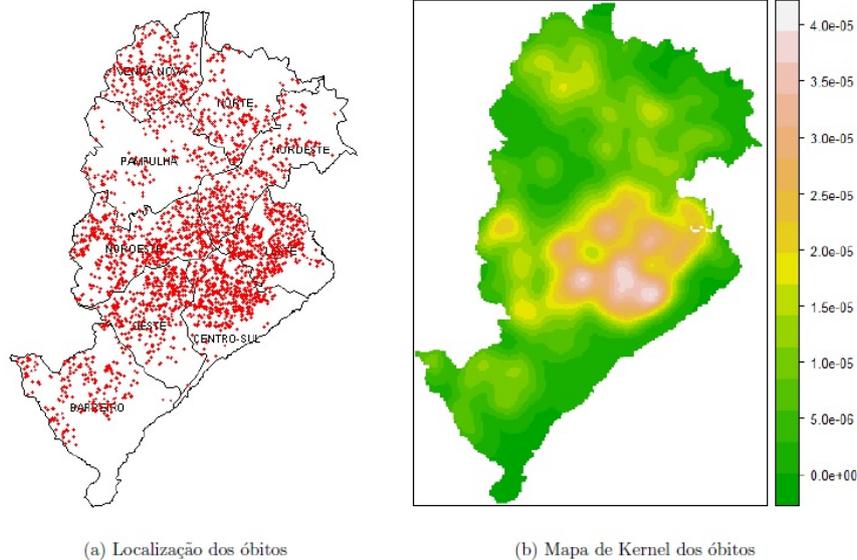


Figura 1.3: Distribuição dos casos de óbitos por neoplasias por meio de suavização de Kernel na aplicação realizada por Silva [2012].

no município de Belo Horizonte no período de 2000 a 2006. A Figura 1.3 apresenta as localizações dos óbitos (residência dos indivíduos) e o mapa de kernel correspondente, indicando que a função intensidade poderia ser constante em 4 ou 5 partes.

Finalmente, além de se esperar uma precisão maior da metodologia proposta quando comparada a modelos onde a FI varia continuamente com um PG, é razoável esperar que, em algumas aplicações, o pesquisador tenha uma motivação prática para estimar uma FI constante por partes, definindo assim uma partição do espaço analisado.

# Capítulo 2

## Revisão da teoria

### 2.1 Processo de Poisson

Processos de Poisson são amplamente utilizados para modelar padrões de pontos localizados aleatoriamente em algum espaço  $S$ . Neste trabalho iremos nos restringir ao contexto em que  $S = \mathbb{R}^d, d \geq 2$ .

**Definição 1** (Processo de Poisson em  $\mathbb{R}^d$ , Kingman [1993]). *Seja  $\mu$  uma medida  $\sigma$ -finita em  $S = \mathbb{R}^d$ . Um processo de Poisson  $Y$  em  $S$  com média  $\mu$  é definido como um subconjunto enumerável aleatório  $\Pi \subseteq S$ , de modo que,  $Y(A)$  é o número de pontos de  $\Pi$  no conjunto mensurável  $A \in S$ , tal que:*

1.  $Y(A)$  é uma variável aleatória com distribuição de Poisson com média  $\mu = \mu(A)$ , onde  $0 \leq \mu \leq \infty$  e,
2. para quaisquer subconjuntos mensuráveis disjuntos  $A_1, \dots, A_n \in S$ , as variáveis aleatórias  $Y(A_1), \dots, Y(A_n)$  são independentes.

Portanto, se  $\mu(A) = 0$ ,  $P(Y(A) = 0) = 1$  e se  $\mu(A) = \infty$ ,  $P(Y(A) = \infty) = 1$ .

A medida  $\mu$  pode ser definida como uma função da intensidade  $\lambda$ , que representa a densidade média de pontos do processo. Desta forma  $\mu$  é dado pela integral de  $\lambda$  com respeito a uma medida  $d$ -dimensional:

$$\mu(A) = \int_A \lambda(x) dx.$$

Quando  $\lambda$  varia de acordo com cada localização do espaço, o processo é denominado processo de Poisson não-homogêneo. Se  $\lambda$  for constante, denominamos de Processo de Poisson homogêneo e neste caso  $\mu(A) = \lambda|A|$ , com  $|A|$  sendo a medida de Lebesgue de  $A$  (se  $d = 2$ , representa a área de  $A$ ).

## 2.2 Processo de Cox

Um processo de Cox (Cox [1955]) é um processo de Poisson não-homogêneo onde a função de intensidade evolui estocasticamente. Esses processos têm sido amplamente utilizados em uma variedade de áreas para modelar fenômenos de padrões de pontos.

**Definição 2.** *Defina a função  $\lambda(s) : \mathbb{R}^d \rightarrow \mathbb{R}^+$  como um processo estocástico que assume valores não negativos. Desta forma  $Y$  é dito ser um processo de Cox com função intensidade  $\lambda(s)$  se a distribuição condicional de  $Y$  dado  $\lambda(s)$  é um processo de Poisson com intensidade  $\lambda(s)$ .*

## 2.3 Processo Gaussiano

Na literatura sobre processo de Cox é muito usual encontrar modelos onde a função intensidade depende de um processo Gaussiano (Møller et al. [1998], Diggle [2013], Gonçalves and Gamerman [2018]). Nesta tese, o processo Gaussiano será uma ferramenta importante na determinação da partição aleatória do espaço com intensidades constantes.

**Definição 3.** Um processo Gaussiano estacionário e isotrópico  $\beta$  é um processo estocástico em algum espaço  $S \subset \mathbb{R}^d$ , tal que a distribuição conjunta do processo em qualquer coleção finita de pontos neste espaço é Gaussiana, ou seja,  $\forall n > 1$  e  $s_1, \dots, s_n \in S$ :

$$(\beta(s_1), \dots, \beta(s_n)) \sim N(\mu, \Sigma), \quad (2.1)$$

onde  $\mu$  é o vetor de médias e  $\Sigma$  a matriz de covariância, onde  $\text{cov}(\beta(s), \beta(s)) = \sigma^2$  e  $\text{cov}(\beta(s), \beta(s')) = \sigma^2 \rho(|s - s'|)$ , com  $\rho$  sendo a função de correlação do processo e  $|s - s'|$  a norma do vetor  $(s - s')$ .

Existem diversas escolhas razoáveis para  $\rho$  na literatura, por exemplo, a família exponencial de funções de covariâncias, que tem a forma:

$$\rho(|s - s'|) = \sigma^2 \exp \left\{ -\frac{1}{2\tau^2} |s - s'|^\gamma \right\}, \quad (2.2)$$

com  $\gamma > 1$  para gerar superfícies diferenciáveis. Esta escolha foi utilizada por Gonçalves and Gamerman [2018].

## 2.4 Algoritmo Metropolis–Hastings

O paradigma Bayesiano diz que a inferência é realizada por meio da distribuição *a posteriori* das quantidades desconhecidas no modelo. Porém, na maioria das vezes, obter esta distribuição não é trivial e a solução mais comum para este problema é utilizar métodos de Monte Carlo via Cadeia de Markov (MCMC). A teoria de MCMC é extensa e complexa, uma discussão mais geral sobre o tema pode ser encontrada por exemplo em Gamerman and Lopes [2006]. No caso Bayesiano, a idéia básica deste método é obter uma amostra (aproximada) grande o suficiente da distribuição *a posteriori* de forma que o Teorema Ergódico garanta a convergência da média ergódica de funções integráveis com relação à *posteriori*.

Os algoritmos de MCMC mais utilizados são o *Gibbs sampling* e o *Metropolis-Hastings* (MH). O primeiro consiste em amostrar das distribuições condicionais de blocos de coordenadas da *posteriori*. O segundo, define uma distribuição proposta  $q(\cdot)$  da qual, a cada iteração

da cadeia, é proposto um valor  $\theta^c$  que é aceito com probabilidade  $\alpha$ , tal que a reversibilidade da cadeia com relação à *posteriori* seja garantida. Em ambos os casos, é necessário garantir a irreduzibilidade e a aperiodicidade da cadeia para garantir a validade do Teorema Ergódico. O Algoritmo 1 fornece os detalhes de um MH no contexto Bayesiano, ou seja, quando  $\pi(\cdot) = \pi(\theta|y)$  é a distribuição *a posteriori*, sendo  $\theta$  as quantidades desconhecidas do modelo e  $y$  os dados.

---

**Algorithm 1** METROPOLIS-HASTINGS

---

Inicialize  $\theta$  atribuindo-lhe um valor inicial  $\theta^0$ .

Para  $i = 0, 1, 2, \dots$

1 Proponha  $\theta^c \sim q(\cdot|\theta^i)$ ;

2 Com probabilidade

$$\alpha_{MH}(\theta^i, \theta^c) = 1 \wedge \frac{\pi(\theta^c|y)q(\theta^i|\theta^c)}{\pi(\theta^i|y)q(\theta^c|\theta^i)}$$

faça  $\theta^{i+1} = \theta^c$ . Caso contrário,  $\theta^{i+1} = \theta^i$ .

---

Uma propriedade fundamental do algoritmo MH é que só precisamos conhecer a parte de  $\pi$  que é proporcional a  $\theta$ , uma vez que as constantes são canceladas na expressão da probabilidade de aceitação. Observe também que a cadeia pode permanecer no mesmo estado por muitas iterações sendo que a taxa de aceitação está intimamente ligada às propriedades de convergência da cadeia (ver Roberts et al. [1997]).

## 2.5 Verossimilhança intratável e amostragem retrospectiva

Em muitos modelos estatísticos, o espaço amostral e ou paramétrico é infinito-dimensional. Isso ocorre, por exemplo, na família de modelos proposta nesta tese. Este tipo de situação geralmente leva à intratabilidade analítica da função de verossimilhança e, conseqüentemente, da densidade da *posteriori*. Isso significa que não é possível calcular o valor da probabilidade de aceitação do algoritmo MH a cada iteração da cadeia. Além disso, a dimensionalidade

infinita do espaço paramétrico implica que a cadeia de Markov do algoritmo de MCMC também será infinito-dimensional.

A amostragem retrospectiva é uma possível solução para esse tipo de problema. De forma geral, essa técnica é definida como qualquer algoritmo de simulação estocástica onde a ordem usual dos passos é invertida sem prejudicar o resultado mas tal que possa se tirar grandes vantagens dessa estratégia. Essa técnica é extremamente poderosa no contexto de simulação de variáveis aleatórias de dimensão infinita e no ganho de eficiência em termos de complexidade e custo computacional. Mais especificamente, a ideia é simular apenas uma representação finita dessa variável, de forma que o algoritmo em questão possa ser executado, como por exemplo, o passo de aceitação/rejeição do algoritmo Metropolis-Hastings. Além disto, esta representação finita (e aleatória) deve também permitir que, condicionado nela, qualquer outra parte finita da variável aleatória seja simulada. No contexto desta tese, a verossimilhança é intratável e a amostragem retrospectiva será aplicada no contexto dos algoritmos *Pseudo-Marginal MH* e de Barker, discutidos a seguir.

### 2.5.1 Algoritmo *Pseudo-Marginal Metropolis-Hastings*

O algoritmo *Pseudo-Marginal* foi proposto por Andrieu and Roberts [2009]. Suponha que exista o interesse em implementar um algoritmo Metropolis-Hastings, mas que a densidade *a posteriori* seja analiticamente intratável e, portanto, não seja possível calcular a probabilidade de aceitação do algoritmo.

No contexto Bayesiano, temos que  $\pi(\theta|y) \propto L(\theta, y)\pi(\theta)$ , onde  $L(\theta, y)$  é a função de verossimilhança do modelo e, portanto, podemos reescrever a probabilidade de aceitação do algoritmo Metropolis-Hastings da seguinte forma:

$$\alpha_{MH} = 1 \wedge \frac{L(\theta^c, y)\pi(\theta^c)q(\theta^i|\theta^c)}{L(\theta^i, y)\pi(\theta^i)q(\theta^c|\theta^i)} \quad (2.3)$$

Andrieu and Roberts [2009] mostram que se substituirmos  $L(\theta, y)$  por um estimador não-viesado e quase certamente maior que zero na probabilidade de aceitação  $\alpha_{MH}$ , a distribuição

*a posteriori* é preservada como a distribuição marginal da distribuição invariante da nova cadeia de Markov, aumentada com a componente aleatória do estimador não-viciado da verossimilhança. Por este motivo, o método *Pseudo-Marginal* é visto como uma “aproximação exata” do Metropolis-Hastings. O Algoritmo 2 fornece os detalhes do *Pseudo-Marginal* no contexto Bayesiano.

Seja  $N$  uma variável aleatória com densidade  $\pi_N$  e  $\hat{L}$  uma função tal que,  $E_{\pi_N}[\hat{L}(y, \theta, N)] = L(\theta, y)$ ,  $\forall \theta \in \Theta$ ,  $\forall y \in Y$  e  $\hat{L} > 0$  quase certamente.

---

**Algorithm 2** *Pseudo-Marginal-Metropolis-Hastings*

---

Inicialize a cadeia com  $(\theta, N) = (\theta^0, N^0)$ .

Para  $i = 0, 1, 2, \dots$

**1** Proponha  $\theta^c \sim q(\cdot|\theta^i)$  e  $N^c \sim \pi_N(\cdot|N^i)$ ;

**2** Com probabilidade

$$\alpha_{PM}(\theta^i, \theta^c) = 1 \wedge \frac{\hat{L}(y, \theta^c, N^c)\pi(\theta^c)q(\theta^i|\theta^c)}{\hat{L}(y, \theta^i, N^i)\pi(\theta^i)q(\theta^c|\theta^i)}$$

faça  $(\theta^{i+1}, N^{i+1}) = (\theta^c, N^c)$ . Caso contrário,  $(\theta^{i+1}, N^{i+1}) = (\theta^i, N^i)$ .

---

## 2.5.2 Algoritmo de Barker

O algoritmo Metropolis-Hastings (MH) pertence a uma família de algoritmos que apresenta uma mesma estrutura: propor um valor candidato do parâmetro  $\theta$  a partir de uma distribuição com densidade  $q(\cdot)$  e aceita-lo com probabilidade  $\alpha$  que preserve a reversibilidade da cadeia com relação à distribuição de interesse. Os algoritmos se diferenciam pela forma da probabilidade  $\alpha$ , sendo o Metropolis-Hastings o algoritmo mais eficiente dessa classe. Isso acontece porque o MH maximiza a probabilidade de transição global e, portanto otimiza o *mixing* da cadeia, minimizando assim a variância assintótica de estimadores de MC de funções em  $\mathcal{L}^2(\pi(\theta|y))$  para as quais vale um Teorema Central do Limite (ver Peskun [1973]). Apesar disto, existem outras opções para a probabilidade de aceitação que definem algoritmos quase tão bons quanto o MH.

Algumas propostas na literatura visam simular de um evento com probabilidade  $\alpha$  diretamente, sem recorrer ao cálculo da probabilidade de aceitação. Nos referimos a esta estratégia como simulação de um evento de probabilidade desconhecida. No entanto, no caso do MH, mesmo que um estimador não viciado para a densidade *a posteriori* seja conhecido, não é possível simular um evento com a probabilidade requerida (ver Łatuszyński et al. [2011]).

Uma alternativa para este problema é apresentada por Gonçalves et al. [2017a], onde os autores descrevem uma implementação do algoritmo MCMC usando a probabilidade de aceitação do algoritmo de Barker [Barker, 1965] de forma que um evento dado por esta probabilidade pode ser simulado sem a necessidade de calcular o seu valor. Basicamente, o algoritmo necessita que se conheça um estimador não viciado e quase certamente positivo para a parte proporcional da densidade *a posteriori*. O algoritmo que simula o evento desejado é chamado de *2-coin algorithm*. O algoritmo Barker possui a seguinte probabilidade de aceitação (já no contexto Bayesiano):

$$\alpha_B(\theta^i, \theta^c) = \frac{\pi(\theta^c|y)q(\theta^c|\theta^i)}{\pi(\theta^c|y)q(\theta^c|\theta^i) + \pi(\theta^i|y)q(\theta^i|\theta^c)} \quad (2.4)$$

Pode-se mostrar que

$$\frac{\alpha_{MH}(\theta^i, \theta^c)}{2} < \alpha_B(\theta^i, \theta^c) < \alpha_{MH}(\theta^i, \theta^c),$$

indicando que os dois algoritmos têm desempenhos semelhantes e o algoritmo de Barker é quase tão bom quanto o Metropolis-Hastings.

Segundo Gonçalves et al. [2017a], se um estimador não viciado e quase certamente positivo para a parte proporcional da densidade *a posteriori* é conhecido, a probabilidade de aceitação do algoritmo de Barker's pode ser reescrita na forma

$$\alpha_B = \frac{c_1 p_1}{c_1 p_1 + c_2 p_2},$$

tal que  $c_1$  e  $c_2$  são constantes positivas e conhecidas e  $p_1$  e  $p_2$  são probabilidades desconhecidas mas tais que eventos com essas probabilidades podem ser simulados. Dessa forma, um evento de probabilidade  $\alpha_B$  é simulado da seguinte forma.

---

**Algorithm 3** ALGORITMO *2-coin* PARA AMOSTRAR A PROBABILIDADE DE ACEITAÇÃO DE BARKER

---

- 1** Amostre  $C_1 \sim Ber\left(\frac{c_1}{c_1+c_2}\right)$
  - 2** Se  $C_1 = 1$ , amostre  $C_2 \sim Ber(p_1)$ ;
    - Se  $C_2 = 1$ ; retorne 1;
    - Se  $C_2 = 0$ ; volte ao passo **1**;
  - 3** Se  $C_1 = 0$ , amostre  $C_2 \sim Ber(p_2)$ ;
    - Se  $C_2 = 1$ ; retorne 0;
    - Se  $C_2 = 0$ ; volte ao passo **1**.
- 

O algoritmo *2-coin* retorna 1 (aceita a proposta  $\theta^c$ ) com probabilidade  $\frac{c_1 p_1}{c_1 p_1 + c_2 p_2}$  e 0 (rejeita a proposta  $\theta^c$ ) com probabilidade  $\frac{c_2 p_2}{c_1 p_1 + c_2 p_2}$ .

Esta abordagem pode ser vista como uma alternativa ao algoritmo *Pseudo-Marginal* descrito anteriormente, principalmente em cenários onde a simulação de um evento de probabilidade proporcional a  $\pi(\theta|y)$  é possível.

# Capítulo 3

## Modelo Bayesiano para processos de Cox Level-Set espaciais

Considere um processo de Poisson (PP) não homogêneo  $Y = \{Y(s) : s \in S\}$ , com função intensidade  $\lambda_S = \{\lambda(s) : s \in S\}$ , tal que  $\lambda(s) : S \rightarrow \mathbb{R}^+$ , onde  $S$  é alguma região compacta em  $\mathbb{R}^2$  e  $\mathbf{S}_K = \{S_1, \dots, S_K\}$  é uma partição aleatória finita de  $S$ . Defina  $\lambda = (\lambda_1, \dots, \lambda_K)$  como um vetor aleatório não negativo. Baseado no trabalho de Dunlop et al. [2016],  $\mathbf{S}_K$  é definida como uma função de conjuntos de níveis que determina regiões (interfaces) do espaço que possuem intensidades constantes dadas por  $\lambda$ . Portanto, para cada sub-região  $S_k \in S$ , temos um processo de Poisson homogêneo com intensidade  $\lambda_k$ . Seja  $c = (c_1, \dots, c_{K-1}) \in \mathbb{R}^{K-1}$ , com  $K \in \mathbb{N}$  fixo, tal que  $-\infty = c_0 < c_1 < \dots < c_{K-1} < c_K = \infty$ . Considere o seguinte modelo Bayesiano hierárquico para  $Y$ :

$$(Y|\lambda_S) \sim PP(\lambda_S), \quad (3.1)$$

$$\lambda(s) = \sum_{k=1}^K \lambda_k I_k(s), \quad (3.2)$$

$$S_k = \{s \in S : c_{k-1} < \beta(s) < c_k\}, \forall k \quad (3.3)$$

$$\beta \sim PG(\mu, \Sigma), \quad (3.4)$$

$$c \sim \text{prior} \quad (3.5)$$

$$\lambda_k \stackrel{\text{ind}}{\sim} \text{Gama}(\alpha_k, \gamma_k), \forall k \quad (3.6)$$

onde  $I_k(s)$  é a função indicadora de  $s \in S_k$  e  $\beta$  é um processo Gaussiano com funções média e covariância dadas por  $\mu$  e  $\Sigma$ , respectivamente. Dessa forma, condicionado em  $(\beta, c, \lambda)$ ,  $Y_k$  é um processo de Poisson homogêneo em  $S_k$  com intensidade constante  $\lambda_k$ . Note que o processo Gaussiano  $\beta$  e o vetor  $c$  determinam somente a partição  $S_k$ , não existindo uma relação direta entre eles e o valor da função intensidade. Além disso  $\beta$ ,  $c$  e  $\lambda_k$  são mutuamente independentes *a priori*.

O modelo proposto acima pode ser estendido para um modelo espaço-temporal, de tal forma que os dados podem ser visto como uma série temporal de processos pontuais, considerando o tempo discreto. A dependência temporal do modelo pode ser definida através do processo Gaussiano  $\beta$ .

## Fronteira das sub-regiões

O modelo proposto apresenta uma restrição com respeito as vizinhanças das sub-regiões. A região 1 só faz fronteira com a 2, a região  $k$  só faz fronteira com as regiões  $k - 1$  e  $k + 1$ , para  $k = 2, \dots, K - 1$ , e a sub-região  $K$  só faz fronteira com a  $K - 1$ . Isto ocorre devido a forma que o modelo define as regiões e pelo fato do processo Gaussiano ser quase certamente contínuo. Note, porém, que não existe uma restrição com respeito à ordenação dos  $\lambda$ 's.

Na teoria isto parece ser algo muito restritivo, porém na prática não é necessariamente um problema, uma vez que basta que exista uma pequena área das sub-regiões sequentes necessárias para que o modelo seja válido. Por exemplo, o modelo não seria capaz de estimar o caso da Figura (3.1), pois a região  $S_1$  faz fronteira com a  $S_3$ . Porém, o que ocorrerá se ajustarmos nosso modelo para este caso é a identificação de um dos três casos presentes na Figura (3.2) que, na prática, seria bom o suficiente. Além disso, uma estimativa da FI produzida pela sua média à *posteriori* seria ainda menos impactada devido à suavização das fronteiras imposta pela maior incerteza acerca dessas.

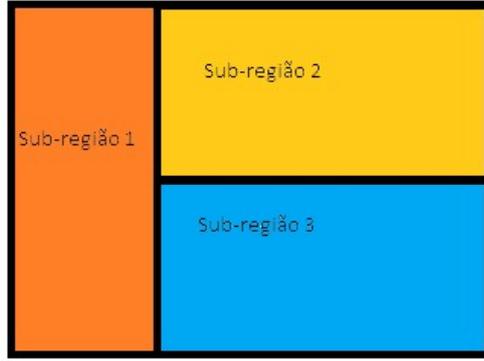


Figura 3.1: Exemplo de configuração não contemplada pelo modelo proposto.



Figura 3.2: Situações que o modelo identifica.

### 3.1 Inferência para o modelo proposto

O principal objetivo do processo de inferência para a classe de modelos que estamos trabalhando é a estimação da função intensidade. Defina  $\{s_i\}_{i=1}^{|Y|}$  como as localizações de  $Y$  e  $\theta = \{\lambda, c, \beta\}$  como o vetor de todos os parâmetros do modelo, onde  $|Y|$  representa o número de pontos em  $Y$ . Note que o modelo é não-paramétrico devido à dimensão infinita de  $\beta$ .

A função de verossimilhança do modelo proposto é obtida através da densidade de  $Y$  com respeito à medida de um  $PP(1)$  em  $S$  [ver Gonçalves and Franklin, 2019] é dada por:

$$L(\theta, Y) \propto \exp \left\{ - \sum_{k=1}^K \lambda_k \mu_k \right\} \prod_{k=1}^K (\lambda_k)^{|Y_k|}, \quad (3.7)$$

onde  $\mu_k$  e  $|Y_k|$  são a área e o número de observações da região  $S_k$ , respectivamente.

A função de verossimilhança do modelo proposto não é identificável, uma vez que para cada ponto do espaço paramétrico temos um conjunto não-enumerável de valores para os quais a função de verossimilhança assume o mesmo valor. Para facilitar o entendimento, redefina  $\beta$  como  $\beta = \mu + \sigma\beta^*$  com  $\beta^* \sim N(0, I_\rho)$ , onde  $I_\rho$  é a matriz de covariância com  $\sigma^2 = 1$ , construída de tal forma que as correlações do processo  $\beta$  sejam mantidas. Qualquer transformação do tipo  $\mu^* = a\mu + m$ ,  $\sigma^* = a\sigma$  e  $c_k^* = m + ac_k$ ,  $\forall m \in \mathbb{R}^d, a \in \mathbb{R}, \forall k$  definem a mesma partição  $\mathbf{S}_K$ , fazendo com que  $L(\mu^*, \sigma^*, c^*, \lambda) = L(\mu, \sigma, c, \lambda)$ . Este problema pode ser resolvido, sem perder a flexibilidade do modelo, fixando os parâmetros  $c = (c_1, \dots, c_{K-1})$  ou fixando os hiperparâmetros  $(\mu, \sigma^2)$  do processo Gaussiano. Iremos adotar a segunda restrição por facilitar a implementação computacional do algoritmo de MCMC.

Outro problema de identificabilidade é o *label-switching*, em que as sub-regiões e seus respectivos  $\lambda$ 's podem se inverterem, resultando na mesma verossimilhança. Uma forma de resolver este problema seria impor uma ordenação para os  $\lambda$ 's, porém devido às características do modelo e da informação mínima que se espera de qualquer conjunto de dados, acredita-se que o problema de *label-switching* não ocorrerá.

Sob o paradigma Bayesiano, a inferência de  $\theta$  é baseada na distribuição *a posteriori*  $\pi(\theta|Y)$ , que tem densidade proporcional à densidade conjunta de  $(Y, \theta)$ , dada por:

$$\begin{aligned} \pi(\theta, Y) &= L(\theta, Y)\pi(\theta) \\ &= L(\theta, Y) \left[ \prod_{k=1}^K \pi(\lambda_k) \right] \left[ \prod_{k=1}^{K-1} \pi(c_k) I_{(c_1 < \dots < c_{k-1})} \right] \pi_{PG}(\beta) \\ &\propto \exp \left\{ - \sum_{k=1}^K \lambda_k \mu_k \right\} \left[ \prod_{k=1}^K (\lambda_k)^{|Y_k|} \pi(\lambda_k) \right] \left[ \prod_{k=1}^{K-1} \pi(c_k) I_{(c_1 < \dots < c_{k-1})} \right] \pi_{PG}(\beta), \end{aligned} \tag{3.8}$$

com  $\pi(\lambda_k)$ ,  $\pi(c_k)$  e  $\pi_{PG}$  representando as densidades *a priori* de  $\lambda_k$ ,  $c_k$  e  $\beta$ , respectivamente.

Devido à complexidade de  $\pi(\theta|Y)$ , será proposto um algoritmo MCMC para gerar uma amostra (aproximada) desta distribuição. O algoritmo consiste em simular uma cadeia de Markov que tem a distribuição *a posteriori* conjunta como sua distribuição invariante e para a qual seja válido o Teorema Ergódico. Note, porém, que (3.8) é analiticamente intratável, ou seja, não é possível calcular seu valor pontualmente. Isto inviabiliza a execução de um algoritmo Metropolis usual. Além disto, a simulação direta de qualquer distribuição condi-

cional completa é também inviável. Uma maneira de solucionar este problema seria o uso de aproximações discretas mas, como discutido anteriormente, existem diversas desvantagens que nos motivam a desenvolver uma metodologia de inferência exata. A solução proposta aqui será baseada em uma combinação de técnicas recentes de simulação estocástica.

Um dos motivos da intratabilidade de (3.8) é que o processo Gaussiano  $\beta$  tem dimensão infinita. Uma solução para lidar com esta questão é o uso da técnica de amostragem retrospectiva, descrita no Capítulo 2, que nos permite executar o algoritmo MCMC revelando o valor do processo  $\beta$  apenas em uma quantidade finita (e aleatória) de localizações.

Outro motivo da intratabilidade é que os  $\mu_k$ 's não podem ser calculados. Uma possível solução seria avaliar o processo Gaussiano em um número suficientemente grande de localizações e aproximar esses valores. Porém, isto seria inviável computacionalmente, devido ao alto custo em se simular um processo gaussiano. A solução para resolver este problema será a construção de um algoritmo *Pseudo-Marginal Metropolis* que, como descrito anteriormente, substitui a função de verossimilhança por um estimador não viesado da mesma na expressão da probabilidade de aceitação do Metropolis. Desta forma, a distribuição *a posteriori* é preservada como a distribuição marginal da nova cadeia de Markov. No nosso caso, este estimador não-viesado da verossimilhança é obtido através de uma técnica conhecida como *Poisson Estimator*, que será explicada em mais detalhes na próxima seção.

## 3.2 *Pseudo-Marginal* para o modelo proposto e o *Poisson Estimator*

Para a construção do algoritmo *Pseudo-Marginal* é necessário encontrar um estimador não-viesado para a verossimilhança, o que significa obter um estimador não-viesado de  $M = \exp\left\{-\sum_{k=1}^K \lambda_k \mu_k\right\}$ , pois  $L(\theta, Y) = aM$ , onde  $a$  é conhecido. Um estimador não-viesado de  $M$  é derivado a partir dos estimadores não-viesados dos  $\mu_k$ 's. Porém, como  $M$  é uma função não linear dos  $\mu_k$ 's, esta derivação não é trivial e, conforme mostrado a seguir o estimador

desejado é função de uma quantidade aleatória de  $G_k$ 's. Utilizando o *Poisson Estimator* (ver Beskos et al. [2006], Fearnhead et al. [2008]), obtemos o seguinte estimador para  $M$ .

**Proposição 1.** *Um estimador não-viesado para  $M$  é dado por:*

$$\hat{M} = e^{-\mu_S \lambda_m} \prod_{k=1}^K \left( \frac{\delta \lambda_M - \lambda_k}{\delta \lambda_M - \lambda_m} \right)^{|N_k|}, \quad (3.9)$$

onde  $N = g(N^*, \lambda^*)$ , com  $\lambda^* = (\delta \lambda_M - \lambda_m)$ ,  $\lambda_M = \max_k \{\lambda_k\}$  e  $\lambda_m = \min_k \{\lambda_k\}$ .  $N^*$  é um processo de Poisson com intensidade 1 no cilindro de base  $S$  e altura de 0 a  $+\infty$ . A função  $g$  é a projeção dos pontos de  $N^*$  em  $S$  que têm altura menor que  $\lambda^*$ . Portanto,  $N$  é um processo de Poisson homogêneo em  $S$  com intensidade  $\lambda^*$ .

Note que os pontos de  $N$  irão determinar a representação finita do processo Gaussiano  $\beta$  no contexto da amostragem retrospectiva. O estimador  $\hat{M}$  tem as seguintes propriedades.

**Proposição 2.** *A variância do estimador  $\hat{M}$  é finita e uma função decrescente em  $\delta$ .*

*Prova das Proposições 1 e 2:* Ver apêndice 7. □

### 3.2.1 Relação do *Poisson Estimator* com a constante $\delta$

A escolha de  $\delta$  é determinante na eficiência do algoritmo. Pensando no custo computacional é importante escolher um  $\delta$  que não seja muito alto, pois quanto maior o  $\delta$ , maior o número de localizações de  $\beta$  necessários para calcular o estimador. Por outro lado, se o valor de  $\delta$  for muito baixo o estimador pode apresentar uma variância alta, prejudicando o *mixing* do MCMC. Desta forma, é importante escolher um  $\delta$  que leve em conta todas estas informações.

Com o objetivo de analisar essa relação, foi realizado estudos de simulação considerando alguns cenários em uma região  $S = (0, 10) \times (0, 10)$ .

- Cenário 1:  $K = 2$ , com  $(\lambda_1 = 2, \mu_1 = 70)$  e  $(\lambda_2 = 5, \mu_2 = 30)$ .
- Cenário 2:  $K = 2$ , com  $(\lambda_1 = 2, \mu_1 = 76)$  e  $(\lambda_2 = 10, \mu_2 = 24)$ .
- Cenário 3:  $K = 3$ , com  $(\lambda_1 = 1.55, \mu_1 = 60)$ ,  $(\lambda_2 = 4, \mu_2 = 24)$  e  $(\lambda_3 = 9, \mu_3 = 16)$ .

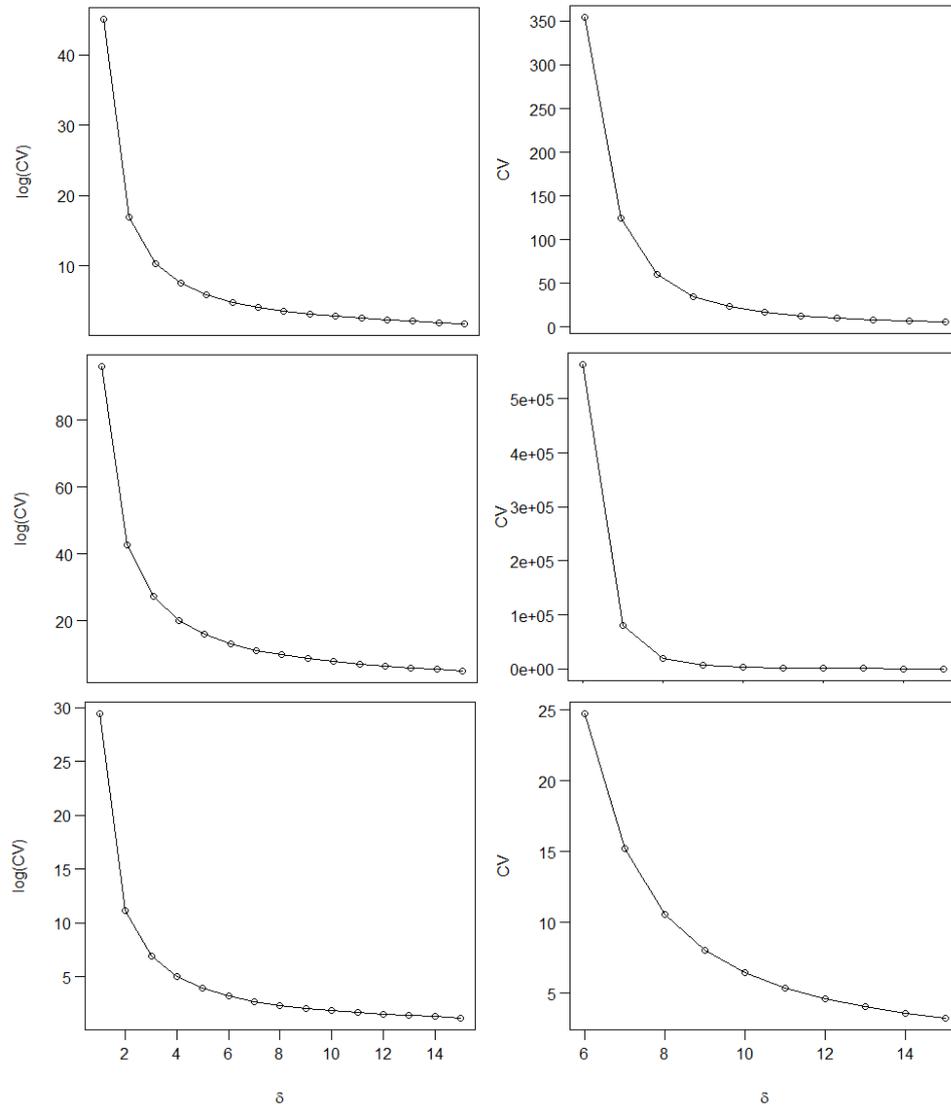


Figura 3.3: Gráficos de  $\delta \times \log(CV[\hat{M}])$  e gráficos de  $\delta \times CV[\hat{M}]$  para os cenários 1, 2 e 3, respectivamente.

Os resultados mostram que, de forma geral, o coeficiente de variação (CV) é maior para maiores diferenças entre os valores dos  $\lambda_k$ 's. Os exemplos apresentados fornecem uma boa ideia da escala de variação do estimador em função dos  $\lambda_k$ 's. Em situações reais, estimações

$\delta$	$CV(\hat{M})$ - cenário 1	$CV(\hat{M})$ - cenário 2	$CV(\hat{M})$ - cenário 3
2	2.131667e+07	3.387645e+18	7.078779e+04
3	3.235796e+04	3.387645e+18	9.851781e+02
4	1.808042e+03	5.988528e+08	1.461673e+02
5	3.540936e+02	8.886111e+06	4.959360e+01
6	1.241383e+02	5.631939e+05	2.472082e+01
7	5.978586e+01	8.034761e+04	1.519689e+01
8	3.489072e+01	1.888558e+04	1.060249e+01
9	2.307146e+01	6.168881e+03	8.029813
10	1.662144e+01	2.531924e+03	6.432753
11	1.273170e+01	1.225508e+03	5.364151
12	1.020428e+01	6.708016e+02	4.607498
13	8.464894	4.034276e+02	4.047609
14	7.212192	2.611801e+02	3.618526
15	6.276299	1.793203e+02	3.280184

Tabela 3.1: Coeficiente de variação de  $\hat{M}$  com diferentes  $\delta$ 's para os cenários 1, 2 e 3, respectivamente.

empíricas dos  $\lambda_k$ 's podem ser utilizadas para se escolher o valor de  $\delta$  a ser utilizado. Resta ainda verificar a performance do algoritmo MCMC proposto em função da escolha do  $\delta$ , como será feito no Capítulo 5.

Uma boa estratégia é, sempre que  $K$  for maior que 3, agrupar os  $k$ 's em grupos de 2 (e, se  $k$  for ímpar, um grupo de 3), que tenham os  $\lambda_k$ 's correspondentes mais próximos, e utilizar um *Poisson Estimator* para cada termo na fatoração correspondente de  $M$ .

# Capítulo 4

## MCMC

A inferência para o modelo proposto será realizada via MCMC, mais especificamente, um algoritmo *Pseudo-Marginal Metropolis*, dada a intratabilidade analítica da função de verossimilhança. Como o vetor  $(\theta, N^*)$  de quantidades desconhecidas possui componentes de natureza bem distinta - tanto topológicas quanto em relação ao papel desempenhado no modelo, um algoritmo que amostra todo o vetor conjuntamente dificilmente será eficiente. Por este motivo, considera-se a amostragem por blocos sendo que, para cada bloco, a probabilidade de aceitação é proporcional à probabilidade do algoritmo que amostra todo o vetor conjuntamente. Os seguintes blocos são definidos

$$\{N^*\} \quad \{\beta\} \quad \{\lambda\} \quad \{c\}$$

A probabilidade de aceitação de um movimento  $(\theta, N^*) \rightarrow (\dot{\theta}, \dot{N}^*)$  é dada por

$$1 \wedge \frac{\hat{\pi}(\dot{\theta}, \dot{N}^*) q(\theta, N^* | \dot{\theta}, \dot{N}^*)}{\hat{\pi}(\theta, N^*) q(\dot{\theta}, \dot{N}^* | \theta, N^*)}, \quad (4.1)$$

onde

$$\hat{\pi}(\theta, N^*) = e^{-\mu_S \lambda_m} \left[ \prod_{k=1}^K \left( \frac{\delta \lambda_m - \lambda_k}{\delta \lambda_M - \lambda_m} \right)^{|N_k|} (\lambda_k)^{|Y_k|} \pi(\lambda_k) \right] \pi(c) \pi_{PG}(\beta) \quad (4.2)$$

e  $N^*$  é proposto de um  $PP(1)$  no cilindro de base  $S$  e altura de 0 a  $+\infty$  ou qualquer outra distribuição que constitua uma cadeia de Markov em  $N^*$  que tenha o  $PP(1)$  como distribuição invariante.

## 4.1 Amostragem do $N^*$

Por ser um processo de Poisson em um volume infinito, o processo  $N^*$  tem infinitos pontos quase certamente e, portanto não pode ser amostrado em sua totalidade. Note, porém, que a probabilidade de aceitação do *Pseudo-Marginal Metropolis* depende de  $N^*$  apenas através dos pontos com altura menor que  $\lambda^*$ . Desta forma, esta componente é amostrada retrospectivamente, apenas abaixo do valor de  $\lambda^*$ , conforme necessário.

Outro aspecto importante relacionado a esta componente é o fato de definirmos a cadeia de Markov em função de  $N^*$  e não diretamente em função de  $N$ . O motivo é que, por ter distribuição dependente de  $\lambda^*$ , a componente  $N$  só poderia ser amostrada conjuntamente com  $\lambda^*$ , caso contrário, não teríamos um estimador não viciado para  $M$ . Além disso, como explicado anteriormente, deseja-se evitar a amostragem conjunta de componentes muito distintas. Por fim, note que a distribuição de  $N^*$  não depende de nenhuma outra componente do modelo e, por isso, essa componente pode ser amostrada separadamente das demais.

A probabilidade de aceitação em (4.1)-(4.2) sugere que propor  $N^*$  diretamente de um  $PP(1)$  levará a uma taxa de aceitação muito baixa, prejudicando severamente a convergência da cadeia. Alternativamente, definimos uma distribuição proposta que define uma cadeia de Markov com distribuição invariante dada por um  $PP(1)$ . Note que tal escolha preserva a probabilidade de aceitação em (4.1)-(4.2). A ideia é que, a cada iteração, proponha-se adicionar ou remover um ponto ou preservar os pontos atuais de  $N^*$  abaixo de  $\lambda^*$ . No entanto, o número de pontos de  $N^*$  nessa região será tipicamente muito alto e, portanto, o *mixing* dessa componente será muito lento. Para resolver esse problema, dividimos a região  $S$  em uma grade regular de  $m$  regiões e definimos cada um dos  $m$  cilindro de zero a  $\lambda^*$ , com as respectivas bases, como um bloco no MCMC, no qual a distribuição proposta descrita anteriormente é utilizada. Note que, quanto maior o valor de  $m$ , menor será a taxa de aceitação (quando se adiciona ou remove um ponto) deste passo.

Resultados empíricos sugerem que o número de regiões  $m$  aproximadamente igual a metade do número médio de pontos de  $N$  é uma escolha robusta.

Defina  $n_j$  como sendo o número atual de pontos na  $j$ -ésima das  $m$  regiões e  $A$  como sendo o volume de cada uma dessas regiões, ou seja,  $\lambda^* \mu(S)/m$ . A distribuição proposta com relação ao número de pontos é dada por:

$$\begin{aligned} q(n_j, n_j + 1) &= \frac{A}{A + n_j + 1}; \\ q(n_j, n_j - 1) &= \begin{cases} \frac{n_j}{A + n_j}, & \text{se } n_j > 0; \\ 0, & \text{se } n_j = 0; \end{cases} \\ q(n_j, n_j) &= 1 - \{q(n_j, n_j + 1) + q(n_j, n_j - 1)\}. \end{aligned} \quad (4.3)$$

Cálculos simples mostram que a distribuição  $Poisson(A)$  é reversível para a cadeia definida acima (ver Apêndice 7). A invariância do  $PP(1)$  para a cadeia é garantida estabelecendo que: i) quando se propõe adicionar um ponto, esse é amostrado uniformemente no volume da respectiva região; ii) quando se propõe retirar um ponto, esse é escolhido aleatoriamente entre os pontos existentes.

Para cada uma das  $m$  regiões, a probabilidade de aceitação de um movimento  $N^{*(i)} \rightarrow N^{*(c)}$  é

$$\alpha_{N^*} = 1 \wedge \left\{ \prod_{k=1}^K \left( \frac{\delta \lambda_m - \lambda_k}{\delta \lambda_M - \lambda_m} \right)^{|N_k|^{(c)} - |N_k|^{(i)}} \right\}, \quad (4.4)$$

onde  $N^{(\cdot)} = g(N^{*(\cdot)}, \lambda^*)$ . Note que o cálculo dessa probabilidade pode ser massivamente simplificado dadas as características da distribuição proposta.

Por fim, note que a atualização de  $N^*$  acima de  $\lambda^*$  é feita de forma retrospectiva de um  $PP(1)$ . Na prática, isso significa que, caso um valor de  $\lambda^*$  maior que o anterior seja amostrado ao longo da cadeia, os pontos de  $N^*$  com altura entre esses dois valores é amostrado diretamente de um  $PP(1)$ .

## 4.2 Amostragem dos $\lambda$ 's via *Pseudo-Marginal*

Propomos uma amostragem conjunta dos  $\lambda$ 's com distribuição proposta dada por um passeio aleatório Gaussiano. A matriz de covariância da proposta, deve ser tal que a taxa

de aceitação varie entre 0.234 e 0.4. Sendo esta em torno de 0.4 para  $K = 2$  e 0.234 para  $K \geq 5$  [Roberts et al., 1997].

A probabilidade de aceitação de um movimento  $\lambda^{(i)} \rightarrow \lambda^{(c)}$  é

$$\alpha_{\lambda}^* = 1 \wedge \left\{ \frac{e^{-\mu_S \lambda_m^{(c)}} \prod_{k=1}^K \left( \frac{\delta \lambda_m^{(c)} - \lambda_k^{(c)}}{\delta \lambda_M^{(c)} - \lambda_m^{(c)}} \right)^{|N_k|^{(c)}} \left( \lambda_k^{(c)} \right)^{|Y_k|}}{e^{-\mu_S \lambda_m^{(i)}} \prod_{k=1}^K \left( \frac{\delta \lambda_m^{(i)} - \lambda_k^{(i)}}{\delta \lambda_M^{(i)} - \lambda_m^{(i)}} \right)^{|N_k|^{(i)}} \left( \lambda_k^{(i)} \right)^{|Y_k|}} \right\},$$

### 4.3 Amostragem dos $\lambda$ 's via Barker

Uma alternativa ao algoritmo *Pseudo-Marginal Metropolis* é o Algoritmo de Barker descrito no Capítulo 2. Sua implementação é, no entanto, inviável, dado o alto custo computacional que teria cada loop do algoritmo 2-coin. Porém, sua implementação é viável para atualizar os parâmetros  $\lambda_k$  individualmente em alguns casos. O uso do algoritmo de Barker apenas para esses parâmetros significa que, para tais, a variável auxiliar  $N^*$  do algoritmo *Pseudo-Marginal* é integrada fora e, portanto, a atualização desse bloco caracteriza o algoritmo de MCMC como um *Collapsed Gibbs Sampling* [Liu, 1994]. Não é óbvio qual das alternativas para amostrar  $\lambda$  é mais vantajosa. Por um lado, o passo via Barker elimina totalmente a variabilidade extra de  $N^*$  levando a um melhor *mixing* da cadeia. Por outro lado, o custo do algoritmo 2-coin pode ser bem mais alto que o custo do *Pseudo-Marginal*. Comparações empíricas serão apresentadas e discutidas no Capítulo 5.

A densidade condicional completa de  $\lambda_k$  marginalizada com respeito a  $N^*$  é dada por

$$\pi(\lambda_k | \cdot) \propto e^{-\mu_k \lambda_k} \lambda_k^{|Y_k|}. \quad (4.5)$$

Neste caso, um passeio aleatório uniforme é mais adequado que um Gaussiano pois evita a forte penalização do custo computacional do algoritmo 2-coin imposta pelas caudas da proposta Gaussiana [ver Gonçalves et al., 2017a]. A probabilidade de aceitação de um movimento  $\lambda_k^{(i)} \rightarrow \lambda_k^{(c)}$  é

$$\alpha_B = \frac{(\lambda_k^{(c)})^{|Y_k|} e^{-\mu_k \lambda_k^{(c)}}}{(\lambda_k^{(c)})^{|Y_k|} e^{-\mu_k \lambda_k^{(c)}} + (\lambda_k^{(i)})^{|Y_k|} e^{-\mu_k \lambda_k^{(i)}}}, \quad (4.6)$$

que pode ser reescrita como

$$\alpha_B = \begin{cases} \frac{1}{1 + \left(\frac{\lambda_k^{(i)}}{\lambda_k^{(c)}}\right)^{|Y_k|} e^{-\mu_k(\lambda_k^{(i)} - \lambda_k^{(c)})}}, & \text{se } \lambda_k^{(c)} < \lambda_k^{(i)}; \\ \frac{\left(\frac{\lambda_k^{(c)}}{\lambda_k^{(i)}}\right)^{|Y_k|} e^{-\mu_k(\lambda_k^{(c)} - \lambda_k^{(i)})}}{\left(\frac{\lambda_k^{(c)}}{\lambda_k^{(i)}}\right)^{|Y_k|} e^{-\mu_k(\lambda_k^{(c)} - \lambda_k^{(i)})} + 1}, & \text{se } \lambda_k^{(c)} \geq \lambda_k^{(i)}, \end{cases} \quad (4.7)$$

que na formulação do algoritmo 2-coin apresentado na seção 2.5.2 tem:

- $c_1 = 1$ ,  $p_1 = 1$ ,  $c_2 = \left(\frac{\lambda_k^{(i)}}{\lambda_k^{(c)}}\right)^{|Y_k|}$  e  $p_2 = e^{-\mu_k(\lambda_k^{(i)} - \lambda_k^{(c)})}$ , se  $\lambda_k^{(c)} < \lambda_k^{(i)}$ ;
- $c_1 = \left(\frac{\lambda_k^{(c)}}{\lambda_k^{(i)}}\right)^{|Y_k|}$ ,  $p_1 = e^{-\mu_k(\lambda_k^{(c)} - \lambda_k^{(i)})}$ ,  $c_2 = 1$  e  $p_2 = 1$ , se  $\lambda_k^{(c)} \geq \lambda_k^{(i)}$ .

Isto significa que, para executar o passo Barker via algoritmo 2-coin é necessário simular eventos de probabilidade  $M_k = e^{-\mu_k d_k}$  para  $d_k > 0$  conhecido. Para tal, derivamos um estimador não-viesado para  $M_k$ , que esteja em  $[0, 1]$  quase certamente, via *Poisson Estimator*.

**Proposição 3.** *Um estimador não-viesado para  $M_k$  é dado por:*

$$\hat{M}_k = \left(1 - \frac{1}{\delta^*}\right)^{|Z_k|},$$

onde  $Z$  é um processo de Poisson homogêneo com intensidade  $\delta^* d_k$  em  $S$  e  $|Z_k|$  é o número de pontos de  $Z$  pertencentes à região  $S_k$ , com  $\delta^* > 1$ .

*Prova:* Ver apêndice 7. □

## 4.4 Amostragem do processo Gaussiano

Primeiramente, note que a amostragem do processo Gaussiano (PG) latente  $\beta$  deve ser feita de forma retrospectiva, dada sua dimensionalidade infinita. Isso significa que o PG

será sempre amostrado em uma quantidade finita de localizações -  $\{Y, N\}$ , de forma que a decisão de se aceitar ou não a superfície de  $\beta$  proposta possa ser tomada. Além disso,  $\beta$  pode ser amostrado em qualquer outra localização, como será necessário nos passos de  $\lambda$  e  $N$ , condicionado às demais localizações já amostradas.

Temos ainda o problema do termo  $\pi_{PG}(\beta)$  aparecer na expressão da probabilidade de aceitação do algoritmo *Pseudo-Marginal*. Para resolver esse problema, adotamos *a priori* de  $\beta$  como distribuição proposta, o que permite cancelar o termo  $\pi_{PG}(\beta)$ . Note porém, que esta escolha levará a taxas de aceitação muito baixas, por ser uma proposta independente e não conter informação sobre *a posteriori* contida nos dados. Para mitigar esse problema, escolhemos, a cada iteração do MCMC, um conjunto finito e aleatório de localizações onde o PG não será atualizado. Essas localizações são escolhidas uniformemente dentre as localizações atuais de  $N$ , de forma que o número de localizações escolhidas é uma proporção fixa  $\mathbf{p}_N$  de  $|N|$ . A escolha dessa proporção impacta diretamente no *mixing* da cadeia (quanto maior a proporção, maior a taxa de aceitação) e, por essa proposta assemelhar-se a um tipo de passeio aleatório infinito-dimensional, essa proporção deve ser escolhida de forma que a taxa de aceitação de  $\beta$  seja próxima de 0.234 [Roberts et al., 1997].

A probabilidade de aceitação de um movimento  $\beta^{(i)} \rightarrow \beta^{(c)}$  é

$$\alpha_\beta = 1 \wedge \left\{ \prod_{k=1}^K \left( \frac{\delta\lambda_m - \lambda_k}{\delta\lambda_M - \lambda_m} \right)^{|N_k|^c - |N_k|^i} (\lambda_k)^{|Y_k|^c - |Y_k|^i} \right\}, \quad (4.8)$$

## 4.5 Amostragem do $c$

A amostragem do vetor  $c$  é feita de forma conjunta, propondo de um passeio aleatório uniforme de mesmo raio para todas as componentes do vetor. Caso a ordenação dessas componentes seja preservada na proposta, um movimento  $c^{(i)} \rightarrow c^{(c)}$  é aceito com probabilidade

$$\alpha_c = 1 \wedge \left\{ \prod_{k=1}^K \left( \frac{\delta\lambda_m - \lambda_k}{\delta\lambda_M - \lambda_m} \right)^{|N_k|^{(c)} - |N_k|^{(i)}} \right\}. \quad (4.9)$$

## 4.6 Aspectos computacionais

### Custo computacional do Processo Gaussiano

O custo computacional associado à simulação de um Processo Gaussiano é um aspecto extremamente relevante em algoritmos que envolvam essa tarefa. A simulação de uma distribuição normal  $n$ -dimensional tem um custo  $\mathcal{O}(n^3)$ . No nosso caso esta dimensão é dada por  $|Y| + |N|$  e, tipicamente, será muito alta levando a um alto ou até mesmo impraticável custo computacional. Existem diversas abordagens na literatura, principalmente de geoestatística, para lidar com esse problema e, dentre as mais apelativas, estão aquelas exatas, ou seja, que definem um processo (muitas vezes Gaussiano) válido para modelar a componente original e, portanto, preservam a validade do paradigma Bayesiano. Nós recorreremos a uma dessas estratégias que, basicamente, zera a correlação entre pontos a partir de uma certa distância. Esta estratégia pode ter um impacto negativo em contextos onde o modelo depende mais fortemente do processo Gaussiano, como em modelos geoestatísticos (quando o PG é parcialmente observado) e processos de Cox em que a FI é uma função contínua do PG como em Gonçalves et al. [2017b]. Porém, esse não é o caso do modelo aqui proposto e, portanto, é razoável esperar que a eficiência do modelo não seja comprometida.

A nova função de covariância de  $\beta$  é definida em função de um raio  $R$  fixo de forma que  $cov(\beta(s), \beta(s)) = \sigma^2$ ,  $cov(\beta(s), \beta(s')) = \sigma^2 \rho(|s - s'|)$  e

$$\rho(s, s') = \begin{cases} \exp\left\{-\frac{1}{2\tau^2}|s - s'|^\gamma\right\}, & \text{se } |s - s'| \leq R \\ 0, & \text{se } |s - s'| > R \end{cases}.$$

Como a abordagem tratada aqui é baseada em zerar covariâncias a partir de uma certa distância utilizando uma matriz de covariância exponencial, que é originalmente positiva-definida, isso implica que a matriz apresentada também seja positiva-definida (ver Banerjee et al. [2014]), fazendo com que o processo resultante seja um processo Gaussiano válido.

Como dito anteriormente, os parâmetros do PG serão fixos. Em particular, iremos adotar  $\mu = 0$ ,  $\sigma^2 = 1$ . Além disso, para que a definição dos demais parâmetros da função de cor-

relação e outras quantidades fixas do modelo/algoritmo, como  $\delta$  e  $R$ , possa ser padronizada e baseada nos exemplos apresentados nesta tese, iremos sempre considerar uma escala para  $S$  que tenham escala igual ou parecida com uma área de lado 10. Neste caso, e baseado nos resultados de Gonçalves et al. [2017b], adotamos  $\tau^2 = 2$  e  $\gamma = 3/2$ . A escolha do raio  $R$  será baseada na distância máxima possível entre dois pontos no espaço  $S$ , uma escolha razoável seria escolher  $R$  de tal forma que ele representa entre 5% e 10% desta distância máxima. Por exemplo, se  $S$  for  $(0, 10) \times (0, 10)$ , a maior distância entre dois pontos seria  $\sqrt{200} = 14.14$  e o valor do raio neste caso estaria entre 1 e 2.

## Passos virtuais do MCMC

Apesar da estratégia descrita acima ajudar a controlar o custo relacionado à simulação do processo Gaussiano, esse custo ainda pode ser severamente prejudicado se houver a acumulação de muitas localizações onde o PG é simulado. Em princípio, isso pode ocorrer devido a sucessivas rejeições do passo Metropolis do  $\beta$  e nos passos de  $\lambda$  e  $N$ , que tipicamente vão requerer a simulação de  $\beta$  em novas simulações.

A dimensionalidade infinita do algoritmo e o contexto de amostragem retrospectiva oferecem uma solução sofisticada e muito eficiente para esse problema. Basicamente, são adicionados passos virtuais ao algoritmo MCMC nos quais  $\beta$  é atualizado em  $S \setminus \{Y, N\}$  via *Pseudo-Marginal* e propondo do PG *a priori*. Como a probabilidade de aceitação desse passo não depende de  $\beta$  nas localizações a serem amostradas, a proposta é sempre aceita. Note, porém, que  $\beta$  não precisa ser inicialmente amostrado nessas localizações (amostragem retrospectiva) e, portanto, na prática, esse novo passo consiste simplesmente na eliminação de todos os valores do  $\beta$  atual já amostrados em  $S \setminus \{Y, N\}$ . Esse passo virtual será implementado após todo passo do *Gibbs sampling* onde o conjunto de localizações amostradas de  $\beta$  em  $S \setminus \{Y, N\}$  é não vazio e, com isso, a acumulação de pontos não ocorrerá. Mais especificamente, isso ocorrerá após toda atualização dos blocos  $\lambda$  e  $N^*$ .

## Outros aspectos importantes

A escolha do valor de  $K$  é extremamente relevante para a aplicação da metodologia proposta nesta tese. Sugerimos que essa escolha seja baseada em uma estimação empírica, de baixo custo computacional, da função intensidade. Em particular, utilizamos a função *Kernel2d* do pacote *Splancs* [Rowlingson et al., 2012], disponível no *software R* [Team et al., 2013], que utiliza suavização via kernel. A escolha do  $K$  deve considerar o *trade-off* entre ajuste e parcimônia. Deve-se ter em mente o “alcance” da distribuição de Poisson, ou seja, que parte do espaço de estados é confortavelmente acomodada pelo desvio-padrão dessa distribuição. Além disso, note que a adequação de um modelo com FI constante por partes também pode ser empiricamente avaliada com base nessa estimativa.

Outra importante etapa da metodologia proposta é a escolha dos valores iniciais do algoritmo MCMC. Uma escolha robusta que se mostrou eficiente nos exemplos a serem apresentados no Capítulo 5 é a seguinte. Os valores iniciais das intensidades são definidos como  $\lambda_k = \frac{|Y|}{\mu_S}$ ,  $\forall k$ . Já os valores iniciais de  $\beta$  e  $c$  são escolhidos dentro da escala da distribuição normal padrão e de forma que os valores de  $\beta$  são iguais para todos pontos que pertençam a mesma sub-região inicial, sendo as sub-regiões definidas com base na estimação empírica citada acima.

Apesar do processo de estimação dos parâmetros  $c$ 's estar bem estabelecido, é razoável argumentar que esses parâmetros não são fortemente identificados pelos dados e que, portanto, o processo de inferência não seja prejudicado se esses forem fixados. Naturalmente, os valores a serem fixados devem estar de acordo com a escala do PG (normal padrão) e condizentes com o valor de  $K$  fixado. Por exemplo, os valores de  $c$  podem ser os que dividem a reta real em  $K$  intervalos de mesma probabilidade sob a distribuição normal padrão.

No que diz respeito à escolha de  $\delta$ , a proposição 2 diz que a variância do *Poisson Estimator* é uma função crescente de  $\delta$  e sabe-se que, de forma geral, quanto maior for essa variância, pior é o *mixing* da cadeia *Pseudo-Marginal* [Andrieu and Vihola, 2015]. Porém, quanto maior o valor de  $\delta$ , maior é o custo computacional de cada iteração do MCMC, já que o

número médio de pontos em  $N$  é  $\mu_S(\delta\lambda_M - \lambda_m)$ . Desta forma, a escolha de  $\delta$  deve levar em conta esse *trade-off*, além dos resultados empíricos da seção 3.2.1.

Finalmente, espera-se que a coordenada  $\lambda$  seja a mais impactada negativamente pela variância do estimador da função de verossimilhança. Isso se dá por esse estimador ser mais impactado, de forma geral, por essa coordenada. Dessa forma é razoável adotar a estratégia de ser fazer múltiplas atualizações de  $\lambda$  a cada iteração do *Gibbs sampling* [Green and Mira, 2001, Zuanetti and Milan, 2016].

# Capítulo 5

## Estudos de simulação

Neste capítulo são apresentados uma coleção de exemplos simulados com a metodologia proposta nesta tese. Os objetivos gerais são investigar sua eficiência com relação a modelagem e inferência. Os objetivos específicos são vários e incluem:

1. Análise de sensibilidade com relação às escolhas de  $R$ ,  $\delta$  e  $K$ ;
2. Comparação entre os algoritmos Barker e *Pseudo-Marginal* para  $\lambda$ ;
3. Avaliar a possibilidade de se fixar os valores de  $c$ .

Todos os conjuntos de dados simulados consideram  $S = (0, 10) \times (0, 10)$ . As comparações de diferentes algoritmos são baseadas no tempo por amostra efetiva, definido como

$$\Upsilon = \frac{\text{Tempo total do algoritmo}}{\text{tamanho efetivo de amostra}}. \quad (5.1)$$

No caso do objetivo específico 3, foram testadas duas configurações distintas para  $c$  nos estudos simulados 1 e 2 e resultados praticamente indistinguíveis foram obtidos. No caso do estudo 1, foram utilizados os valores 0 e 0.5 para  $c_1$  e, no estudo 2, os valores  $c_1 = -1, c_2 = 1$  e  $c_1 = -0.5, c_2 = 0.5$ .

Com relação ao objetivo 2, os estudos de simulações nos quais foram implementados os dois algoritmos de atualização de  $\lambda$  evidenciaram um custo computacional significativamente maior quando o algoritmo Barker era utilizado (ver apêndice 7). Além disso, para se obter a versão desse algoritmo a ser utilizada para comparação com o algoritmo *Pseudo-Marginal*, é necessário fazer escolhas não triviais de parâmetros relacionados à sua implementação. Portanto, recomenda-se, de forma geral, o uso do algoritmo *Pseudo-Marginal* para amostrar  $\lambda$ .

No caso da análise de sensibilidade da especificação do raio  $R$ , estudos simulados mostraram que escolhas dentro da faixa recomendada, de 5% a 10%, retornaram resultados praticamente indistinguíveis.

As escolhas dos valores iniciais das regiões e do valor de  $K$  são feitas através de uma análise visual dos gráficos produzidos pela estimação empírica da FI via Kernel (ver Figura 7.1). Em todos os estudos, é gerada uma cadeia de tamanho 1000, com um *burn-in* de 200.

As regiões estimadas *a posteriori* serão apresentadas através de dois gráficos, que reproduzem a média e a moda *a posteriori* da função intensidade. Para isso foi feito uma grade com  $41 \times 41 = 1681$  áreas regulares na região  $S$ . O primeiro gráfico é produzido baseando-se no valor da média *a posteriori* do centróide de cada área. Já para a produção do gráfico da moda, é verificado a qual região  $k$  o centróide  $i$ , para  $i = 1, \dots, 1681$ , pertenceu mais vezes e, então, esse centróide é vinculado ao valor da média *a posteriori* de  $\lambda_k$ .

## 5.1 Estudo de simulação 1

Para avaliar o desempenho do modelo quando  $K = 2$  e realizar uma análise de sensibilidade para  $R$  e  $\delta$ , foi gerado um conjunto de dados com 290 pontos em  $S$ , com  $\lambda_1 = 5$ ,  $\lambda_2 = 2$ ,  $\mu_1 = 30$  e  $\mu_2 = 70$ . Para o mesmo conjunto de dados foram considerados os seguintes cenários:

- **Cenário 1:**  $\delta = 5$  e  $R = 1$ ;

- **Cenário 2:**  $\delta = 8$  e  $R = 1$ ;
- **Cenário 3:**  $\delta = 12$  e  $R = 1$ ;
- **Cenário 4:**  $\delta = 8$  e  $R = 1.5$ .

A análise de sensibilidade para o  $\delta$  é baseada na comparação dos cenários 1, 2 e 3 e a análise para o raio  $R$ , na comparação dos cenários 2 e 4. A Tabela 5.1 traz o valor de  $\Upsilon$  para as cadeias dos  $\lambda_k$ 's e de  $\hat{l}(\theta, y)$ , nos quatros cenários, sendo

$$\hat{l}(\theta, y) = \log(\hat{M}) + \sum_{k=1}^K |Y_k| \log(\lambda_k). \quad (5.2)$$

De forma geral, o cenário 2 apresenta os melhores resultados. O aumento de  $\delta = 5$  para  $\delta = 8$  ocasionou uma melhora no *mixing* das cadeias que compensou o aumento do custo computacional. Já o aumento de  $\delta = 8$  para  $\delta = 12$  não gerou uma melhora do *mixing* que compensasse o aumento do custo computacional, como pode ser visto na Figura 7.2 no apêndice.

Em relação à comparação na análise de sensibilidade do  $R$ , o cenário 2 é mais eficiente computacionalmente do que cenário 4. Além disso, as estimativas obtidas para a FI nos dois cenários são bem parecidas, sendo um pouco mais suave, como era de se esperar, no cenário 4 (ver Figura 7.3 no apêndice).

Os trace-plots das cadeias dos parâmetros do Cenário 2 estão disponíveis na Figura 7.4, indicando convergência da cadeia. As taxas de aceitação de cada coordenada do MCMC e os valores dos parâmetros fixos são apresentados na primeira coluna das Tabelas 7.1 e 7.2 no apêndice, respectivamente.

A região verdadeira e a função intensidade estimada pela média e moda são apresentados na Figura 5.1. A média e o desvio padrão *a posteriori* de  $(\lambda_1, \lambda_2)$ , são  $(5.286, 2.065)$  e  $(0.443, 0.168)$ , respectivamente. Note que o modelo foi eficiente na estimação dos parâmetros e na identificação das regiões verdadeiras.

		Cenário 1	Cenário 2	Cenário 3	Cenário 4
Tempo (min)		37	154	507	688
$\Upsilon$ (seg)	$\lambda_1$	0.179	0.361	1.308	1.74
	$\lambda_2$	0.077	0.192	0.634	3.326
	$\hat{\mathbf{I}}(\theta, \mathbf{y})$	1.414	1.357	13.615	31.960

Tabela 5.1: Valores do *Tempo* (min) e de  $\Upsilon$  nos cenários 1, 2, 3 e 4.

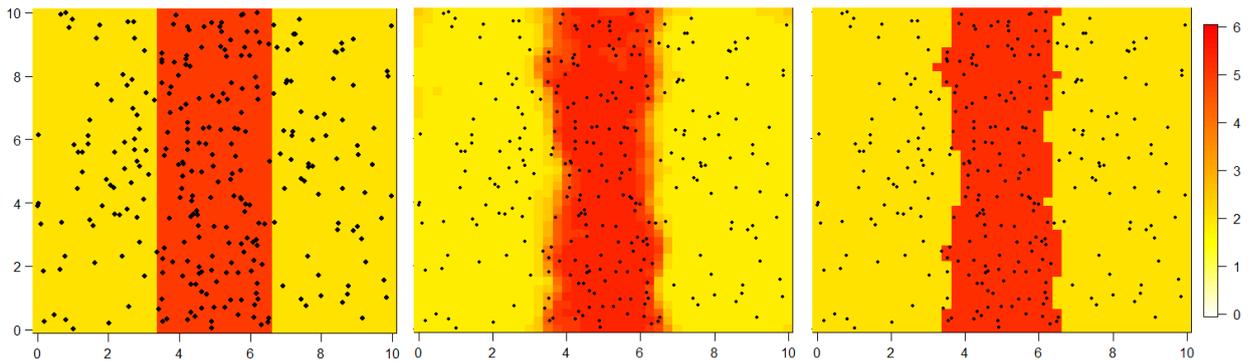


Figura 5.1: Região verdadeira (esquerda) e os gráficos das estimações da média (meio) e da moda (direita) *a posteriori* da FI para o cenário 2.

## 5.2 Estudo de simulação 2

Para avaliar o desempenho do modelo quando  $K = 3$  e realizar uma análise de sensibilidade para a escolha do valor de  $K$ , foi gerado um conjunto de dados com 333 pontos em  $S$ , onde  $\lambda_1 = 9$ ,  $\lambda_2 = 4$ ,  $\lambda_3 = 1.55$ ,  $\mu_1 = 16$ ,  $\mu_2 = 24$  e  $\mu_3 = 60$ . Para esse mesmo conjunto de dados o modelo proposto foi ajustado com  $K = 2$ ,  $K = 3$  e  $K = 4$ .

As taxas de aceitação de cada coordenada do MCMC e os valores dos parâmetros fixos do caso  $K = 3$  podem ser vistos na segunda coluna das Tabelas 7.1 e 7.2 no apêndice, respectivamente. As regiões e as funções intensidades estimadas pela média *a posteriori* para os três casos são apresentadas na Figura 5.3 e a região verdadeira pode ser vista na Figura 5.2. A média e o desvio padrão *a posteriori* do vetor  $\lambda$  são apresentados na Tabela 5.2.

Note que para o caso  $K = 3$ , o modelo foi eficiente na estimação dos parâmetros e na identificação das sub-regiões verdadeiras. Já para o caso  $K = 2$ , o modelo identificou corretamente a região com maior área (com intensidade verdadeira igual a 1.55), mas a segunda região foi identificada como a junção das duas regiões com áreas menores. Já os resultados para  $K = 4$  sugerem fortemente que o modelo com  $K = 3$  é adequado, isso pode ser notado tanto pelo gráfico da estimação *a posteriori* da função intensidade, quanto pela proximidade das estimações do  $\lambda_3$  e  $\lambda_4$ .

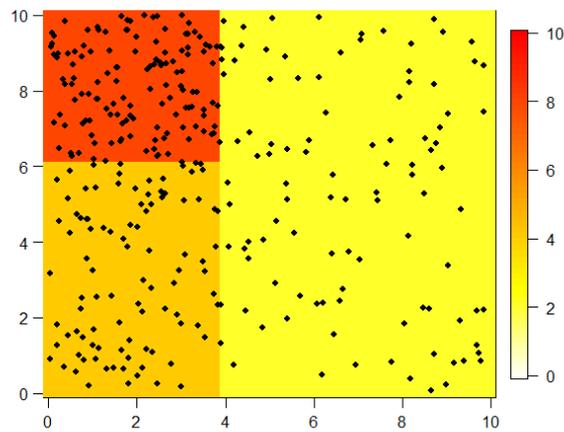


Figura 5.2: Região e FI real para o estudo de simulação 2.

Veja que a estrutura do modelo proposto nesta tese nos permite identificar com mais facilidade se um  $K$  maior do que o necessário está sendo utilizado. Portanto, o ajuste com modelo com mais de um  $K$  pode ser considerado, juntamente com a análise empírica da FI estimada via kernel, para escolher o  $K$  definitivo a ser utilizado.

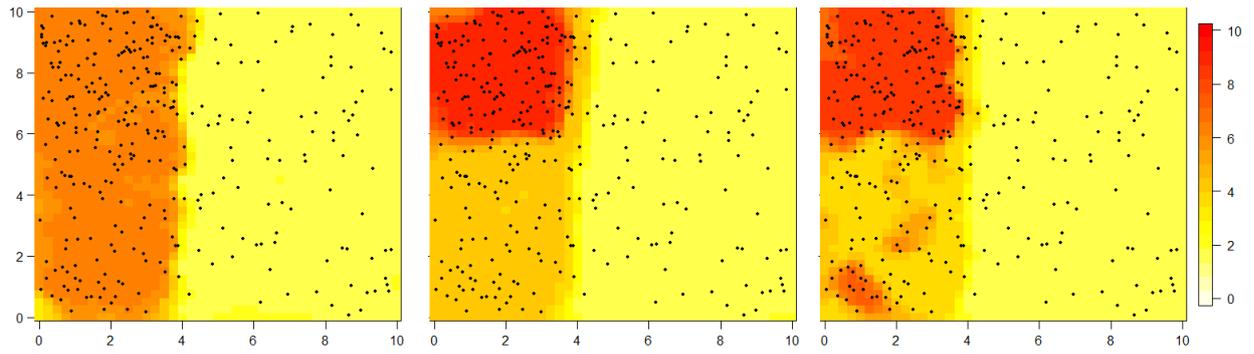


Figura 5.3: Gráficos da estimação da média *a posteriori* nos casos em que  $K = 2$  (esquerda),  $K = 3$  (meio) e  $K = 4$  (direita).

		Caso $K = 2$	Caso $K = 3$	Caso $K = 4$
$\lambda_1$	Média	6.102	9.385	9.425
	Desvio	0.382	0.768	0.836
$\lambda_2$	Média	1.602	3.768	3.842
	Desvio	0.165	0.374	0.386
$\lambda_3$	Média		1.561	1.611
	Desvio		0.1645	0.2685
$\lambda_4$	Média			1.661
	Desvio			0.234

Tabela 5.2: Média e desvio-padrão *a posteriori* dos  $\lambda$ 's para os casos com  $K = 2$ ,  $K = 3$  e  $K = 4$ .

### 5.3 Estudo de simulação 3

Para avaliar o desempenho do modelo quando  $K = 2$ , mas  $S$  é dividido em mais que duas sub-regiões, foi gerado um conjunto de dados com 408 pontos, onde  $\lambda_1 = 10$ ,  $\lambda_2 = 2$ ,  $\mu_1 = 26$  e  $\mu_2 = 74$ .

As taxas de aceitação de cada coordenada do MCMC e os valores dos parâmetros fixos podem ser vistos na terceira coluna das Tabelas 7.1 e 7.2 no apêndice, respectivamente.

A região verdadeira e a função intensidade estimada pela média e moda *a posteriori* são apresentados na Figura 5.4. A média e o desvio padrão *a posteriori* de  $(\lambda_1, \lambda_2)$ , são  $(10.081, 1.987)$  e  $(0.654, 0.169)$ , respectivamente. Note que o modelo foi eficiente na estimação dos parâmetros e na identificação das sub-regiões verdadeiras.

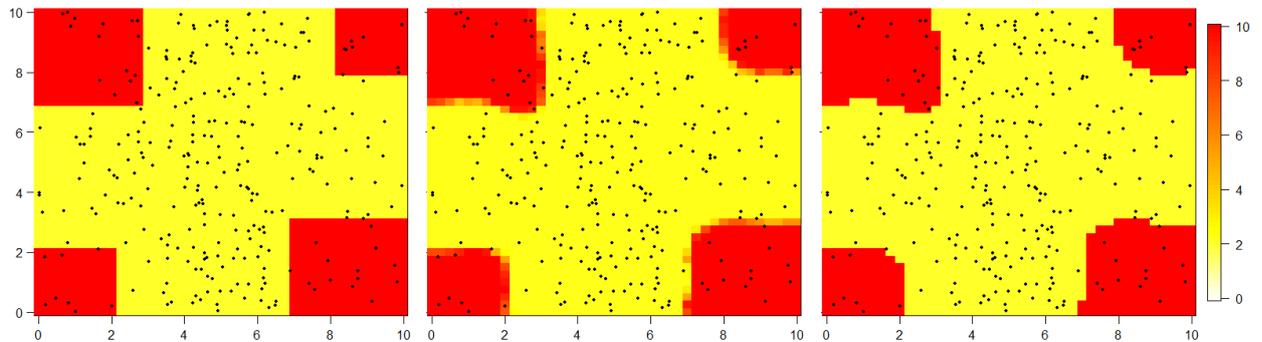


Figura 5.4: Região real (esquerda) e os gráficos das estimações da média (meio) e da moda (direita) *a posteriori* da FI para o estudo simulado 3.

## 5.4 Estudo de simulação 4

Para avaliar o desempenho do modelo quando  $K = 4$  foi gerado um conjunto de dados com 408 pontos em  $S$ , onde  $\lambda_1 = 2$ ,  $\lambda_2 = 3.52$ ,  $\lambda_3 = 5$ ,  $\lambda_4 = 10$ ,  $\mu_1 = 30$ ,  $\mu_2 = 24$ ,  $\mu_3 = 25$  e  $\mu_4 = 21$ .

As taxas de aceitação de cada coordenada do MCMC e os valores dos parâmetros fixos podem ser vistos na quarta coluna das Tabelas 7.1 e 7.2 no apêndice, respectivamente.

A região verdadeira e a função intensidade *a posteriori* estimada pela média e moda são apresentados na Figura 5.5. A média e o desvio padrão *a posteriori* de  $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ , são  $(2.164, 3.216, 5.275, 11.089)$  e  $(0.283, 0.380, 0.467, 0.898)$ , respectivamente. Note que o modelo foi eficiente na estimação dos parâmetros e na identificação das sub-regiões verdadeiras.

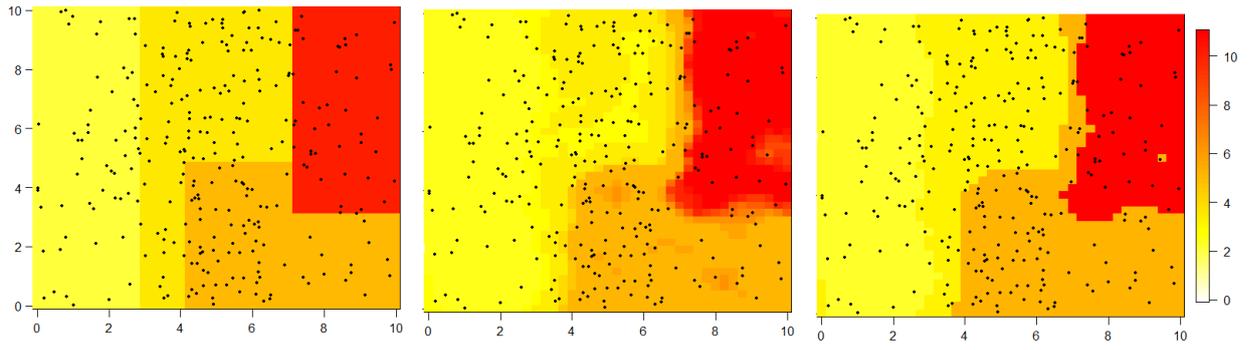


Figura 5.5: Região real (esquerda) e os gráficos das estimações da média (meio) e da moda (direita) *a posteriori* da FI para o Simulação 4.

# Capítulo 6

## Aplicação

Neste capítulo o modelo proposto nesta tese será aplicado em três conjuntos de dados reais. As mesmas especificações de tamanho de cadeia e *burn-in* dos estudos simulados são consideradas nos exemplos deste capítulo.

### 6.1 Aplicação 1

Nesta seção será analisado um conjunto de dados referente às localizações de árvores em Lansing Woods, Condado de Clinton, Michigan, EUA. Esses dados estão disponíveis no pacote `spatstat` do R (Baddeley et al. [2015]) e consiste nas localizações de 448 carvalhos brancos (White oak) em uma área com tamanho original ( $924 \times 924$  pés), que redimensionamos para  $(0, 10) \times (0, 10)$ .

Para a escolha do valor de  $K$ , foi feita uma análise visual do gráfico reproduzido pela estimação empírica da FI via Kernel (ver Figura 7.10). Como não ficou claro se o mais indicado seria um  $K = 3$  ou  $K = 4$ , o modelo foi ajustado para os dois casos.

A média e o desvio padrão *a posteriori* do vetor  $\lambda$  são apresentados na Tabela 6.1. Note que, para o caso  $K = 4$ , o modelo identificou que a região  $S$  é particionado em 3 regiões,

isso pode ser notado pela proximidade das estimações do  $\lambda_1$  e  $\lambda_2$ , sugerindo que o modelo com  $K = 3$  é mais adequado.

		Caso $K = 3$	Caso $K = 4$
$\lambda_1$	Média	2.337	2.738
	Desvio	0.252	0.295
$\lambda_2$	Média	4.839	2.225
	Desvio	0.373	0.295
$\lambda_3$	Média	7.771	5.069
	Desvio	0.594	0.406
$\lambda_4$	Média		7.972
	Desvio		0.609

Tabela 6.1: Média e o desvio padrão *a posteriori* dos  $\lambda$ 's para os casos com  $K = 3$  e  $K = 4$  da Aplicação 1.

As taxas de aceitação de cada coordenada do MCMC e os valores dos parâmetros fixos para o caso  $K = 3$  podem ser vistos nas primeiras colunas das Tabelas 7.3 e 7.4, respectivamente.

A função intensidade *a posteriori* estimada no caso  $K = 3$  pela média e moda são apresentados na Figura 6.1. Note que mesmo o modelo sendo constante por partes, existe uma suavidade entre as fronteiras das partições.

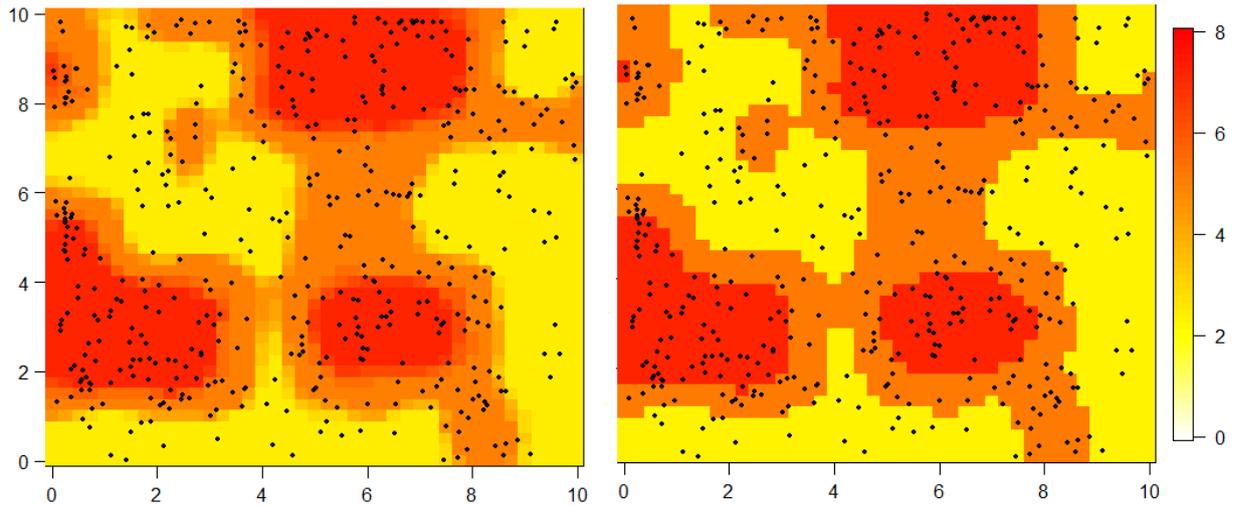


Figura 6.1: Gráficos das estimações da média (esquerda) e da moda (direita) *a posteriori* da FI para a Aplicação 1 (caso  $K = 3$ ).

### 6.1.1 Comparação da qualidade da predição entre a metodologia proposta e um modelo em que a FI varia continuamente

Este estudo tem como objetivo comparar a metodologia proposta nesta tese com uma que considera um modelo em que a função intensidade é uma função contínua de um processo Gaussiano. Em particular, consideramos a metodologia proposta em Gonçalves and Gamerman [2018], que também executa inferência exata na classe de modelos proposta.

Essa comparação é feita baseada na estimação da esperança *a posteriori* da FI integrada em determinadas regiões. Em particular, consideramos as regiões  $S$ ,  $S_1 = (5, 7) \times (8, 10)$  e  $S_2 = (8, 10) \times (4.5, 6.5)$ . Esse tipo de estimação é útil, por exemplo, quando se pretende fazer predição de replicações do processo observado. Apesar da intratabilidade analítica da função de interesse, estimadores de Monte Carlo não-viesados podem ser obtidos utilizando a metodologia apresentada em Gonçalves and Gamerman [2018] (seção 4.3), baseada em

variáveis auxiliares uniformemente distribuídas nas áreas de interesse. O estimador não-viesado para a FI integrada em uma região  $S_j$  é o seguinte:

$$\mu(S_j)\lambda(U), \tag{6.1}$$

sendo  $\mu(S_j)$  a área de  $S_j$  e  $U \sim \mathcal{U}(S_j)$ . Com o intuito de diminuir a variância *a posteriori* do estimador obtido, a área de interesse é dividida em quadrados regulares de lado 1 e uma variável uniforme em cada quadrado é introduzida. Os resultados são apresentados na tabela 6.2.

As estimativas pontuais são relativamente parecidas para os dois modelos, porém, mais próximas do número real de pontos nas respectivas regiões para o modelo proposto nesta tese. Esse modelo também apresenta menores variâncias, como já se esperava, dado sua estrutura mais simples.

	Y	Modelo proposto		Modelo contínuo	
		Média	Desvio	Média	Desvio
$S$	448	451.12	23.36	449.82	25.84
$S_1$	27	31.08	2.38	22.25	3.19
$S_2$	9	9.37	1.03	12.39	4.49

Tabela 6.2: Média e desvio-padrão *a posteriori* do estimador.

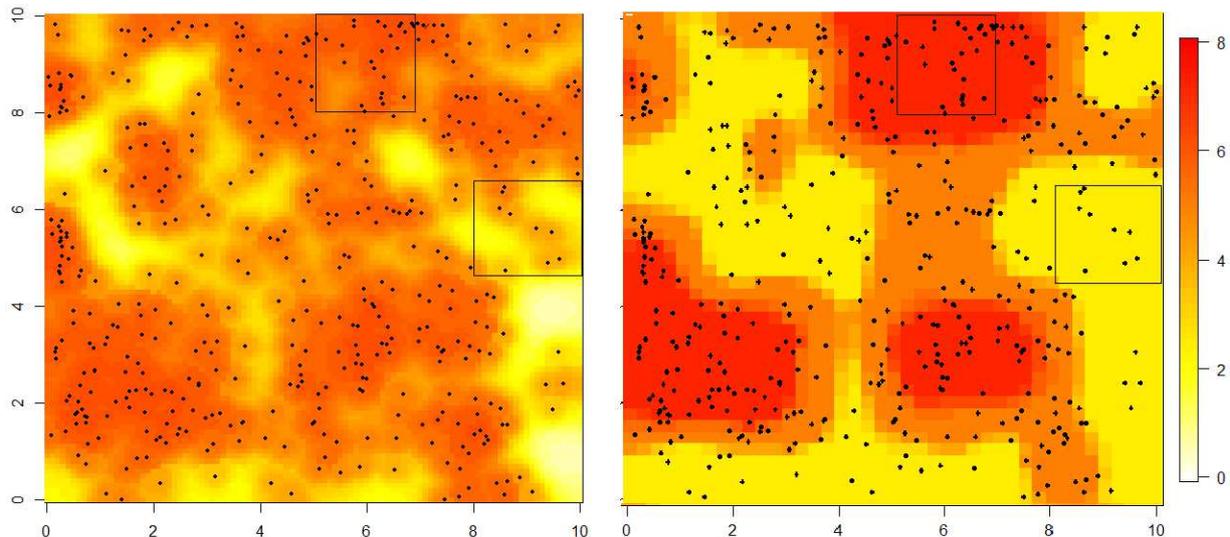


Figura 6.2: Gráficos da FI estimada pelo modelo proposto e pelo modelo contínuo de Gonçalves and Gamerman [2018].

## 6.2 Aplicação 2

Nesta seção será analisado o conjunto de dados que registram as localizações de 584 pinheiros (*Pinus palustris*) em uma região de  $200 \times 200$  metros no sul da Geórgia (EUA). Esses dados também estão disponíveis no pacote `spatstat` do R (Baddeley et al. [2015]), com área  $S$  redimensionada para  $(0, 10) \times (0, 10)$ .

A escolha dos valores iniciais das regiões e do valor de  $K$  são feitas através de uma análise visual dos gráficos reproduzidos pela estimação empírica da FI via Kernel (ver Figura 7.10). Como na aplicação anterior, houve uma dúvida em relação à escolha do  $K$ . À vista disso, o modelo foi ajustado para  $K = 3$  e  $K = 4$ . A média e o desvio padrão *a posteriori* da FI para os dois casos são apresentados na Tabela 6.3 e, como os valores estimados para dois  $\lambda$ 's intermediários do caso  $K = 4$  são parecidos, considerando-se as escalas das respectivas distribuições Poisson, optamos pelo modelo com  $K = 3$ .

		Caso $K = 3$	Caso $K = 4$
$\lambda_1$	Média	3.814	2.833
	Desvio	0.252	0.390
$\lambda_2$	Média	5.002	4.117
	Desvio	0.551	0.461
$\lambda_3$	Média	11.667	5.381
	Desvio	0.721	0.688
$\lambda_4$	Média		10.686
	Desvio		0.697

Tabela 6.3: Média e desvio-padrão *a posteriori* dos  $\lambda$ 's para os casos com  $K = 3$  e  $K = 4$  da Aplicação 2.

As taxas de aceitação de cada coordenada do MCMC e os valores dos parâmetros fixos para o caso  $K = 3$  podem ser vistos na segunda coluna das Tabelas 7.3 e 7.4, respectivamente. A função intensidade *a posteriori* estimada (caso  $K = 3$ ) pela média e moda são apresentados na Figura 6.3.

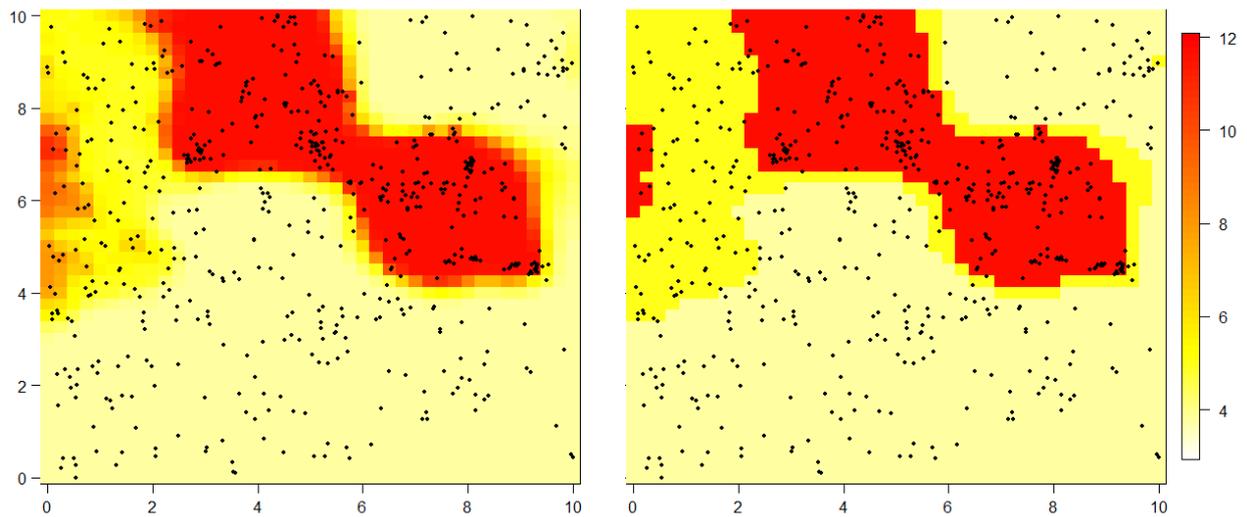


Figura 6.3: Gráficos das estimações da média (esquerda) e da moda (direita) *a posteriori* da FI para o Aplicação 2 (caso  $K = 3$ ).

## 6.3 Aplicação 3

Nesta seção será analisado o conjunto de dados que registram as localizações de 678 partículas decorrentes da fabricação de um filtro de bronze pelo método de sinterização, obtido por Bernhardt et al. [1997]. Para a coleta dos dados, um filtro é cortado em uma área retangular  $18 \times 7mm$ , identificando a posição e o raio de todas as partículas. Essas informações estão disponíveis no pacote `spatstat` do R (Baddeley et al. [2015]), conforme mostra a Figura 6.4.

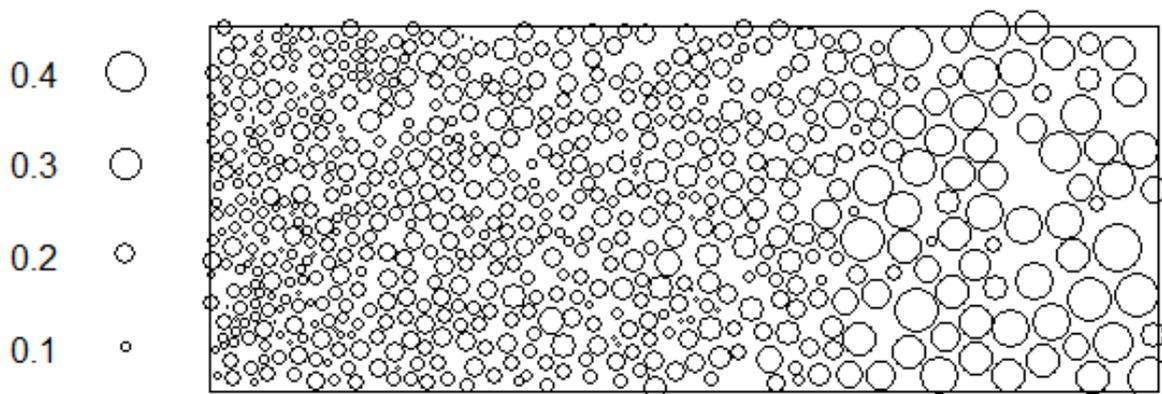


Figura 6.4: Distribuição das partículas de bronze disponível do pacote `spatstat` do R, com os tamanhos da partículas proporcionais ao seu raio.

Para manter a relação da escala original, redimensionamos a área  $S$  para  $(0, 10) \times (0, 4)$ . A escolha dos valores iniciais das regiões e do valor de  $K$  são feitas através de uma análise visual dos gráficos reproduzidos pela estimação empírica da FI via Kernel (ver Figura 7.10). O gráfico reproduzido empiricamente sugere um  $K = 4$ . No entanto, como a região central apresentou cores parecidas, considerando as escalas das respectivas distribuições Poisson, ajustamos o modelo para  $K = 3$ ,  $K = 4$  e  $K = 5$ .

A média e o desvio padrão *a posteriori* da FI para os dois casos são apresentados na Tabela 6.4. Note que os dados foram bem ajustados para o caso  $K = 5$ , isso pode ser notado pela diferença nas estimações dos  $\lambda$ 's, sugerindo que esta configuração é a mais adequada.

		<b>Caso <math>K = 3</math></b>	<b>Caso <math>K = 4</math></b>	<b>Caso <math>K = 5</math></b>
$\lambda_1$	Média	5.824	6.427	6.408
	Desvio	0.803	0.685	0.736
$\lambda_2$	Média	15.636	15.712	14.691
	Desvio	0.993	1.602	1.428
$\lambda_3$	Média	27.191	19.949	19.355
	Desvio	1.467	1.937	1.942
$\lambda_4$	Média		27.131	24.012
	Desvio		1.616	1.801
$\lambda_5$	Média			35.359
	Desvio			2.852

Tabela 6.4: Média e desvio-padrão *a posteriori* dos  $\lambda$ 's para os casos com  $K = 3$ ,  $K = 4$  e  $K = 5$  da Aplicação 3.

As taxas de aceitação de cada coordenada do MCMC e os valores dos parâmetros fixos para o caso  $K = 5$  podem ser vistos na terceira coluna das Tabelas 7.3 e 7.4, respectivamente.

A função intensidade *a posteriori* estimada (caso  $K = 4$ ) pela média e moda são apresentados na Figura 6.5. Note que existe uma divisão natural onde cada partição possui uma maior concentração de partículas com um mesmo raio, além disso o número de partículas diminuem à medida que os diâmetros aumentam. Esses tipos de informações podem ser úteis para avaliar o processo de sinterização.

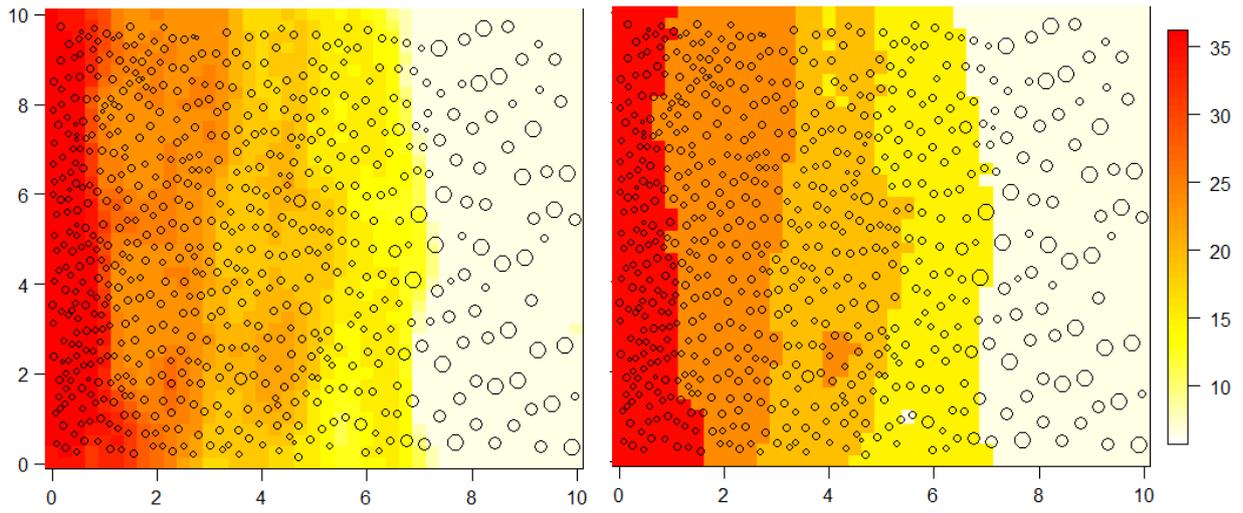


Figura 6.5: Gráficos das estimações da média (esquerda) e da moda (direita) *a posteriori* da FI para o Aplicação 3 (caso  $K = 5$ ).

# Capítulo 7

## Considerações Finais

Nesta tese, foi proposta uma metodologia Bayesiana para modelagem e inferência sobre padrões de pontos espaciais para os quais seja razoável assumir as propriedades probabilísticas do processo de Poisson, com a suposição adicional de que a função intensidade seja constante por partes. O modelo proposto consiste de um processo de Cox onde a FI é definida pelas curvas de nível de um processo Gaussiano latente, o que permite estruturas bem flexíveis para particionar o espaço considerado. Além disso, a metodologia para se fazer inferência teve como principal paradigma, a inexistência de qualquer erro/aproximação baseado em discretizações do espaço contínuo.

Foi proposto um algoritmo MCMC infinito-dimensional e eficiente o bastante para ser aplicado em problemas de tamanho razoável. O algoritmo utiliza técnicas recentes de simulação estocástica, como *Pseudo-Marginal Metropolis*, amostragem retrospectiva e *Poisson Estimator*. Em particular, a amostragem retrospectiva permitiu uma considerável redução no custo computacional através de passos virtuais ao longo do algoritmo.

Através de estudos simulados, mostrou-se que a metodologia proposta permite uma solução eficiente para modelar e estimar diferentes configurações de regiões. Foram também realizadas análises de sensibilidade com respeito à escolha de algumas especificações relacionadas ao modelo e ao MCMC -  $\delta$ ,  $R$  e  $K$ , tendo sido obtidos resultados esclarecedores.

Os estudos aplicados apresentaram resultados satisfatórios, reforçando que, na prática, podem existir problemas onde os dados são adequadamente acomodados por um processo com função intensidade constante por partes. Por meio de uma comparação entre a metodologia proposta e uma na qual a FI varia continuamente como função de um PG, foram obtidas previsões mais precisas com a primeira.

Todos os códigos foram cuidadosamente desenvolvidos na linguagem Ox Doornik [2009] e os gráficos foram construídos através do software R (Team et al. [2013]). Os códigos estão disponíveis mediante solicitação à autora desta tese.

O trabalho aqui proposto motiva o desenvolvimento de pesquisas futuras que considerem, por exemplo, a extensão da metodologia proposta para o caso espaço-temporal, de tal forma que os dados possam ser vistos como uma série temporal em tempo discreto de processos espaciais, referente a uma região em comum. A dependência temporal do modelo seria definida através do processo Gaussiano. Além disso, outro ponto a ser estudado seria a viabilidade de se estimar o parâmetro  $K$ , apesar dos resultados aqui apresentados mostrarem total viabilidade em se aplicar a metodologia proposta quando este é fixado.

# Apêndice

## Prova da Proposição 1

*Demonstração.* Defina  $I_{nk} = I_k(s_n)$  como a função indicadora do  $n$ -ésimo ponto de  $N$   $s_n \in S_k$  e  $I = (I_1, \dots, I_{|N|})$ , onde  $I_n = (I_{n1}, \dots, I_{nK}) \sim Mult\left(1, \frac{\mu_1}{\mu_S}, \dots, \frac{\mu_K}{\mu_S}\right)$  e, portanto,  $E(I_{nk}) = \frac{\mu_k}{\mu_S}$  e  $\mu_S I_{nk}$  é um estimador não-viesado para  $\mu_k$ . Assim:

$$\begin{aligned}
 E_{|N|,I}[\hat{M}] &= E_{|N|,I} \left[ e^{-\mu_S \lambda_m} \prod_{k=1}^K \left( \frac{\delta \lambda_M - \lambda_k}{\delta \lambda_M - \lambda_m} \right)^{|N_k|} \right] \\
 &= E_{|N|,I} \left[ e^{-\mu_S \lambda_m} \prod_{k=1}^K \left( \frac{\delta \lambda_M - \lambda_k}{\delta \lambda_M - \lambda_m} \right)^{\sum_{n=1}^{|N|} I_k(s_n)} \right] \\
 &= E_{|N|,I} \left[ e^{-\mu_S \lambda_m} \prod_{n=1}^{|N|} \left( \frac{\delta \lambda_M - \sum_{k=1}^K I_{nk} \lambda_k}{\delta \lambda_M - \lambda_m} \right) \right] \\
 &= e^{-\mu_S \lambda_m} E_{|N|} \left[ E_{I||N|} \prod_{n=1}^{|N|} \left[ \left( \frac{\mu_S \delta \lambda_M - \sum_{k=1}^K \mu_S I_{nk} \lambda_k}{\mu_S (\delta \lambda_M - \lambda_m)} \right) \right] \right] \\
 &= e^{-\mu_S \lambda_m} E_{|N|} \left[ \left( \frac{\mu_S \delta \lambda_M - \sum_{k=1}^K \mu_k \lambda_k}{\mu_S (\delta \lambda_M - \lambda_m)} \right)^{|N|} \right] \\
 &= e^{-\mu_S (\lambda_m + \delta \lambda_M - \lambda_m)} \sum_{j=0}^{\infty} \frac{\left( \mu_S \delta \lambda_M - \sum_{k=1}^K \mu_k \lambda_k \right)^j}{j!} \\
 &= e^{-\sum_{k=1}^K \mu_k \lambda_k} = M.
 \end{aligned}$$

□

## Prova da Proposição 2

*Demonstração.* Vamos calcular a variância de  $\hat{M}_1 = \prod_{k=1}^K \left( \frac{\delta\lambda_M - \lambda_k}{\delta\lambda_M - \lambda_m} \right)^{|N_k|}$ .

Para cada  $k = 1, \dots, K$ , temos que  $N_k \sim PP(\lambda^*)$  e, portanto,  $|N_k| \sim Poisson(\mu_k \lambda^*)$ .

Temos que

$$\begin{aligned} Var[\hat{M}_1] &= E[\hat{M}_1^2] - E[\hat{M}_1] \\ &= \prod_{k=1}^K E \left[ \left( \frac{\delta\lambda_M - \lambda_k}{\delta\lambda_M - \lambda_m} \right)^{2|N_k|} \right] - \prod_{k=1}^K E \left[ \left( \frac{\delta\lambda_M - \lambda_k}{\delta\lambda_M - \lambda_m} \right)^{|N_k|} \right] \\ E \left[ \left( \frac{\delta\lambda_M - \lambda_k}{\delta\lambda_M - \lambda_m} \right)^{2|N_k|} \right] &= \sum_{j=0}^{\infty} \left( \frac{\delta\lambda_M - \lambda_k}{\delta\lambda_M - \lambda_m} \right)^{2j} \frac{\exp\{-\mu_k(\delta\lambda_M - \lambda_m)\} [\mu_k(\delta\lambda_M - \lambda_m)]^j}{j!} \\ &= \sum_{j=0}^{\infty} \frac{\left[ \frac{(\delta\lambda_M - \lambda_k)^2 \mu_k}{(\delta\lambda_M - \lambda_m)} \right]^j}{j!} \exp\{-\mu_k(\delta\lambda_M - \lambda_m)\} \\ &= \exp \left\{ \frac{\mu_k(\delta\lambda_M - \lambda_k)^2}{\delta\lambda_M - \lambda_m} - \mu_k(\delta\lambda_M - \lambda_m) \right\} \\ E \left[ \left( \frac{\delta\lambda_M - \lambda_k}{\delta\lambda_M - \lambda_m} \right)^{|N_k|} \right] &= \sum_{j=0}^{\infty} \left( \frac{\delta\lambda_M - \lambda_k}{\delta\lambda_M - \lambda_m} \right)^j \frac{\exp\{-\mu_k(\delta\lambda_M - \lambda_m)\} [\mu_k(\delta\lambda_M - \lambda_m)]^j}{j!} \\ &= \sum_{j=0}^{\infty} \frac{[\mu_k(\delta\lambda_M - \lambda_k)]^j}{j!} \exp\{-\mu_k(\delta\lambda_M - \lambda_m)\} \\ &= \exp \{ \mu_k(\delta\lambda_M - \lambda_k) - \mu_k(\delta\lambda_M - \lambda_m) \} \\ &= \exp \{ -\mu_k(\lambda_k - \lambda_m) \} \end{aligned}$$

Assim,

$$\prod_{k=1}^K E \left[ \left( \frac{\delta\lambda_M - \lambda_k}{\delta\lambda_M - \lambda_m} \right)^{2|N_k|} \right] = \exp \left\{ - \sum_{k=1}^K \mu_k(\delta\lambda_M - \lambda_m) \left[ 1 - \left( \frac{\delta\lambda_M - \lambda_k}{\delta\lambda_M - \lambda_m} \right)^2 \right] \right\}$$

$$\prod_{k=1}^K E \left[ \left( \frac{\delta\lambda_M - \lambda_k}{\delta\lambda_M - \lambda_m} \right)^{|N_k|} \right] = \exp \left\{ - \sum_{k=1}^K \mu_k(\lambda_k - \lambda_m) \right\}$$

Portanto,

$$\begin{aligned} \text{Var}(\hat{M}_1) &= \exp \left\{ - \sum_{k=1}^K \mu_k (\delta \lambda_M - \lambda_m) \left[ 1 - \left( \frac{\delta \lambda_M - \lambda_k}{\delta \lambda_M - \lambda_m} \right)^2 \right] \right\} \\ &\quad - \exp \left\{ -2 \sum_{k=1}^K \mu_k (\lambda_k - \lambda_m) \right\} \end{aligned}$$

e, conseqüentemente,

$$\text{Var}(\hat{M}) = \exp \{ -2\mu_S \lambda_m \} \text{Var}(\hat{M}_1).$$

Finalmente,

$$\frac{\partial \text{Var}(\hat{M}_1)}{\partial \delta} = \exp(\kappa) \left[ \sum_{k=1}^K \mu_k \lambda_M \left( \left( \frac{\delta \lambda_M - \lambda_k}{\delta \lambda_M - \lambda_m} \right)^2 - 2(\delta \lambda_M - \lambda_k) - 1 \right) \right] < 0$$

onde  $\kappa \in \mathbb{R}$ .

□

### Prova da Proposição 3

*Demonstração.* Defina  $I_{ik} = I_k(z_i)$  como a função indicadora do  $i$ -ésimo ponto de  $Z$   $z_i \in S_k$  e  $I = (I_1, \dots, I_{|Z|})$ , onde  $I_i = (I_{i1}, \dots, I_{iK}) \sim \text{Mult} \left( 1, \frac{\mu_1}{\mu_S}, \dots, \frac{\mu_K}{\mu_S} \right)$  e, portanto,  $E(I_{ik}) = \frac{\mu_k}{\mu_S}$  e  $\mu_S I_{ik}$  é um estimador não-viesado para  $\mu_k$ . Assim:

$$\begin{aligned} E_{|Z|, I}[\hat{M}_k] &= E_{|Z|, I} \left[ \left( 1 - \frac{1}{\delta^*} \right)^{|Z_k|} \right] \\ &= E_{|Z|, I} \left[ \left( \frac{\delta^* \mu_S d_k - \mu_S d_k}{\delta^* \mu_S d_k} \right)^{|Z_k|} \right] \\ &= E_{|Z|} \left[ E_{I||Z|} \left[ \prod_{i=1}^{|Z|} \left( \frac{\delta^* \mu_S d_k - \mu_S I_{ik} d_k}{\delta^* \mu_S d_k} \right) \right] \right] \\ &= E_{|Z|} \left[ \left( \frac{\delta^* \mu_S d_k - \mu_k d_k}{\delta^* \mu_S d_k} \right)^{|Z|} \right] \\ &= e^{-\delta^* \mu_S d_k} \sum_{j=0}^{\infty} \frac{(\delta^* \mu_S d_k - \mu_k d_k)^j}{j!} = e^{-\mu_k d_k} = M_k \end{aligned}$$

□

## Prova reversibilidade da cadeia de $N^*$ :

*Demonstração.* Dizemos que uma distribuição  $\pi$  é reversível para a cadeia de Markov com distribuição de transição  $q$  se

$$\pi(x)q(x, y) = \pi(y)q(y, x) \quad \forall x, y \in \mathcal{X},$$

onde  $\mathcal{X}$  é o espaço de estados da cadeia.

No nosso caso, a distribuição  $\pi$  é uma Poisson( $A$ ) e a distribuição de transição definida em (4.3) é tal que

$$\pi(n)q(n, n+1) = \frac{A^n e^{-A}}{n!} \frac{A}{A+n+1} = \frac{A^{n+1} e^{-A}}{n!(A+n+1)}$$

e

$$\pi(n+1)q(n+1, n) = \frac{A^{n+1} e^{-A}}{(n+1)!} \frac{n+1}{A+(n+1)} = \frac{A^{n+1} e^{-A}}{n!(A+n+1)},$$

o que prova o resultado desejado.

□

## Resultados estudos de simulação

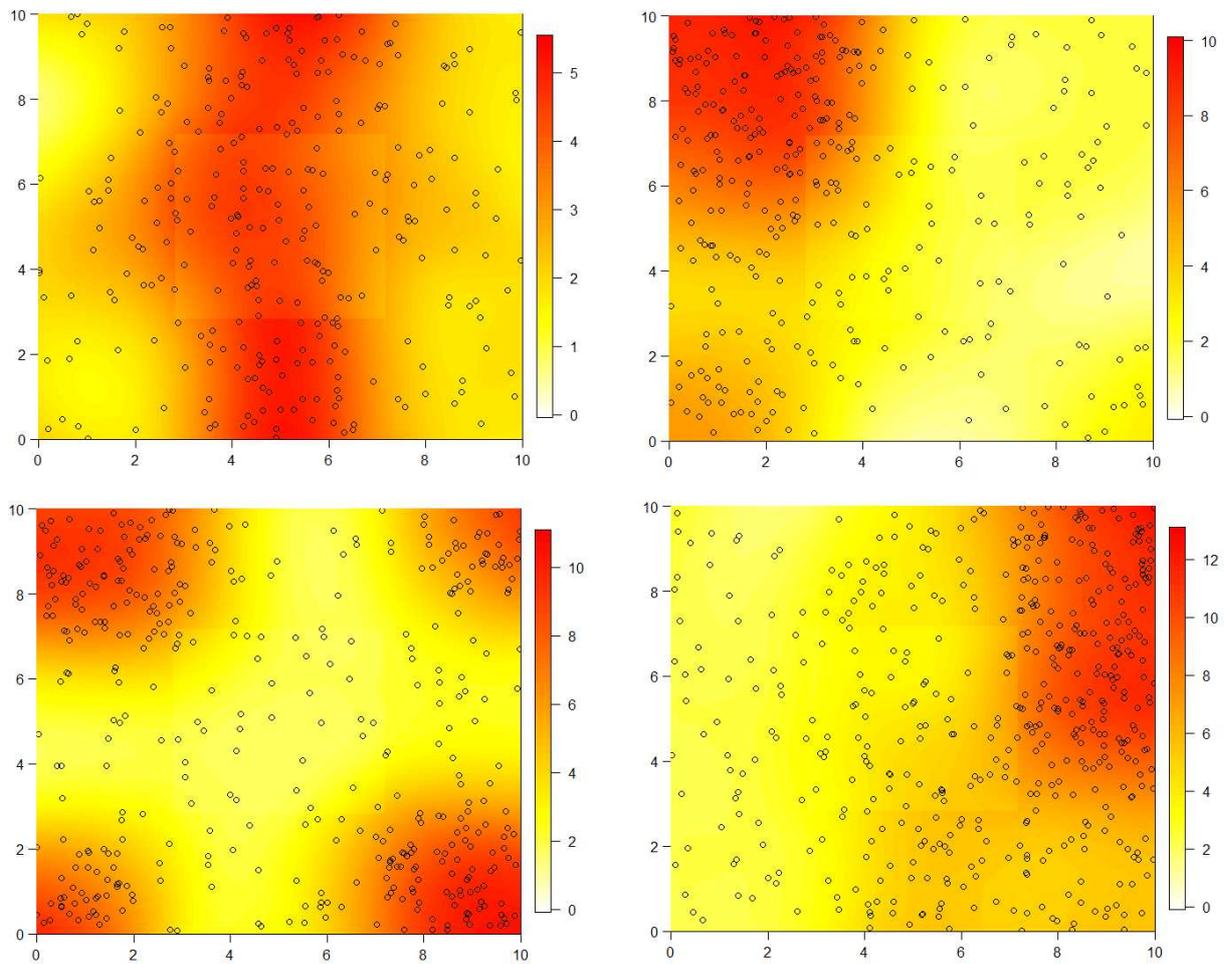


Figura 7.1: Estimação empírica via Kernel da FI para a escolha do valor de  $K$  e das regiões iniciais da Simulação 5.1 (esquerda-1º linha), Simulação 5.2 (direita-1º linha), Simulação 5.3 (esquerda-2º linha) e Simulação 5.4 (direita-2º linha).

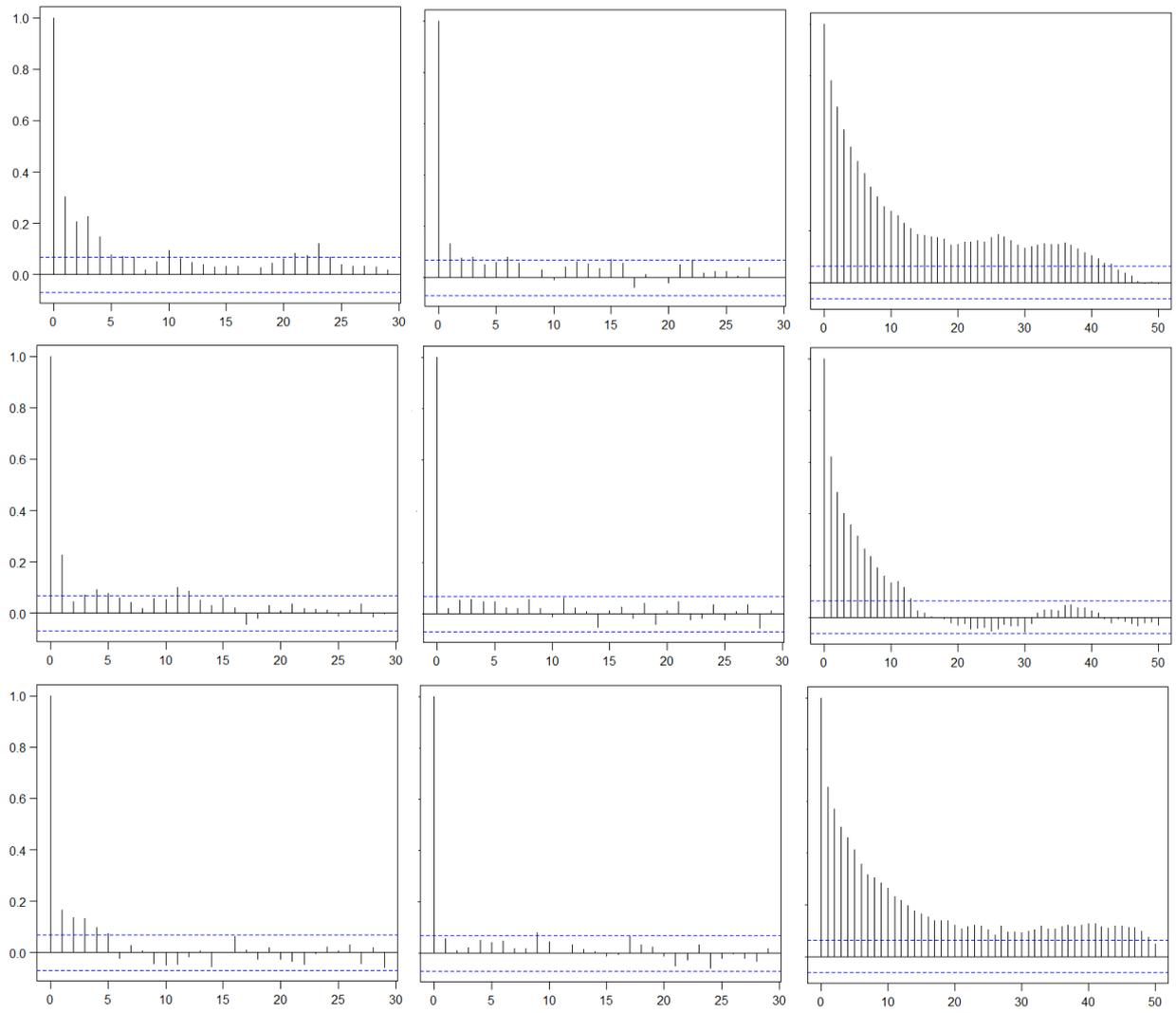


Figura 7.2: Autocorrelação de  $\lambda_1$ ,  $\lambda_2$  e  $\log(\hat{i}(\theta, y))$  referentes aos cenários 1 (linha 1), 2 (linha 2) e 3 (linha 3) do estudo de Simulação 5.1.

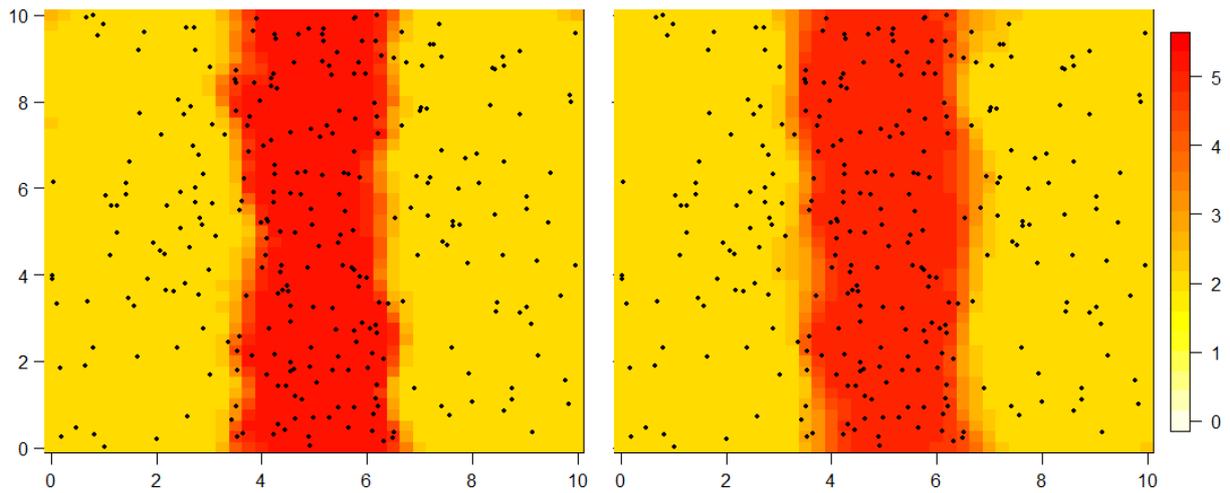


Figura 7.3: Estimação da média *a posteriori* da FI para os cenários 2 (esquerda) e 4 (direita) do estudo de Simulação 5.1.

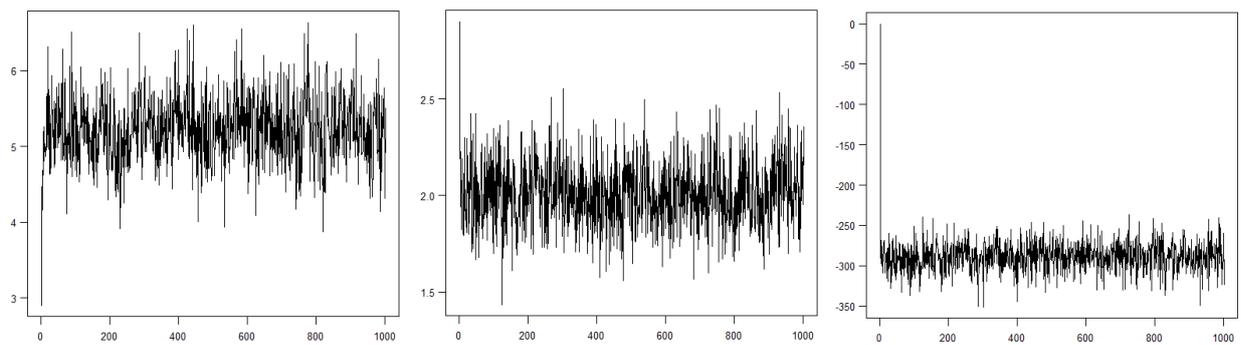


Figura 7.4: Trace Plots para as cadeias de  $\lambda_1$  (esquerda),  $\lambda_2$  (meio) e  $\log(\hat{l}(\theta, y))$  (direita) referente ao cenário 2 do estudo de Simulação 5.1.

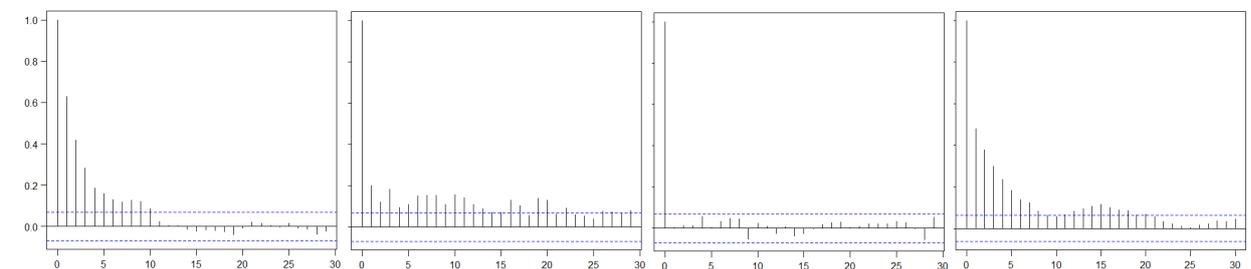


Figura 7.5: Autocorrelação de  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  e  $\log(\hat{l}(\theta, y))$  referentes ao caso  $K = 3$  do estudo de Simulação 5.2.

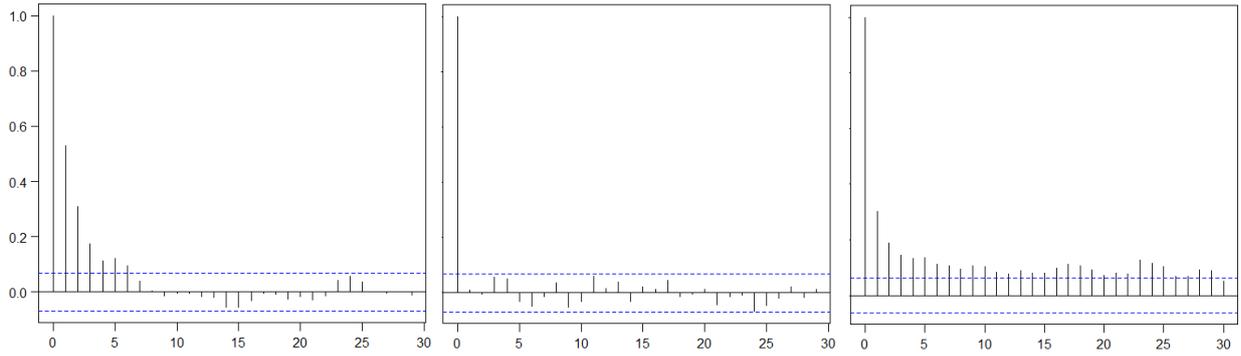


Figura 7.6: Autocorrelação de  $\lambda_1$ ,  $\lambda_2$  e  $\log(\hat{l}(\theta, y))$  referentes ao estudo de Simulação 5.3.

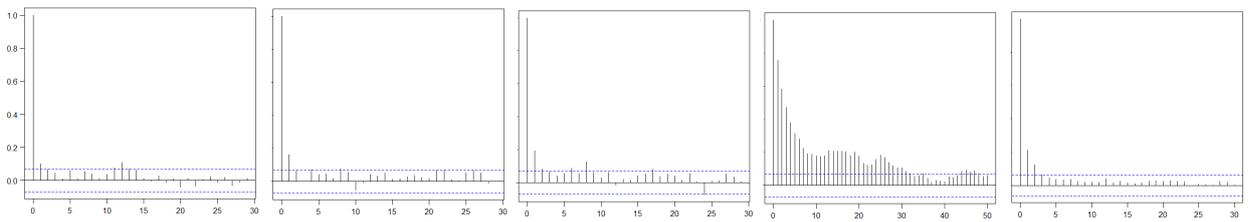


Figura 7.7: Autocorrelação de  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$ ,  $\lambda_4$  e  $\log(\hat{l}(\theta, y))$  referentes ao estudo de Simulação 5.4.

	<b>Estudo 1(cenário 2)</b>	<b>Estudo 2</b>	<b>Estudo 3</b>	<b>Estudo 4</b>
$\beta$	0.272	0.261	0.200	0.238
<b>N</b>	0.854	0.839	0.828	0.837
$\lambda$	0.336	0.353	0.390	0.331

Tabela 7.1: Taxa de aceitação das cadeias dos parâmetros referentes aos estudos de simulações 5.1, 5.2, 5.3 e 5.4.

	<b>Estudo 1</b> (cenário 2)	<b>Estudo 2</b>	<b>Estudo 3</b>	<b>Estudo 4</b>
<b>Variância da proposta <math>\lambda</math></b>	0.16	0.10	0.10	0.12
<b>c</b>	0	(-0.5, 0.5)	0	(-0.7, 0, 0.7)
<b>m</b>	1936	2704	2916	2500
<b>pN</b>	0.5	0.30	0.50	0.90
<b><math>\delta</math></b>	8	6	6	5

Tabela 7.2: Valores dos parâmetros que são fixos nos estudos de simulações 5.1, 5.2, 5.3 e 5.4.

### Comparação algoritmo Barker e Pseudo Marginal:

	Barker	<i>Pseudo-Marginal</i>
$\lambda_1$	29.567	1.662
$\lambda_2$	41.861	0.390
$\log(L(\theta, y))$	9.705	9.763

Figura 7.8: Valores de  $\frac{\text{Tempo}}{n_{eff}}$ , onde  $n_{eff}$  representa o tamanho efetivo da amostra, referente ao estudo de Simulação 5.1 para os métodos Barker e PM.

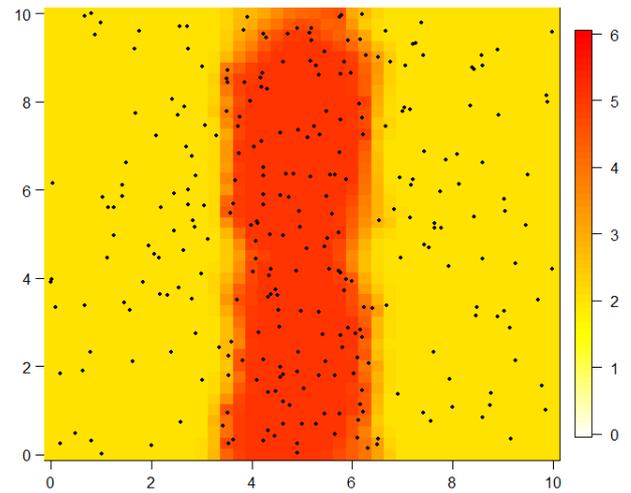


Figura 7.9: FI estimada pelo modelo Barker para o exemplo de simulação 5.1.

## Resultados Aplicação

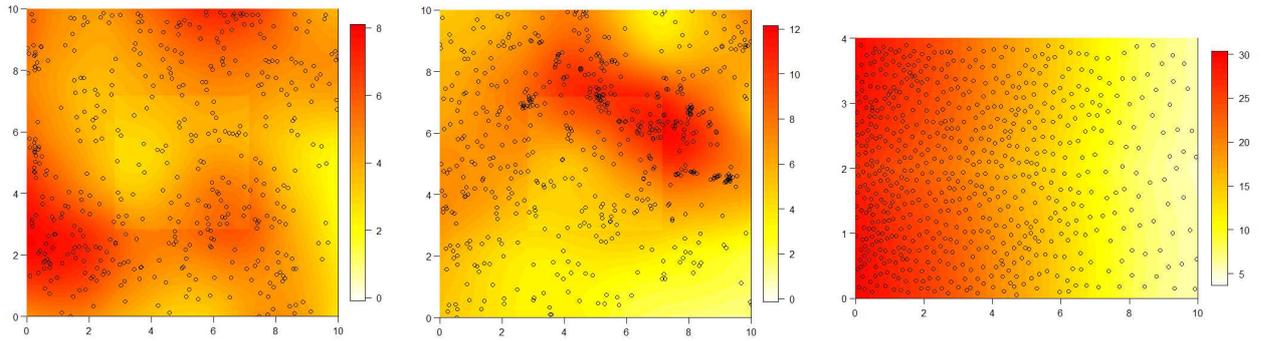


Figura 7.10: Estimação empírica via Kernel da FI para a escolha do valor de  $K$  e das regiões iniciais da Aplicação 6.1 (esquerda), Aplicação 6.2 (meio) e Aplicação 6.3 (direita).

	Aplicação 1	Aplicação 2	Aplicação 3
$\beta$	0.288	0.201	0.280
$\mathbf{N}$	0.834	0.831	0.849
$\lambda$	0.373	0.353	0.383

Tabela 7.3: Taxa de aceitação das cadeias dos parâmetros nos estudos aplicados 6.1, 6.2 e 6.3.

	Aplicação 1	Aplicação 2	Aplicação 3
Variância da proposta $\lambda$	0.12	0.15	0.3
$\mathbf{c}$	$(-0.5, 0.5)$	$(-0.5, 0.5)$	$(-0.7, 0, 0.7)$
$\mathbf{m}$	2500	2500	1296
$\mathbf{pN}$	0.80	0.6	0.75
$\delta$	7	3	2.5

Tabela 7.4: Valores dos parâmetros que são fixos nos estudos aplicados 6.1, 6.2 e 6.3.

# Referências Bibliográficas

- C. Andrieu and G. O. Roberts. The pseudo-marginal approach for efficient monte carlo computations. *The Annals of Statistics*, pages 697–725, 2009.
- C. Andrieu and M. Vihola. Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *The Annals of Applied Probability*, 25:1030–1077, 2015.
- A. Baddeley, E. Rubak, and R. Turner. *Spatial point patterns: methodology and applications with R*. Chapman and Hall/CRC, 2015.
- S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, 2014.
- A. A. Barker. Monte carlo calculations of the radial distribution functions for a proton electron plasma. *Australian Journal of Physics*, 18(2):119–134, 1965.
- R Bernhardt, F Meyer-Olbersleben, and B Kieback. Fundamental investigation on the preparation of gradient structures by sedimentation of different powder fractions under gravity. In *Proc. of the 4th Int. Conf. on Composite Engineering*, pages 147–148, 1997.
- A. Beskos, O. Papaspiliopoulos, and G. O. Roberts. Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, pages 1077–1098, 2006.
- A. Brix and P. J. Diggle. Spatiotemporal prediction for log-gaussian cox processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):823–841, 2001.
- D. R. Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 129–164, 1955.

- P. J. Diggle. *Statistical analysis of spatial and spatio-temporal point patterns*. CRC Press, 2013.
- J. A. Doornik. An object-oriented matrix programming language ox 6. 2009.
- M. M. Dunlop, M. A. Iglesias, and A. M. Stuart. Hierarchical bayesian level set inversion. *Statistics and Computing*, pages 1–30, 2016.
- P. Fearnhead, O. Papaspiliopoulos, and G. O. Roberts. Particle filters for partially observed diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):755–777, 2008.
- D. D. R. Frade. *Relações entre fatores ambientais e espécies florestais por metodologias de processos pontuais*. PhD thesis, Universidade de São Paulo, 2014.
- D. Gamerman and H. F. Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press, 2006.
- F. B. Gonçalves and D. Gamerman. Exact bayesian inference in spatiotemporal cox processes driven by multivariate gaussian processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):157–175, 2018.
- F. B. Gonçalves, K. Łatuszyński, and G. O. Roberts. Barker’s algorithm for bayesian inference with intractable likelihoods. *Brazilian Journal of Probability and Statistics.*, 31:732–745, 2017a.
- F. B. Gonçalves, K. G. Łatuszyński, and G. O. Roberts. Exact bayesian inference for diffusion driven cox processes. *To be published*, 2017b.
- F. B. Gonçalves and P. Franklin. On the definition of likelihood function. *arXiv:1906.10733*, 2019.
- P. J. Green and A. Mira. Delayed rejection in reversible jump metropolis–hastings. *Biometrika*, 88(4):1035–1053, 2001.
- A. Hildeman, D. Bolin, J. Wallin, and J. B. Illian. Level set cox processes. *Spatial Statistics*, 28:169–193, 2018.

- J. C. Kingman. *Poisson processes*. Wiley Online Library, 1993.
- K. Łatuszyński, I. Kosmidis, O. Papaspiliopoulos, and G. O. Roberts. Simulating events of unknown probabilities via reverse time martingales. *Random Structures & Algorithms*, 38(4):441–452, 2011.
- J. S. Liu. The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89:958–966, 1994.
- J. Møller, A. R. Syversveen, and R. P. Waagepetersen. Log gaussian cox processes. *Scandinavian journal of statistics*, 25(3):451–482, 1998.
- P. H. Peskun. Optimum monte-carlo sampling using markov chains. *Biometrika*, 60(3):607–612, 1973.
- V. Rao and Y. W. Teh. Fast mcmc sampling for markov jump processes and extensions. *The Journal of Machine Learning Research*, 14(1):3295–3320, 2013.
- G. O. Roberts, A. Gelman, Walter R. G., et al. Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability*, 7(1):110–120, 1997.
- B. Rowlingson, P. Diggle, and M. R. Bivand. Package splancs. *gen*, 14(1), 2012.
- V. L. Silva. *Associação espaço-temporal entre mortalidade por neoplasia de proximidade de antenas de telefonia celular em Belo Horizonte*. PhD thesis, Universidade Federal de Minas Gerais, 2012.
- D. Simpson, J. B. Illian, F. Lindgren, S. H. Sørbye, and H. Rue. Going off grid: Computationally efficient inference for log-gaussian cox processes. *Biometrika*, 103(1):49–70, 2016.
- R. C. Team et al. R: A language and environment for statistical computing. 2013.
- D. A. Zuanetti and L. A. Milan. Data-driven reversible jump for qtl mapping. *Genetics*, 202(1):25–36, 2016.