

Thiago Mota Carneiro

# **Pacote GeoPoisson: Implementação e Aplicações**

Natal - RN

29 de Julho de 2019

Thiago Mota Carneiro

## **Pacote GeoPoisson: Implementação e Aplicações**

Monografia de Graduação apresentada ao Departamento de Estatística do Centro de Ciências Exatas e da Terra da Universidade Federal do Rio Grande do Norte como requisito parcial para a obtenção do grau de Bacharel em Estatística.

Universidade Federal do Rio Grande do Norte

Centro de Ciências Exatas e da Terra

Departamento de Estatística

Orientador: Prof. Dr. Fidel Ernesto Castro Morales

Coorientador: Prof. Dr. Antonio Hermes Marques da Silva Jr.

Natal - RN

29 de Julho de 2019

Universidade Federal do Rio Grande do Norte - UFRN  
Sistema de Bibliotecas - SISBI  
Catalogação de Publicação na Fonte. UFRN - Biblioteca Setorial Prof. Ronaldo Xavier de Arruda - CCET

Carneiro, Thiago Mota.

Pacote GeoPoisson: implementação e aplicações / Thiago Mota  
Carneiro. - 2019.  
42f.: il.

Monografia (Bacharelado em Estatística) - Universidade  
Federal do Rio Grande do Norte, Centro de Ciências Exatas e da  
Terra, Departamento de Estatística. Natal, 2019.

Orientador: Fidel Ernesto Castro Morales.

Coorientador: Antonio Hermes Marques da Silva Júnior.

1. Estatística - Monografia. 2. Estatística aplicada -  
Monografia. 3. Inferência bayesiana - Monografia. 4. Estatística  
espacial - Monografia. I. Morales, Fidel Ernesto Castro. II.  
Silva Júnior, Antonio Hermes Marques da. III. Título.

RN/UF/CCET

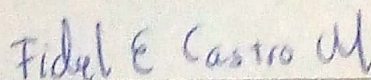
CDU 519.2

Thiago Mota Carneiro

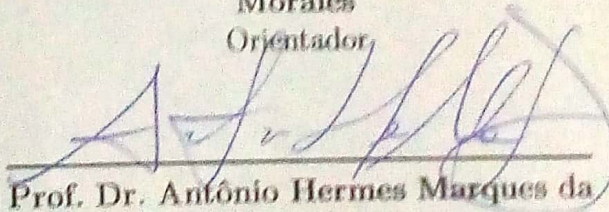
## Pacote GeoPoisson: Implementação e Aplicações

Monografia de Graduação apresentada no Departamento de Estatística do Centro de Ciências Exatas e da Terra da Universidade Federal do Rio Grande do Norte como requisito parcial para a obtenção do grau de Bacharel em Estatística.

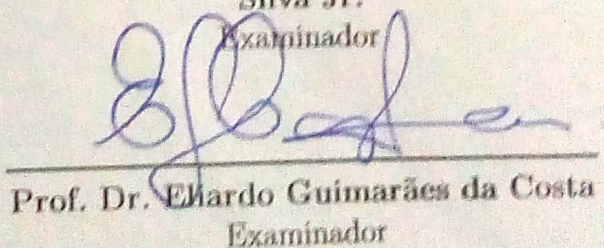
Aprovado em                      de                      de



Prof. Dr. Fidel Ernesto Castro  
Morales  
Orientador



Prof. Dr. Antônio Hermes Marques da  
Silva Jr.  
Examinador



Prof. Dr. Elardo Guimarães da Costa  
Examinador

Natal - RN  
29 de julho de 2019

*A Luiz e Nice*

# Agradecimentos

A Deus, que tudo dispôs com medida, quantidade e peso.

Aos meus pais, Luiz e Nice, a quem fui agraciado a ter ao meu lado todos esses anos.

Aos meus familiares e namorada, que me apoiaram com palavras, bons conselhos e visitas.

Ao professor orientador Fidel, que me deu minhas primeiras aulas de estatística quando eu ainda estava em outro curso e aceitou ser meu orientador nesse projeto.

Ao professor coorientador Hermes, que além de boas aulas me deu a oportunidade de participar desse projeto.

Ao professor Eliardo, que gentilmente aceitou ser parte da banca examinadora.

Aos professores Pledson e Dione, que muito me ajudaram na minha transição para o curso de Estatística.

Ao professor André, que, além de ótimas aulas em outras disciplinas, ministrou individualmente Séries Temporais para mim, permitindo que eu me formasse antecipadamente.

Ao professor Marcus Nunes, que disponibilizou o modelo em que esta monografia foi escrita.

A todos os colegas do PET e ao tutor, professor Moisés. Suas contribuições tornaram minha formação em Estatística mais rica.

A todos os demais professores e funcionários que me ajudaram em minha jornada.

*"A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey,  
strongly believes he has seen a mule."*

*Autor Desconhecido*

# Resumo

As aplicações de modelos geoestatísticos têm crescido rapidamente nas últimas décadas. Modelos geoestatísticos de contagem via Processo de Poisson têm se mostrado muito versáteis em ciências ambientais na previsão de eventos anômalos (excesso de chuvas, concentrações letais de CO no ar). No entanto, o modelo de contagem via Processo de Homogêneo é limitada para aplicações, restringindo as situações em que se adequa aos fenômenos. Para tal fim, foi desenvolvido o modelo geoestatístico de contagem via Processo de Poisson não homogêneo.

Este estudo se propõe a implementar em um pacote na linguagem R a metodologia de estimação do modelo acima. O uso das funções do pacote é exemplificado em dados obtidos de 29 estações de medições pluviométricas da ANA (Agência Nacional de Águas) para os estados do Piauí e Maranhão, durante 30 anos (1980 - 2010), e pretende criar intervalos de credibilidade, por meio de MCMC - Simulação de Monte Carlo via Cadeia de Markov (amostragem de Gibbs com passos de Metropolis-Hastings), para os parâmetros do modelo. Além disso, o pacote retorna um mapa de interpolação espacial de chuvas anômalas esperadas durante o mesmo período em regiões vizinhas não observadas.

**Palavras-chave:** Geoestatística. Poisson Não Homogêneo. MCMC.



# Abstract

The application of geostatistical models have been growing rapidly in the last decades. Geostatistical models for counting via Poisson process have been a versatile tool for environmental sciences in predicting anomalous events (excessive rains, lethal CO concentration). However, the counting via homogeneous Poisson process model is applicable to a very narrow scope of phenomena. For thus a geostatistical model was developed for counting via non homogeneous Poisson process.

We propose a new package in R software with a set of functions to estimate the model above. The usage of the package is exemplified with data obtained from 29 measuring stations in Piaui and Maranhao - Brazil, from 1980 to 2010 published by ANA (Agencia Nacional de Aguas - Brazilian National Water Agency). The main function obtains credibility intervals to the parameters of the model in a Bayesian approach through MCMC, specifically the Metropolis-Hastings algorithm within Gibbs sampling. In addition, we made a function which returns an interpolation map of anomalous events in neighboring unobserved regions.

**Keywords:** Geostatistics .Non-Homogeneous Poisson. MCMC.

# Lista de ilustrações

Figura 2.1 – Temperatura no Brasil em 25 de abril de 2016. PEGORIM, J.	15
Figura 2.2 – Taxa de homicídios a cada 100.000 em cada Estado brasileiro em 2016. FONTES, G.	15
Figura 2.3 – Casos de Ebola no centro e sul da África. Walsh MG, Haseeb M. 2015.	16
Figura 3.1 – Prioris de cada pesquisador. Disponível em Bolstad e Curran (2016)	21
Figura 3.2 – Posterioris de cada pesquisador. Disponível em Bolstad e Curran (2016)	22
Figura 6.1 – Observações da estação 2 acima do limite no tempo em dias.	36
Figura 6.2 – Taxa de aceitação do amostrador de Gibbs para $\alpha$ .	38
Figura 6.3 – Taxa de aceitação do amostrador de Gibbs para $\phi$ .	39
Figura 6.4 – Taxa de aceitação do amostrador de Gibbs para $W$ .	39
Figura 6.5 – Número esperado de dias com chuvas acima do limite no período de 30 anos.	40

# Lista de tabelas

Tabela 6.1 – Quantis empíricos da simulação das distribuições a posteriori dos parâ-	
metros.	37

# Sumário

<b>Lista de ilustrações</b>	8
<b>Lista de tabelas</b>	9
<b>1 INTRODUÇÃO</b>	12
<b>1.1 Objetivos</b>	12
1.1.1 Objetivos Específicos	12
<b>1.2 Estrutura dos capítulos</b>	12
<b>2 NOÇÕES DE ESTATÍSTICA ESPACIAL</b>	14
<b>2.1 Tipos de Dados Espaciais</b>	14
<b>2.2 Campos Aleatórios</b>	16
2.2.1 Introdução	16
2.2.2 Variograma e Estruturas de Covariância Espacial	17
2.2.3 Campo Aleatório Gaussiano	19
<b>2.3 Processos espaço-temporais</b>	19
<b>3 INFERÊNCIA BAYESIANA VIA MCMC</b>	20
<b>3.1 O Paradigma Bayesiano</b>	20
<b>3.2 Inferência Bayesiana</b>	22
<b>3.3 Simulação de Monte Carlo por Cadeias de Markov (MCMC)</b>	23
3.3.1 Amostrador de Gibbs	24
3.3.2 Algoritmo de Metropolis-Hastings	25
3.3.3 Amostrador de Gibbs com passos de Metropolis-Hastings	26
<b>4 MODELO GEOPOISSON NÃO HOMOGÊNEO</b>	27
<b>4.1 Introdução ao Modelo</b>	27
<b>4.2 A Parte Temporal do Modelo: Contagem por Processo de Poisson Não Homogêneo</b>	27
<b>4.3 A parte espacial do modelo</b>	28
<b>4.4 Estimação dos parâmetros</b>	28
<b>5 PACOTE GEOPOISSON</b>	30
<b>5.1 Estrutura Geral</b>	30
<b>5.2 Criação de pacotes no R</b>	30
<b>5.3 A Função Principal: GeoPoisson</b>	31
<b>5.4 Funções Auxiliares à função GeoPoisson</b>	32

<b>5.5</b>	<b>Interpolador</b>	32
<b>5.6</b>	<b>Utilização das funções</b>	34
<b>6</b>	<b>RESULTADOS</b>	36
<b>6.1</b>	<b>Aplicação</b>	36
<b>6.2</b>	<b>Diagnóstico das funções do amostrador de Gibbs com passos de</b>	
	<b>Metropolis-Hastings</b>	37
<b>6.3</b>	<b>Mapa gerado do Interpolador</b>	38
<b>7</b>	<b>CONSIDERAÇÕES FINAIS</b>	41
	<b>REFERÊNCIAS</b>	42

# 1 Introdução

## 1.1 Objetivos

O principal objetivo deste trabalho é desenvolver um pacote no R chamado GeoPoisson. Utilizando a metodologia do modelo geoestatístico para dados de contagem proposto por [Morales et al. \(2017\)](#), a função principal deste pacote estima os parâmetros. Além disso, baseando-se na interpolação descrita no mesmo artigo, há outra função que realiza a *krigagem* do processo em uma região de interesse em um intervalo de tempo  $t$ .

### 1.1.1 Objetivos Específicos

Os objetivos específicos deste trabalho são:

- Desenvolver funções para estimar os parâmetros do modelo proposto por [Morales et al. \(2017\)](#).
- Desenvolver funções que, a partir das estimativas obtidas no item anterior, gerem um mapa de interpolação espacial, prevendo valores para regiões não observadas.
- Organizar as funções num pacote da linguagem R.
- Apresentar o funcionamento do pacote em dados reais.

## 1.2 Estrutura dos capítulos

Este trabalho está dividido em sete capítulos. No capítulo 1 (Introdução), fala-se sobre os objetivos e a estrutura do trabalho.

O capítulo 2 é uma breve introdução a estatística espacial, expondo tipos de dados, seus usos e exemplos. Discorre-se brevemente sobre campos aleatórios: suas estruturas de correlação espacial, propriedades, variograma e o campo aleatório Gaussiano, um dos mais populares. Por fim, há uma curta descrição do que é um processo espaço-temporal - cujo entendimento é necessário para se compreender o modelo de [Morales et al. \(2017\)](#), exposto no capítulo 4 e, conseqüentemente, sua implementação em formato de pacote no R e as estimativas de seus parâmetros na aplicação nos capítulos 5 e 6 respectivamente.

O capítulo 3 descreve sucintamente tópicos de inferência bayesiana e simulação de Monte Carlo via Cadeias de Markov (MCMC). O primeiro tópico é discutido para expor a ideia geral da inferência bayesiana. O segundo tópico é uma explanação do método de simulação estocástica conhecido como MCMC. Este é o método de estimação implementado

no pacote para os parâmetros do modelo de interesse do capítulo 4. O MCMC é uma alternativa para modelos espaço-temporais nos quais o processo de estimação não possui solução analítica ou aproximações não iterativas dos parâmetros, ou ainda quando estas são muito complicadas.

O capítulo 4 explica o modelo geoestatístico Poisson não homogêneo de [Morales et al. \(2017\)](#). Este capítulo depende diretamente de conceitos expostos nos capítulos anteriores e é a base teórica do pacote.

No capítulo 5, é apresentado o pacote GeoPoisson, a parte original do trabalho. Para entender como funciona um pacote do R, há uma breve discussão de como elaborar um pacote em R, baseada no artigo de [Leisch \(2008\)](#). As funções do pacote GeoPoisson são expostas e explicadas, mostrando ao usuário como se deve utilizá-lo e que resultados esperar.

O capítulo 6 apresenta uma aplicação do pacote a dados reais - medições intensidade pluviométrica no interior do Piauí são modelados no pacote, que estima e apresenta percentis empíricos simulados (via MCMC) dos parâmetros. Ademais, serão apresentados gráficos do processo de convergência, a taxa de rejeição do amostrador de Gibbs (que deve estar de acordo com o proposto em [Neto\(2007\)](#) e [Roberts et al.\(1997\)](#). Finalmente, se exemplificará a função de interpolação espacial para gerar um mapa com valores preditos a partir das observações fornecidas.

O capítulo 7 retoma o que foi realizado em todo o trabalho, além propor novas funções para adicionar no pacote. Há ideias de melhorar o diagnóstico de convergência com auxílio de outros pacotes.

## 2 Noções de Estatística Espacial

### 2.1 Tipos de Dados Espaciais

Um processo espacial é definido por:  $\{Z(\mathbf{s}) : \mathbf{s} \in G \subset \mathbb{R}^d\}$ , em que  $Z$  denota um atributo observado na região  $G$ , por exemplo, a temperatura numa cidade brasileira, o número de ocorrências policiais num bairro. O local em que  $Z$  é observado é denotado por  $\mathbf{s}$ , um vetor de coordenadas com  $d \times 1$  dimensões. O mais comum é se estudar processos em espaço bidimensional,  $d = 2$ , e  $\mathbf{s} = [x, y]^T$  são coordenadas cartesianas. A análise de processos espaciais tem aplicações em diversas disciplinas, por exemplo meteorologia, geologia, geografia, entre outros. Dados espaciais são realizações do processo espacial  $Z(\mathbf{s})$  e podem ser diferenciados em três tipos por meio de características do domínio  $G$ :

**1 - Dados Geoestatísticos:** O domínio  $G$  é um conjunto contínuo e fixo. Por contínuo, entende-se que  $Z(\mathbf{s})$  pode ser observado em qualquer lugar em  $G$ , i.e., entre duas amostras de lugares  $\mathbf{s}_1$  e  $\mathbf{s}_2$  pode-se encontrar infinitas amostras de lugares (infinito não enumerável). Como os pontos são não estocásticos, diz-se que  $G$  é fixo. Um exemplo de dados geoestatísticos é a medição de temperaturas na superfície do Brasil, por poder ser mensurada em qualquer ponto do país (o domínio  $G$  é contínuo). Ao criar mapas das variáveis, é comum descrever certos intervalos de valores da característica de interesse por meio da intensidade das cores num espectro de duas cores nos extremos. Isto porque seria impossível expor todas as diferenças de temperatura nos infinitos pontos em que esta pode ser observada. Na Figura 2.1<sup>[1]</sup>, há um espectro de cores que vai do azul (temperatura mais baixa) até o vermelho (temperatura mais alta).

**2 - Dados Regionais:** são dados em que o domínio  $G$  é fixo e discreto, i.e., não estocástico e contável. O número de lugares pode ser finito ou infinito, neste último caso, deve ser enumerável. As localizações são sempre regiões ou áreas. Um exemplo é a taxa de homicídios em cada estado brasileiro como na Figura 2.2<sup>[2]</sup>. Neste caso, o Rio Grande do Norte inteiro é representado por apenas um valor e uma cor (considerando que a intensidade da cor representa a taxa de homicídios). Os estados vizinhos também assumem valores únicos e o interesse pode ser monitorar o perigo por região.

<sup>1</sup> <https://www.climatempo.com.br/noticia/2016/05/03/vai-esfriar-mais-em-maio-1620>. Acessado em 17 de julho de 2019.

<sup>2</sup> <https://www.gazetadopovo.com.br/politica/republica/dois-a-cada-tres-estados-tiveram-aumento-de-homicidios-rio-nao-esta-entre-eles-bkewyn3izxthv7aona9vkrzl3/>. Acessado em 17 de julho de 2019.



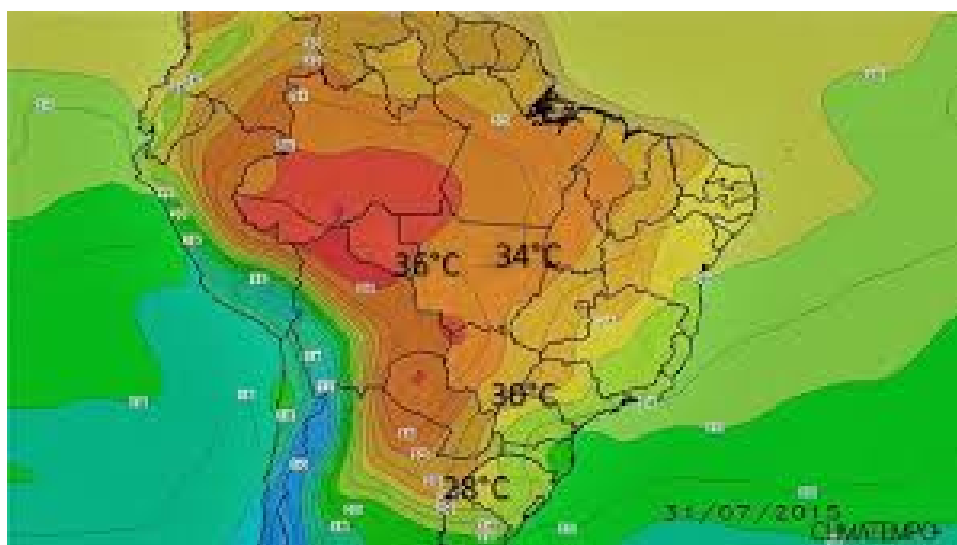


Figura 2.1 – Temperatura no Brasil em 25 de abril de 2016. PEGORIM, J.

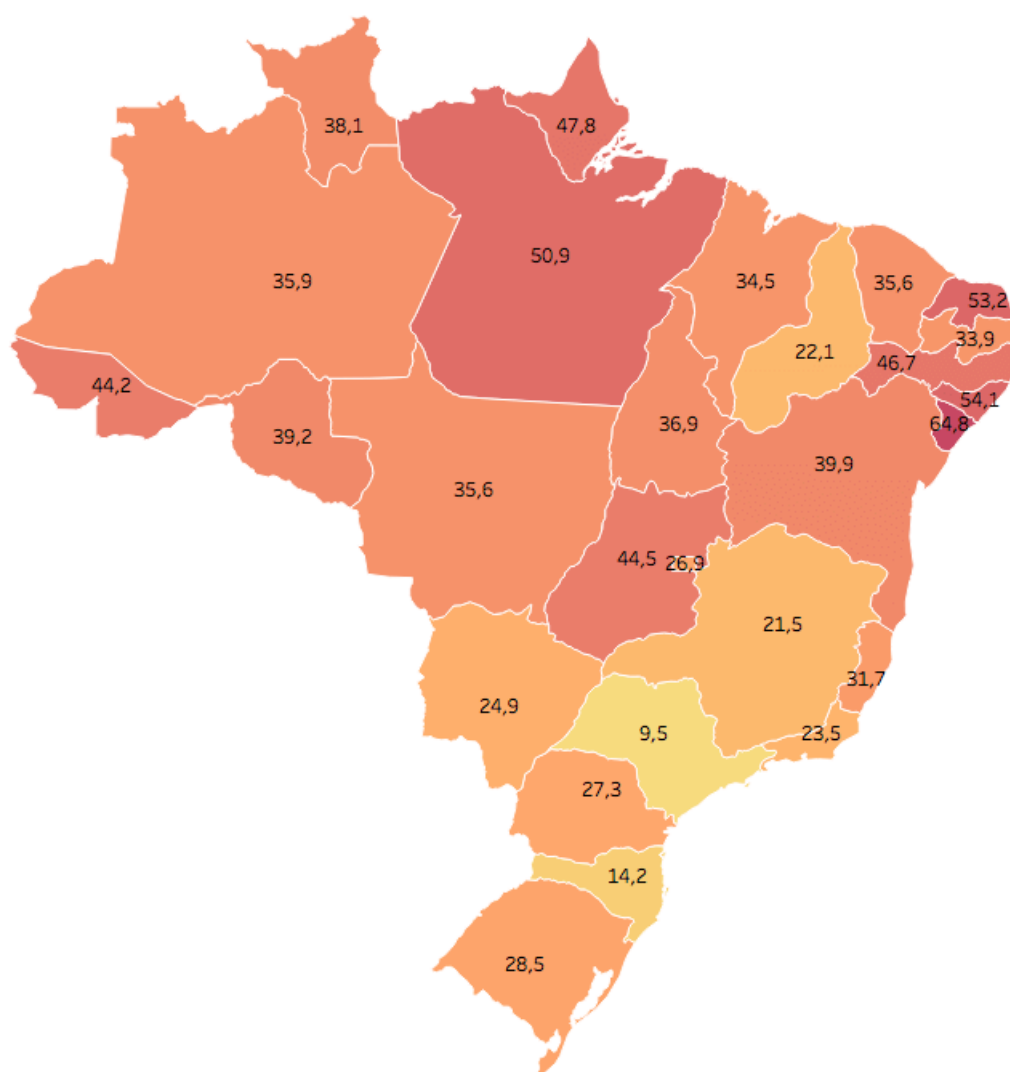


Figura 2.2 – Taxa de homicídios a cada 100.000 em cada Estado brasileiro em 2016. FONTES, G.

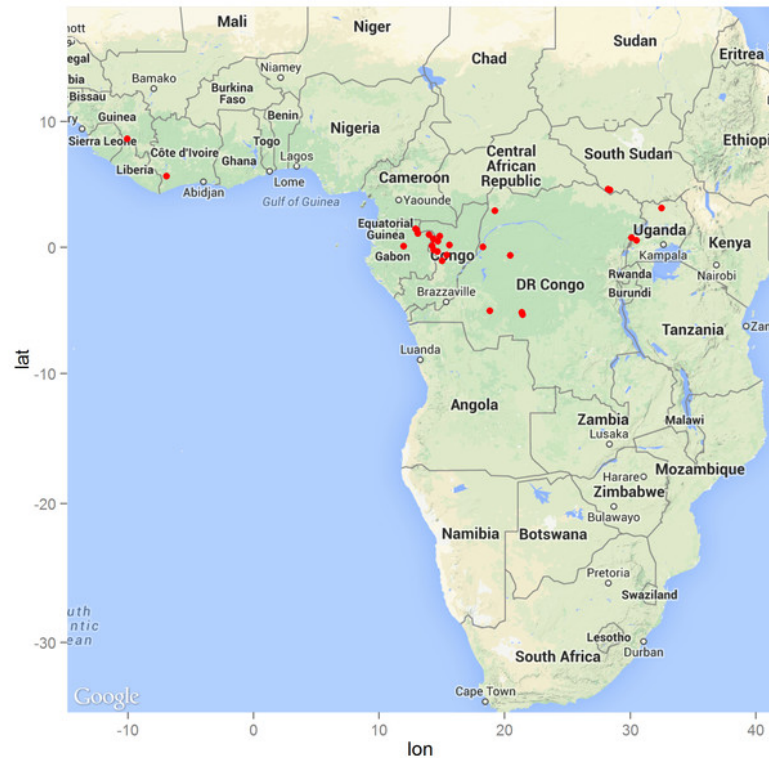


Figura 2.3 – Casos de Ebola no centro e sul da África. Walsh MG, Haseeb M. 2015.

**3 - Processos pontuais:** nesses dados a variável é medida de forma binária: um evento ocorreu ou não, ou ainda, uma medida ultrapassou um limite ou não. O interesse no estudo desse tipo de dado está em detectar anomalias ou erradicar doenças. Um exemplo de aplicação de padrões pontuais é a localização de ocorrências de ebola no centro-oeste africano, representado na Figura 2.3<sup>3</sup>.

## 2.2 Campos Aleatórios

### 2.2.1 Introdução

Campos aleatórios são processos estocásticos definidos num espaço  $G \{Z(\mathbf{s}) : \mathbf{s} \in G \subset \mathbb{R}^d\}$ . Em geral, o espaço é um subconjunto de  $\mathbb{R}^2$ . O campo é descrito pela distribuição de  $\mathbf{Z}(\mathbf{s}) = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))$ . Um campo aleatório é estritamente estacionário se a distribuição conjunta de  $(Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))$  é a mesma que a de  $(Z(\mathbf{s}_1 + \mathbf{k}), \dots, Z(\mathbf{s}_n + \mathbf{k}))$ <sup>4</sup>. A condição de estacionariedade fraca de um campo é atendida quando a média é

<sup>3</sup> <https://peerj.com/articles/735/>. Acessado em 17 de julho de 2019

<sup>4</sup> Obs: qualquer translação num espaço  $d$  dimensional será representada com a adição de uma letra com negrito. Quando se lê  $\mathbf{s} + \mathbf{k}$ , isto quer dizer: a soma do vetor  $d$  dimensional  $\mathbf{s}$  com outro vetor de mesma dimensão e módulo (ou distância euclidiana) igual a um valor real  $k$ . Poderá haver acréscimos similares com outras letras. O leitor deve seguir esta mesma interpretação.

igual a um  $m$  constante para todo  $\mathbf{s} \in G$  e a sua covariância só dependa da distância  $k$  (geralmente euclidiana) entre  $\mathbf{s}$  e  $\mathbf{s} + \mathbf{k}$ , ou seja,  $\text{Cov}(Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{k})) = C(k)$ . Se a função de covariância depender apenas da distância euclidiana entre dois pontos  $\mathbf{s}_i$  e  $\mathbf{s}_j$ , diz-se que o campo é isotrópico. A isotropia é uma importante propriedade de alguns campos, costumando ser um pressuposto muito forte e útil em modelagem de dados espaciais.

## 2.2.2 Variograma e Estruturas de Covariância Espacial

Um variograma é definido como:

$$E(Z(\mathbf{s} + \mathbf{k}) - Z(\mathbf{s}))^2 = \text{Var}(Z(\mathbf{s} + \mathbf{k}) - Z(\mathbf{s})),$$

assumindo que  $E(Z(\mathbf{s} + \mathbf{k}) - Z(\mathbf{s})) = 0$ .

Quando o variograma depende apenas de  $k$ , o processo será intrinsecamente estacionário e será possível escrevê-lo somente como  $2\gamma(k)$ , sendo  $\gamma(k)$  um semivariograma. A importância de estudá-los está em especificar o processo geoespacial para uma espaço  $G$ .

A fim de estimar um variograma, assumindo média constante, é possível utilizar o método dos momentos:

$$2\widehat{\gamma}(\mathbf{k}) = \frac{1}{|N(\mathbf{k})|} \sum_{N(\mathbf{k})} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2,$$

em que  $N(\mathbf{k})$  o número de pares de pontos com mesmas distâncias euclidianas entre si e  $|N(\mathbf{k})|$ , a respectiva cardinalidade. Há outros métodos mais apropriados para estimar o variograma, estes serão omitidos por não concernerem aos objetivos deste trabalho.

Uma função  $\text{Cov}(\mathbf{s}_i, \mathbf{s}_j)$  é uma função de covariância espacial válida se ela é positiva definida, i.e., se para qualquer inteiro positivo  $n, \mathbf{s}_j \in G$  e  $c_j \in \mathbb{R}$ :

$$\sum_{i,j} c_i c_j \text{Cov}(\mathbf{s}_i, \mathbf{s}_j) > 0; i, j = 1, \dots, n$$

De acordo com [Schabenberger e Gotway \(2005\)](#), alguns dos modelos de covariâncias são:

- Família esférica: seja uma função núcleo a indicadora de uma esfera:

$$K_d(u) = \begin{cases} 1, & \text{se } u < k \\ 0, & \text{c.c.} \end{cases}$$

e um campo aleatório de ruído branco. A função de covariância de sua convolução será:

$$C(k) = \text{Cov}[Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{k})] = \sigma^2 \int_{\mathbf{u}} K_d(\mathbf{u}) K_d(\mathbf{u} + \mathbf{h}) d\mathbf{u}.$$

Especificamente para  $d = 3$  (modelo esférico):

$$C(k) = \begin{cases} \sigma^2(1 - \frac{3}{2}\phi k + \frac{1}{2}(\phi k)^3), & 0 < k \leq \frac{1}{\phi}; \\ 0, & k \geq \frac{1}{\phi}. \end{cases}$$

- Matérn: esta é uma classe de funções de covariância definidas da seguinte forma:

$$C(k) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{\nu}k\phi)^\nu K_\nu(2\sqrt{\nu}k\phi), \quad k > 0, \nu > 0.$$

Sendo  $K_\nu$  a função de Bessel modificada do segundo tipo e ordem  $\nu > 0$ . O parâmetro  $\phi$  é responsável pelo alcance da dependência espacial e a suavidade é maior à medida em que  $\nu$  cresce. Um célebre caso particular desta classe é o modelo gaussiano, quando  $\nu \rightarrow \infty$ .

- Exponencial: este é um caso particular da classe Matérn para  $\nu = \frac{1}{2}$ .

$$C(k) = \sigma^2 \exp(-|\phi k^\nu|), \quad k > 0, 0 < \nu \leq 2.$$

- Quadrático racional:

$$C(k) = \sigma^2 \left( 1 + \frac{k^2}{1 + \phi k^2} \right), \quad k > 0.$$

- Ondular:

$$C(k) = \sigma^2 \frac{\text{sen}(\phi k)}{\phi k}.$$

Seja  $\rho(k)$  a correlação entre dois pontos no espaço  $G$  com distância euclidiana  $k$ . O efeito pepita é o valor de  $\lim_{k \rightarrow 0^+} \rho(k) = c_0$ . Em teoria, este valor deveria ser zero, não podendo haver descontinuidade (já que se define o campo aleatório como contínuo). Em muitos casos, o valor é diferente de zero e se atribui este fenômeno a um erro de medida, sendo esperado que repetidas mensurações flutuariam ao redor do valor real.

Foi proposto modelar  $c_0$  como um processo composto por um erro de medida  $c_{EM}$  e um processo ruído branco (média zero e variância constante):

$$c_0 = c_{EM} + c_{RB}.$$

Em alguns problemas, a informação acima é essencial para verificar a validade de modelos de interpolação.

Um patamar é o valor  $\lim_{k \rightarrow \infty} \rho(k) = c_0 + \sigma^2$ . Se o patamar for atingido para um  $d$  finito, chama-se de alcance o inverso de  $\rho(k)$ . Designa-se por alcance efetivo o valor de  $k$  em

que correlação cai para um valor muito pequeno (0,05 ou outro, de acordo com o critério do pesquisador). Para além de  $k$ , considera-se que a correlação espacial é desprezível.

### 2.2.3 Campo Aleatório Gaussiano

O campo aleatório gaussiano é aquele em que todas as suas distribuições de dimensão finita  $F(z(\mathbf{s}_1), \dots, z(\mathbf{s}_n))$  de um vetor de características de interesse  $\mathbf{z} = (z(\mathbf{s}_1), z(\mathbf{s}_2), \dots, z(\mathbf{s}_n))$  são distribuições normais multivariadas com vetor de médias  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)$  para as  $n$  localizações  $(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n)$  e matriz de covariâncias  $\boldsymbol{\Sigma}$  de dimensão  $n \times n$ , em que cada elemento é a covariância da característica de interesse nas coordenadas. Em outras palavras, cada ponto do espaço tem uma função de densidade de probabilidade associada à característica de interesse:

$$f(\mathbf{z}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left(\frac{-\mathbf{z}^\top - \boldsymbol{\mu} \boldsymbol{\Sigma}^{-1} \mathbf{z} - \mu}{2}\right).$$

De acordo com Schmidt e Sanso (2006), há equivalência entre estacionariedade forte e fraca para campos aleatórios Gaussianos.

## 2.3 Processos espaço-temporais

Um processo espaço-temporal pode ser descrito como  $\Lambda(\mathbf{s}, t)$  em  $G \times \mathbb{R}^+$  (em que  $G$  se refere ao subespaço de  $\mathbb{R}^d$  - domínio da parte espacial e  $\mathbb{R}^+$  é o domínio do tempo). Para fins práticos, adota-se  $t$  discreto como tempo de coletas da informação de interesse. Quando o processo é estacionário e tem estrutura de covariância separável, especifica-se a covariância dependendo apenas da distância euclidiana e diferença de tempos de medição:

$$\text{Cov}(\mathbf{s}_i, \mathbf{s}_j, t_i, t_j) = C(k)C(h),$$

em que  $k = \|s_i - s_j\|$  e  $h = |t_i - t_j|$ .

Este é o modelo de maior simplicidade na estimação dos parâmetros. No entanto, isso limita muito o que pode ser modelado, já que a propriedade acima não se aplica à maioria dos fenômenos. Neste caso, é comum se olhar para alguma transformação da função do processo espaço temporal, como logaritmo:

$$\log(\Lambda(\mathbf{s}, t)) = W(\mathbf{s}|\theta_w) + H(t|\theta_H),$$

no qual se modela separadamente a parte espacial e a parte temporal do processo e simplificando o trabalho de estimar os parâmetros.

As informações deste capítulos foram obtidas de Schabenberger e Gotway (2005), Gneiting e Guttorp (2005), Schmidt e Sanso (2006) e Cressie (1991).

## 3 Inferência Bayesiana via MCMC

### 3.1 O Paradigma Bayesiano

Suponha que um pesquisador quer determinar o comportamento de um parâmetro  $\theta$  modelado por uma função probabilidade  $f(y | \theta)$ . Assuma que, por sua experiência, ele tenha uma boa noção sobre o valor de  $\theta$ . No entanto, pelo paradigma frequentista, a inferência de  $\theta$  é realizada com base apenas na função de verossimilhança, i. e., coleta-se uma amostra aleatória  $\mathbf{Y} = (Y_1, \dots, Y_n)$ , calculando-se a função de verossimilhança a partir dos resultados observados:  $L(\theta | \mathbf{y})$  e, baseando-se exclusivamente nela, realiza-se o procedimento de inferência (teste de hipótese, estimação pontual e intervalar). Por outro lado, na abordagem Bayesiana, é possível utilizar o conhecimento a priori sobre  $\theta$  através de uma função de probabilidade  $\pi(\theta)$ , chamada de função de distribuição a priori. No paradigma Bayesiano, o pesquisador incorpora sua crença a priori sobre  $\theta$  por meio de  $\pi(\theta)$ , posteriormente coletando uma amostra  $\mathbf{Y}$  e atualiza seu conhecimento sobre o parâmetro por meio do Teorema de Bayes:

$$\pi(\theta | \mathbf{y}) = \frac{L(\theta | \mathbf{y})\pi(\theta)}{\int_L L(\theta | \mathbf{y})\pi(\theta)},$$

em que  $\pi(\theta | \mathbf{y})$  é a função de distribuição a posteriori (caso contínuo).

$$\pi(\theta | \mathbf{y}) = \frac{L(\theta | \mathbf{y})\pi(\theta)}{\sum_L L(\theta | \mathbf{y})\pi(\theta)}$$

(caso discreto).

Mais usualmente, se escreve a fórmula acima por proporcionalidade, a fim de simplificar o trabalho com o núcleo da distribuição posteriori (a busca pela constante que torna a posteriori uma função de distribuição de probabilidade é feita num segundo momento):

$$\pi(\theta | \mathbf{y}) \propto L(\theta | \mathbf{y})\pi(\theta).$$

O processo de escolha da priori (conhecido como elicitaco da priori) causa um impacto relevante na posteriori. O ideal é que o pesquisador já conheça algo do comportamento de  $\theta$ , como simetria ou assimetria, presença ou ausência de caudas pesadas e proponha uma distribuição coerente com estas características. Quando o pesquisador não crê dispor de um conhecimento substantivo, recomenda-se o uso de uma distribuição não informativa para a priori. No caso uniparamétrico, a proposta de Jeffreys é uma função que dependa da informação de Fisher ( $I(\theta)$ ):

$$\pi(\theta) \propto [I(\theta)]^{-1/2}.$$

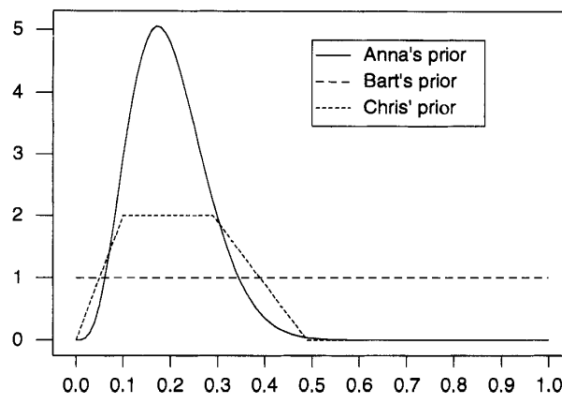


Figura 3.1 – Prioris de cada pesquisador. Disponível em [Bolstad e Curran \(2016\)](#)

Um exemplo de [Bolstad e Curran \(2016\)](#) mostrando a influência da priori na posteriori é o seguinte: imagine que três pesquisadores desejam saber a opinião de moradores da região sobre cassinos. Anna acredita que 20% dos moradores apoiam a construção do cassino e assume um desvio-padrão a priori de 0,08. Ajustando uma distribuição beta, ela toma uma  $Beta(4,8, 19,2)$  como distribuição a priori. Bart não tem nenhuma crença específica e decide assumir uma distribuição  $uniforme(0,1)$  como distribuição a priori. Chris acredita numa priori em forma trapezoidal, com maiores pesos entre 0,1 e 0,3. Ele propõe uma distribuição a priori imprópria (função que não atende às condições de uma genuína função de distribuição de probabilidade) com a seguinte lei de formação:

$$P = \begin{cases} 20p, & \text{se } 0 \leq p \leq 0,10 \\ 0,2, & \text{se } 0,10 \leq p \leq 0,30 \\ 5 - 10p, & \text{se } 0,30 \leq p \leq 0,50. \end{cases}$$

Na visão de muitos bayesianos, o uso de prioris impróprias não é um problema desde que a posteriori seja uma genuína função de distribuição de probabilidade. Os pesquisadores tomam uma amostra de 100 pessoas e descobrem que 26 defendem a construção de um cassino. Os gráficos das prioris e posterioris obtidas por cada pesquisador estão indicados nas Figuras 3.1 e 3.2, respectivamente. Neste exemplo, é possível entender como a escolha inicial do pesquisador interfere na distribuição a posteriori obtida. Por esta razão, deve-se fundamentar o processo de elicitacão em literatura consagrada da área ou adotar prioris não informativas.

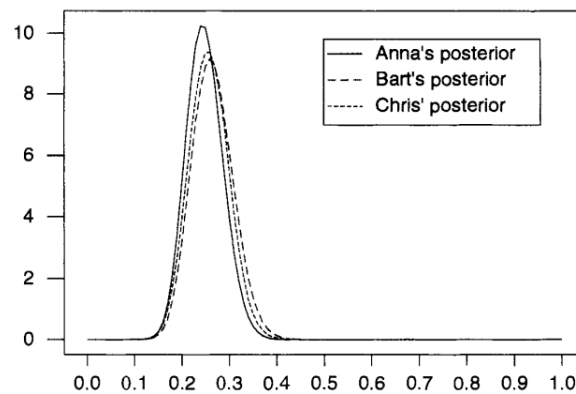


Figura 3.2 – Posterioris de cada pesquisador. Disponível em [Bolstad e Curran \(2016\)](#)

## 3.2 Inferência Bayesiana

É comum que um pesquisador tenha interesse em fazer inferências sobre uma população a partir de amostras aleatórias. No paradigma Bayesiano, o teste de hipóteses, a inferência pontual e intervalar é realizada a partir da função de probabilidade a posteriori e, às vezes, a partir da teoria de decisão. Um problema de decisão tem alguns elementos básicos:

- o espaço paramétrico  $\Theta$ , um conjunto não vazio de possíveis valores de  $\theta$ ;
- o espaço amostral  $\Omega$ , um conjunto não vazio de valores que uma variável aleatória  $X$  pode assumir.  $X$  tem função de distribuição de probabilidade que depende de  $\theta$ ;
- um conjunto  $A$  não vazio de possíveis ações que podem ser tomadas pelo pesquisador;
- uma função de decisão  $\delta : \Omega \rightarrow A$ . As decisões podem ser: aceitar ou rejeitar uma hipótese (para testes de hipótese); escolher um valor  $\theta \in \Theta$  em detrimento de outros como estimador; escolher um determinado intervalo de credibilidade  $[a, b]$ ,  $a, b \in \Theta$  em detrimento de outros.

Cada decisão do pesquisador está associada a uma função de perda  $L : \Theta \times A \rightarrow \mathbb{R}^+$ . O objetivo é tomar uma decisão que incorra na menor perda possível sob determinado critério. A esperança da posteriori da perda é denominado risco da regra de decisão  $R(\delta)$ . Uma regra  $\delta^*$  é dita ótima quando seu risco é o mínimo - e este risco é denominado risco de Bayes.

As funções de perda mais utilizadas para estimação pontual de  $\theta$  são:

- Perda quadrática:  $L(\delta, \theta) = (\delta - \theta)^2$ . O estimador resultante é a esperança da posteriori de  $\theta$ :  $\delta = \hat{\theta} = E_{\theta|y}(\theta)$ ;



- Perda absoluta: neste caso,  $L(\delta, \theta) = |\delta - \theta|$ , sendo o estimador a mediana da posteriori de  $\theta$ .

- Perda 0-1: ocorre se  $L(\delta, \theta) = \lim_{\epsilon \rightarrow 0} I_{|\theta - \delta|}([\epsilon, \infty])$ . O estimador é a moda da posteriori de  $\theta$ .

Esses procedimentos, no entanto, são menos usuais que a informação de intervalos de credibilidade. Neste caso, descreve-se o menor intervalo  $[c_1, c_2]$  da distribuição a posteriori de  $\theta$  que contenha  $(1 - \alpha)\%$  da área total da distribuição. O valor  $1 - \alpha$  é chamado de nível de credibilidade. Muito similar à inferência frequentista, este procedimento tem diferente interpretação. Na inferência clássica, falava-se em intervalos de confiança para o estimador que conteria o valor real do parâmetro em  $(1 - \alpha)\%$  das vezes caso o experimento fosse repetido muitas vezes. Na inferência Bayesiana, fala-se do grau de credibilidade  $(100 - \alpha)\%$  de o valor do parâmetro (uma variável aleatória) estar no intervalo  $[c_1, c_2]$ , convenientemente tomando-se o menor intervalo com este percentual da área. Economizam-se conceitos ao tratar-se o parâmetro como variável aleatória.

### 3.3 Simulação de Monte Carlo por Cadeias de Markov (MCMC)

Em geral, tem-se problemas em encontrar formas fechadas para distribuição a posteriori. Quando isto ocorre, é necessário obter informações sobre a distribuição (como a esperança) de forma aproximada. [Gamerman e Lopes \(2006\)](#) descrevem quatro tipos de aproximações comuns em inferência bayesiana: aproximações assintóticas, quadratura gaussiana, simulação de Monte Carlo, simulações estocásticas. As duas últimas são particularmente interessantes quando se quer estimar um número grande de parâmetros de um modelo complexo. No entanto, apresentam a desvantagem de requererem mais tempo, uma vez que dependem de uma sucessiva geração de números aleatório e, em muitos casos, a convergência para a distribuição a posteriori pode demorar centenas de milhares de iterações. Por esta razão, os autores recomendam evitar essas últimas técnicas quando o modelo for simples o bastante para usar as duas primeiras.

Formalmente, diz-se que um processo estocástico tem a propriedade de Markov se:

$$\forall n \in \mathbb{N}, P(X_n \in A_n | X_{n-1} = x, X_{n-2} \in A_{n-2}, \dots, X_0 \in A_0) = P(X_n \in A_n | X_{n-1} = x).$$

Em termos mais simples, isto equivale a dizer que no processo estocástico  $(X_n)_{n \in \mathbb{N}}$  o valor de  $X_n$  depende apenas de  $X_{n-1}$ .

Seja  $(X_n)_{n \in S_{\mathbb{C}\mathbb{N}}}$  um processo estocástico markoviano indexado por um conjunto finito ou infinito enumerável. Se  $X_n = i$ , diz-se que no momento  $n$  o processo está no  $i$ -ésimo estado. Supondo-se que há uma probabilidade fixa  $p_{ij}$  de o processo passar do  $i$ -ésimo para o  $j$ -ésimo estado, é possível representar todas as possíveis probabilidades de passagem no  $n$ -ésimo tempo por uma matriz:

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Sendo cada  $p_{i,j}$  a probabilidade de ir de um estado  $i$  para um estado  $j$ , e  $i, j$  são estados possíveis num conjunto contínuo ( $i, j \in G \subset \mathbb{R}^d, d \in \mathbb{N}$ ). No caso contínuo, a passagem de um estado  $i$  para um estado  $j$  pode ser representado como:

$$p_{i,j} = \frac{\partial F(i,j)}{\partial j},$$

em que  $F(i,j)$  é uma função contínua de passagem de  $X_n$  de um estado  $i$  a um estado  $j$ . Se as probabilidades de transição invariantes com tempo, a cadeia de Markov é dita homogênea. Uma cadeia é irreduzível quando pode atingir qualquer estado a partir dele mesmo ou de outro estado qualquer num número finito de passos. É recorrente positiva se o tempo médio de retorno a um mesmo ponto é finito para qualquer estado.

Com as propriedades acima, uma cadeia de Markov apresenta distribuição estacionária única. Tomando-se  $\lim_{n \rightarrow \infty} p_{i,j}^n = \pi(j)$ , a distribuição estacionária é única se  $\lim_{n \rightarrow \infty} F(X_n) = \pi(j)$ . Isto garante que, para  $n$  suficientemente grande, a distribuição de  $X_n$  é aproximadamente  $\pi$ .

Uma cadeia de Markov é aperiódica se for possível que esteja num mesmo estado  $i$  para diferentes tempos  $k$  e  $k+1$ . Se uma cadeia é homogênea, irreduzível, recorrente positiva e aperiódica, diz-se que ela é ergódica. Uma cadeia ergódica tem um comportamento limite  $p^n(i,j) \xrightarrow{n \rightarrow \infty} \pi(j)$ . Na prática é possível sortear  $\theta^{(1)}$  da densidade  $p(\theta^{(0)})$ , depois  $\theta^{(2)}$  de  $p(\theta^{(1)})$ , e à medida em que  $n$  aumente, atendidas as condições citadas, os sorteios da distribuição limite são boas aproximações de sorteios de  $\pi$ . Este resultado é o que torna possível métodos de simulação estocástica do tipo MCMC. Dois deles serão descritos mais adiante: o Amostrador de Gibbs e o algoritmo de Metropolis-Hastings.

### 3.3.1 Amostrador de Gibbs

Como visto em [Gamerman e Lopes \(2006\)](#) e [Schmidt e Sanso \(2006\)](#), considere  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$  um vetor de variáveis aleatórias com distribuição conjunta (excetuando-se uma constante)  $\pi(\boldsymbol{\theta})$ . Suponha que seja fácil gerar valores para todos os  $\theta_i$  por meio de  $\pi_i(\theta_i) = \pi(\theta_i | \boldsymbol{\theta}_{-i})$  (a distribuição condicional de  $\theta_i$  tendo todos os parâmetros  $\boldsymbol{\theta}_{(-i)} = (\theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n)$  de  $1, \dots, n$  diferentes de  $i$ ). O amostrador de Gibbs gerará uma sequência  $(\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \dots)$  a partir de uma Cadeia de Markov com distribuição limite  $\pi(\boldsymbol{\theta})$  e um núcleo de transição composto por produtos das condicionais completas.

O algoritmo segue os seguintes passos:

**Passo 1** - Definir valores iniciais de  $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_n^{(0)})$ ;

**Passo 2** - para  $t = 1, 2, \dots, n$ , sortear um valor para  $\boldsymbol{\theta}^{(t)}$  na  $t$ -ésima iteração a partir de  $\boldsymbol{\theta}^{(t-1)}$ , especificamente:

i- sortear  $\theta_1^{(t)}$  da função  $\pi(\theta_1 | \boldsymbol{\theta}_{-i}^{(t-1)})$ ;

ii- sortear  $\theta_2^{(t)}$  da função  $\pi(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_n^{(t-1)})$ ;

iii - para o  $n$ -ésimo parâmetro, sortear  $\theta_n^{(t)}$  da função  $\pi(\theta_n | \theta_1^{(t)}, \dots, \theta_{n-1}^{(t)})$ ;

**Passo 3** - repetir o passo 2 até atingir a distribuição limite  $\pi(\boldsymbol{\theta})$ . Este critério pode ser verificado por um diagnóstico da convergência.

### 3.3.2 Algoritmo de Metropolis-Hastings

Segundo [Gamerman e Lopes \(2006\)](#), quando for muito difícil ou custoso gerar valores para a posteriori  $\pi$ , pode-se tomar amostras desta iterativamente por meio de cadeias de Markov. Em [Turkman e Paulino \(2015\)](#) e [Gamerman e Lopes \(2006\)](#), para o algoritmo de Metropolis-Hastings, escolhe-se uma distribuição instrumental arbitrária  $q(\boldsymbol{\theta}, \boldsymbol{\phi})$  cujo produto com certo  $\alpha(\boldsymbol{\theta}, \boldsymbol{\phi})$  defina uma cadeia reversível  $p(\boldsymbol{\theta} | \boldsymbol{\phi})$  da seguinte forma:

$$p(\boldsymbol{\theta} | \boldsymbol{\phi}) = q(\boldsymbol{\theta} | \boldsymbol{\phi})\alpha(\boldsymbol{\theta} | \boldsymbol{\phi})$$

Hastings demonstrou que, dada a razão  $R = \frac{\pi(\boldsymbol{\phi})q(\boldsymbol{\phi} | \boldsymbol{\theta})}{\pi(\boldsymbol{\theta})q(\boldsymbol{\theta} | \boldsymbol{\phi})}$ , este  $\alpha$  é:

$$\alpha(\boldsymbol{\phi} | \boldsymbol{\theta}) = \min\{1, R\}$$

Os passos do algoritmo são:

**Passo 1** - definir valores iniciais de  $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_n^{(0)})$ ;

**Passo 2** - na  $t$ -ésima iteração, sortear o valor  $\boldsymbol{\phi}^{(t)}$  de  $q(\boldsymbol{\phi} | \boldsymbol{\theta}^{(t)})$ ;

**Passo 3** - encontrar  $\alpha(\boldsymbol{\theta}^{(t)} | \boldsymbol{\phi}^{(t)}) = \min\{1, R^{(t)}\}$ , em que  $R$  é a razão de Hastings para os  $t$ -ésimos  $\boldsymbol{\phi}$  e  $\boldsymbol{\theta}$ ;

**Passo 4** - sortear  $\boldsymbol{\theta}^{(t+1)}$  como:

$$\boldsymbol{\theta}^{(t+1)} = \begin{cases} \boldsymbol{\phi}^{(t)}, & \text{com probabilidade } \alpha(\boldsymbol{\theta}^{(t)} | \boldsymbol{\phi}^{(t)}); \\ \boldsymbol{\theta}^{(t)}, & \text{com probabilidade } 1 - \alpha(\boldsymbol{\theta}^{(t)} | \boldsymbol{\phi}^{(t)}); \end{cases}$$

**Passo 5** - repetir o algoritmo a partir do passo 2 até a convergência.

### 3.3.3 Amostrador de Gibbs com passos de Metropolis-Hastings

Muitas vezes, não é trivial tomar amostras de uma função de distribuição condicional - requisito do amostrador de Gibbs. Nestes casos, [Gamerman e Lopes \(2006\)](#) recomenda que o valor seja sorteado por meio do algoritmo de Metropolis-Hastings dentro de cada passo do amostrador de Gibbs, da seguinte maneira:

**Passo 1** - definir valores iniciais de  $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_n^{(0)})$ ;

**Passo 2** - cada subpasso se assemelha-se ao amostrador de Gibbs. No entanto, o valor não é sorteado da função condicional (uma vez que não se tem forma fechada nesses casos). Deve-se sortear um valor para  $\boldsymbol{\theta}^{(t)}$  na t-ésima iteração a partir de  $\boldsymbol{\theta}^{(t-1)}$ , especificamente:

i- sortear  $\theta_1^{(t)}$  por meio do algoritmo de Metropolis Hastings;

i.i -na t-ésima iteração, sortear valor  $\boldsymbol{\phi}^{(t)} \sim q(\boldsymbol{\phi} | \boldsymbol{\theta}^{(t)})$ ;

i.ii - encontrar  $\alpha(\boldsymbol{\theta}^{(t)} | \boldsymbol{\phi}^{(t)}) = \min 1, R^{(t)}$ , em que  $R$  é a razão de Hastings, para os t-ésimos  $\boldsymbol{\phi}$  e  $\boldsymbol{\theta}$ ;

i.iii - sortear  $\boldsymbol{\theta}^{(t+1)}$  como:

$$\boldsymbol{\theta}^{(t+1)} = \begin{cases} \boldsymbol{\phi}^{(t)}, & \text{com probabilidade } \alpha(\boldsymbol{\theta}^{(t)} | \boldsymbol{\phi}^{(t)}); \\ \boldsymbol{\theta}^{(t)}, & \text{com probabilidade } 1 - \alpha(\boldsymbol{\theta}^{(t)} | \boldsymbol{\phi}^{(t)}); \end{cases}$$

ii- sortear  $\theta_2^{(t)}$  por meio do algoritmo de Metropolis Hastings, como feito em i, mas atualizando  $\theta_1$  como  $\theta_1^{(t)}$  já sorteado;

iii - para o  $n$ -ésimo valor, sortear  $\theta_n^{(t)}$  por meio do algoritmo de Metropolis Hastings, como feito em i, mas atualizando os  $\theta$ s anteriores já sorteados;

**Passo 3** - repetir o passo 2 até atingir a distribuição limite  $\pi(\boldsymbol{\theta})$ . Este critério pode ser verificado por um diagnóstico de convergência.

Neste caso, é importante ter mais iterações que o Amostrador de Gibbs incomum, pois há um certo desperdício nas vezes em que o valor atualizado repete o passo anterior.

## 4 Modelo GeoPoisson Não Homogêneo

### 4.1 Introdução ao Modelo

Em geoestatística, um problema comum é detectar comportamentos anômalos num espaço (excesso de CO no ar, chuvas excessivas numa região). Para tais casos, é interessante pensar em um modelo espaço-temporal de contagem de eventos anômalos. [Morales et al. \(2017\)](#) propõe um modelo espaço-temporal Poisson não homogêneo que se define numa área geográfica de interesse  $G \subset \mathbb{R}^d$  (em geral,  $d = 2$ ) com  $n$  pontos fixos  $(\mathbf{s}_1, \dots, \mathbf{s}_n) \in G$ , com sua fórmula expressa por um  $\Lambda(\mathbf{s}, t)$ . Mais comumente, costuma-se expressar seu logaritmo:

$$\log(\Lambda(\mathbf{s}, t)) = W(\mathbf{s} \mid \boldsymbol{\theta}_w) + H(t \mid \boldsymbol{\theta}_H),$$

em que  $W$  é a parte espacial com propriedades de isotropia e não estacionariedade e  $H$  é a parte temporal, baseada no modelo generalizado de Goel Poisson não homogêneo. A vantagem da notação logarítmica do modelo é facilitar a estimação dos parâmetros. No artigo, este modelo foi satisfatoriamente aplicado a dados de poluição (ozônio) na atmosfera da Cidade do México.

### 4.2 A Parte Temporal do Modelo: Contagem por Processo de Poisson Não Homogêneo

Segundo [Ross et al. \(1996\)](#), um processo de contagem  $H(t), t \geq 0$  é de Poisson não homogêneo com função de intensidade  $\lambda(t)$ , quando atende às seguintes condições:

- $H(0) = 0$ ;
- Há independência entre os incrementos de  $H(t), t \geq 0$ ;
- $P(H(t+h) - H(t) \geq 2) = o_p(h)$ , i.e., a probabilidade de a diferença após o incremento ser maior que 1 é da mesma ordem que o incremento;
- $P(H(t+h) - H(t) = 1) = \lambda(t)h + o_p(h)$ .

No contexto espaço-temporal, há um processo de contagem para cada ponto  $\mathbf{s}_j$  do espaço  $\mathbb{R}^d$ . A equação usada para a intensidade foi a proposta por Goel:

$$\lambda(\mathbf{s}_j, t) = \theta_j \beta \alpha t^{(\alpha-1)} e^{-\beta t^\alpha},$$

e média  $m(\mathbf{s}_j, t) = \theta_j(1 - e^{-\beta t^\alpha})$ , sendo muito útil em modelagens ambientais. A parte temporal do modelo é a função:  $H(t | \boldsymbol{\theta}_H)$ , sendo esta parte um processo de Poisson não homogêneo com função de intensidade como a proposta por Goel (1983).

### 4.3 A parte espacial do modelo

A parte espacial é denotada por:

$$W(\mathbf{s} | \theta_w) \sim GP(\mu(\mathbf{s}), \sigma^2, \rho_\phi),$$

sendo  $GP(\cdot)$  um processo Gaussiano não estacionário e isotrópico de variância  $\sigma^2$  e  $\rho_\phi$  é a correlação entre dois pontos com certa distância euclidiana e sendo  $\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})\boldsymbol{\Psi}$  com  $\boldsymbol{\Psi}$  sendo a média do processo e  $\mathbf{x}(\mathbf{s})$ , o vetor de covariáveis.

### 4.4 Estimação dos parâmetros

$H$  é a parte temporal, um processo generalizado de Poisson não homogêneo de Goel (ou seja,  $\lambda$ , a taxa do processo, não é fixa). Os parâmetros  $\alpha$  e  $\beta$  são parâmetros da parte temporal da equação;  $\sigma^2$ ,  $\phi$  e  $\boldsymbol{\Psi}$  são da parte espacial (processo espacial gaussiano não estacionário e isotrópico). As prioris propostas para cada um dos parâmetros são:

$$\beta \sim G(a_\beta, b_\beta);$$

$$\alpha \sim G(a_\alpha, b_\alpha);$$

$$\boldsymbol{\Psi} \sim N(\mathbf{m}, \mathbf{v});$$

$$\sigma^2 \sim G(a_{\sigma^2}, b_{\sigma^2});$$

$$\phi \sim G(a_\phi \eta, \eta);$$

$$\mathbf{W} \sim N(\mathbf{X}\boldsymbol{\Psi}, \Sigma).$$

Considerando as marginais independentes, tem-se a priori de  $\Phi = (\beta, \alpha, \boldsymbol{\Psi}, \mathbf{W}, \phi, \sigma^2)$  :

$$\pi(\Phi) = \pi(\beta)\pi(\alpha)\pi(\boldsymbol{\Psi})\pi(\mathbf{W} | \phi, \sigma^2)\pi(\phi)\pi(\sigma^2).$$

A função de verossimilhança de  $\Phi$  dadas as  $n$  observações  $\mathbf{D} = (D_1, D_2, \dots, D_n)$  é expressa por:

$$L(\Phi | \mathbf{D})$$

A função de verossimilhança é:

$$L(\Phi | \mathbf{D}) = (\beta\alpha)^{\sum_{j=1}^n (n_j)}$$

$$\begin{aligned} & \times \exp \left( -\beta \sum_{j=1}^n \sum_{i=1}^{n_j} t_{ij}^\alpha + \sum_{j=1}^n n_j W_j - \sum_{j=1}^n e^{W_j} (1 - e^{-\beta T^\alpha}) \right) \\ & \quad \times \prod_{j=1}^n \prod_{i=1}^{n_j} t_{ij}^{\alpha-1}. \end{aligned}$$

A posteriori será expressa por:

$$\pi(\Phi \mid \mathbf{D}) \propto L(\Phi \mid \mathbf{D}) \pi(\beta) \pi(\alpha) \pi(\Psi) \pi(\mathbf{W} \mid \phi, \sigma^2) \pi(\phi) \pi(\sigma^2).$$

Coletando dados  $\mathbf{D}$  de  $n$  estações de observação, sendo os dados os dias em que a média diária ultrapassou o limite aceitável definido pra característica de interesse, encontram-se estimativas para  $\Phi$  a partir da esperança da posteriori.

No entanto, não há forma fechada para a posteriori e não é possível se aproximar de forma não iterativa, sendo necessário tomar amostras das condicionais da distribuição a posteriori  $\pi(\Phi \mid \mathbf{D})$  via MCMC. Para este modelo, deve-se obter estimativas para os parâmetros por meio de um amostrador de Gibbs. Há formas fechadas para as condicionais dos parâmetros  $\beta, \phi$  e  $\Phi$ . Não há formas fechadas para as condicionais  $\alpha, \sigma^2$  e  $\mathbf{W}$ , sendo necessário usar passos de Metropolis-Hastings para gerar valores ao acaso.

## 5 Pacote GeoPoisson

### 5.1 Estrutura Geral

No pacote, há uma função principal (`GeoPoisson`) que estima os parâmetros do modelo via MCMC. Na primeira fase, descartam-se as iterações iniciais (denominado *burn-in*). Após a convergência, as iterações aproveitadas fornecem intervalos de credibilidade para os parâmetros. Dentro da função principal, a taxa de aceitação dos passos de Metropolis-Hastings foi deixada em 44% e 24% para os casos univariado e multivariado, respectivamente, seguindo o indicado em [Neto \(2012\)](#) e [Roberts et al. \(1997\)](#) de forma aproximada.

Há diversas funções auxiliares para simular valores das distribuições condicionais dos parâmetros e uma para transformar os dados das observações de estações numa matriz própria para o programa. Por último, há uma função de interpolação espacial (`Interpolationgrid`). Com base nos valores estimados para os parâmetros na função anterior, são projetados os valores esperados para as variáveis de interesse num *grid* espacial quadrado das proximidades da área monitorada pelas estações. O procedimento é baseado em conhecimentos de distribuição normal multivariada como vista em [Rencher \(1998\)](#).

### 5.2 Criação de pacotes no R

Em [Leisch \(2008\)](#), há um guia sucinto para criação de pacotes no R. No tutorial, o usuário aprende os passos enquanto cria um pacote para regressão linear (`linmod`). Antes de tudo, deve-se preparar as funções. Algumas destas aparecerão na interface do usuário - é o meio para que estes insira dados, peça análises. Outras cumprem papel auxiliar, não sendo importante ou desejável que o usuário tenha acesso (dependendo da complexidade, o ideal é que nem se possa mexer nestas funções, cabendo somente a desenvolvedores estabelecer as rotinas).

O padrão é que todo pacote no R tenha um `print` e um `summary`. A primeira serve para expor, sucintamente, o resultado de uma função; a segunda fornece uma descrição mais especificada do procedimento (para uma função que calcula quantis de uma distribuição, sua função `print` retornaria apenas a média, enquanto que o `summary` retornaria quantis importantes para os dados, desvio-padrão e, em funções mais complexas, até mesmo o número de iterações para a convergência de um algoritmo).

Em seguida, deve-se entender o que são classes e métodos. Uma classe é uma forma



de se padronizar objetos que são usados ou extraídos das funções do pacote. O método é o procedimento para se trabalhar com o objeto padronizado. Um pacote deve ter coerência interna na hora de realizar suas rotinas. Se o usuário tivesse de se preocupar constantemente em transformar "dataframes" em matrizes ou matrizes em polígonos dependendo do que as funções requerem, ou ainda, se o programador precisasse converter qualquer formatação de dados inseridos pelo usuário, o uso do R seria gravemente prejudicado. Por isso, a forma dos dados já deve ser ajustada pelas funções e os manuais devem explicar ao usuário quais objetos são válidos naquele contexto e como ele deve inserir informação nas funções. No R, há 2 sistemas, informalmente conhecidos como *S3* e *S4*. O *S3* é o mais antigo e mais interativo dos dois. O *S4* é mais complexo, moderno e menos interativo.

Por fim, basta deixar as funções no R e colocar no final o comando `package.skeleton()`, inserindo um nome para o pacote e um diretório. No *Linux*, os comandos no terminal: `R CMD check` e `R CMD build`, respectivamente, checam se há erros na construção do pacote e fazem a versão final. O arquivo resultante poderá ser disponibilizado no *Github* para correções e, posteriormente, no *CRAN* para uso da comunidade. Vale ressaltar que o pacote deve ter seus manuais devidamente preenchidos, autoria, tipo de licença, instrução de uso e algum banco de dados em que se aplique o uso.

O arquivo compactado já pode ser instalando no R na aba `Tools > Install Packages` e selecionando o arquivo. Mais detalhes podem ser vistos no artigo de [Leisch \(2008\)](#), que também explica o procedimento para criação de pacotes do R no *Windows*.

### 5.3 A Função Principal: GeoPoisson

Esta função prepara os dados para obtenção de estimativas por meio da estrutura do amostrador de Gibbs. Inicialmente, o usuário terá de fornecer os seguinte argumentos:

- `data`: são os dados das séries temporais das estações. Pode estar em formato `data.frame` ou `matrix`;
- `limuser`: valor limite acima do qual se contabiliza um evento de contagem Poisson (poluição acima do aceitável, chuva acima do normal, etc);
- `c1`: um hiperparâmetro a ser usado nos parâmetros do model, 0,001 por padrão;
- `d1`: um segundo hiperparâmetro, 0,001 por *default*;
- `iter`: número de iterações totais no amostrador de Gibbs;
- `burn_in`: número de iterações descartadas. A convergência costuma demorar, é sugerido ao usuário algum valor superior a 100.000;
- `X`: matriz de covariáveis do modelo;

- `loca`: matriz de coordenadas (longitude e latitude) das estações.

Depois que o usuário fornecer essas informações, o programa ajustará o formato da matriz tomando os dias em que o `limuser` foi superado. Este procedimento será descrito em detalhes no próximo capítulo.

Em seguida, uma série de chutes iniciais serão atribuídos aos parâmetros do modelo. Por fim, as informações entram num laço com o número de iterações equivalendo ao que foi definido na variável "iter". Para os passos de Metropolis-Hastings, haverá averiguação da conformidade da taxa de aceitação: 0,44 e 0,24 para os casos univariado e multivariado respectivamente. Caso a proporção não seja conforme, há uma alteração proposta em  $X$ .

Terminadas as iterações, o usuário terá uma lista com as cadeias de observações dos parâmetros após o `burn_in` e taxas de aceitação a cada 50 iterações dos passos de Metropolis-Hastings. Para usar a função de interpolação, será necessário informar os dados obtidos desta função.

## 5.4 Funções Auxiliares à função GeoPoisson

A lista de funções auxiliares à principal são:

- Funções auxiliares para cada parâmetro, sorteando um valor para cada chamada do amostrador de Gibbs: `samplealpha`, `samplebeta`, `sampleb`, `samplev`, `sampleW`, `samplePsi`. Estas funções usam valores anteriores dos demais parâmetros para simular um novo do parâmetro atual. As amostragens de `samplebeta`, `samplev` e `samplePsi` são feitas pelas distribuições condicionais. As demais funções usam passos de Metropolis-Hastings;
- Uma função que transforma séries temporais das observações de uma característica de interesse em uma matriz com os dias em que o limite estabelecido pelo usuário para esta característica foi superado: `matrixtransf`;
- Funções de cálculo de covariância espacial e correlação `gSigma` e `gCor`;

## 5.5 Interpolador

Até o momento, o programa somente encontrou parâmetros com base nas observações das  $n$  estações. No entanto, é comum que um pesquisador deseje saber o comportamento do mesmo fenômeno num ponto não observado próximo das estações em um certo momento  $t$ . Para esse fim, recorre-se à interpolação espacial. Esta parte do capítulo foi baseada em [Rencher \(1998\)](#) e [Morales et al. \(2017\)](#).

Seja  $\mathbf{W}$  o vetor de  $n$  pontos com as estações e  $\mathbf{W}^{NO}$  o vetor de  $m$  pontos não observados e que se quer estimar indiretamente.

Como foi estabelecido no modelo:

$$\begin{pmatrix} \mathbf{W} \\ \mathbf{W}^{NO} \end{pmatrix} \sim N(\mathbf{A}, \Sigma),$$

sendo:

$$\mathbf{A} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix},$$

e:

$$\Sigma = \begin{pmatrix} \Sigma_{A_1} & \Sigma_{A_{12}}^\top \\ \Sigma_{A_{12}} & \Sigma_{A_2} \end{pmatrix},$$

sendo as matrizes  $\Sigma$  com seus índices as matrizes de variância e covariância das partições  $A_1$  e  $A_2$ ,  $\mathbf{A}_1 = \mathbf{X}_{A_1} \Psi$ ,  $\mathbf{A}_2 = \mathbf{X}_{A_2} \Psi$ , de modo que as matrizes de covariáveis  $\mathbf{X}_{A_1}$  e  $\mathbf{X}_{A_2}$  são associadas com  $\mathbf{W}$  e  $\mathbf{W}^{NO}$ .

Por propriedades das distribuições Normais multivariadas, é sabido que:

$$\mathbf{W}^{NO(r)} | \mathbf{W} \sim N(A_2^*, \Sigma_{A_2}^*).$$

A partir das cadeias de  $\mathbf{W}^{(r)}$ ,  $\Psi^{(r)}$ ,  $\sigma^{2(r)}$  e  $\phi^{(r)}$  - com  $r$  denotando o  $r$ -ésimo dos  $M$  elementos - obtidas da função principal é possível encontrar valores de  $\mathbf{W}^{NO}$ , com os seguintes passos:

- Encontrando as matrizes da covariância das partições:

$$\Sigma_{ij}^r = \sigma^{2(r)} \exp[-\phi^{(r)}] |s_i - s_j|$$

- Encontrando as partições na iteração:

$$\mathbf{A}_1^{(r)} = \mathbf{X}_{A_1} \Psi^{(r)};$$

$$\mathbf{A}_2^{(r)} = \mathbf{X}_{A_2} \Psi^{(r)};$$

- Buscar a partição não observada, ajustada por iteração, e a matriz ajustada de covariâncias:

$$\begin{aligned} \Sigma_{A_2}^{*(r)} &= \Sigma_{A_2}^{(r)} - \Sigma_{A_2}^* \Sigma_{A_1}^{-1(r)} \Sigma_{A_{12}}^{(r)}; \\ \mathbf{A}_2^{*(r)} &= \mathbf{A}_2^{(r)} + \Sigma_{A_{12}} \Sigma_{A_1}^{-1} (\mathbf{W} - \mathbf{A}_1); \end{aligned}$$

- Tomando as amostras de  $\mathbf{W}^{NO}$ :

$$\mathbf{W}^{NO} | \mathbf{W} \sim N(A_2^*, \Sigma_{A_2}^*)$$

- Por fim, interpola-se média de dias acima do limite ( $\lambda$ ):

$$E[\lambda(\mathbf{s}_i, t)] = \frac{1}{M} \sum_{r=1}^M \exp(W_i^{(r)}) (1 - \exp(-\beta^{(r)} t^{\alpha^{(r)}})).$$

A função `Interpolationgrid` recebe os seguintes argumentos:

- `Geodata`: a lista recebida da função `GeoPoisson`;
- `loca`: a matriz de localização das estações;
- `limuser`: limite aceitável da característica de interesse;
- `data`: séries temporais das observações nas estações por dia;
- `size`: o número de quadrados ( $size \times size$ ). Se o usuário inserir 20, significa que ele quer a interpolação para  $20 \times 20$  estações na região vizinha às estações observadas. Recomenda-se moderação (máximo de 30) para esse parâmetro da função;
- `intensity`: indica quão escuro é o *grid* cobrindo o mapa. Valor entre 0 e 1. Por *default*, o valor é de 0,75.

No final, gera-se um mapa com o número esperado de vezes que a característica de interesse ultrapassará o limite definido por região vizinha não observada .

## 5.6 Utilização das funções

Para utilizar as funções, o usuário somente tem acesso direto a duas funções principais: `GeoPoisson` e `Interpolationgrid`. Esta última depende de um objeto que armazene os resultados da primeira.

A função `lstinlineGeoPoisson` depende de 8 parâmetros (já vistos em 5.3). Espera-se que o usuário forneça um *burn-in* elevado o suficiente para garantir a convergência dos parâmetros (de 100.000 até 300.000 em alguns casos).

Inicialmente, é necessário carregar as séries temporais de observações das estações em forma de matriz (cada coluna sendo uma estação):

```
data <- read.table("MatrizDeObservacoes", head=T)
```

Em seguida, é necessário carregar a matriz de localização das estações e a matriz de planejamento  $X$  (covariáveis e localizações).

```
loca <- read.table("LocalizacaoDasEstacoes", head=T)
```

```
X <- read.table("MatrizDePlanejamento", head=T)
```

Depois disso, o usuário pode rodar a função principal (ciente do limite observado em `limuser`, número de iterações e `burn_in`):

```
dados <- GeoPoisson(data, limuser, c1=0.001, d1=0.001, iter, burn_in, X, loca)
```

O usuário receberá um objeto com as cadeias de valores aleatórios nas últimas iterações, isto é, excluindo-se as geradas no *burn-in*. No momento, como o pacote ainda está sendo ajustado, não há um retorno automático do *summary* dos dados. Em sua atualização, o pacote deverá oferecer ao usuário um sumário das estimativas dos parâmetros.

O próximo passo é interpolar os dados num mapa. A função `Interpolationgrid` realizará isso:

```
Interpolationgrid(Geodata, loca, limuser, data, size=21, intensity=0.85)
```

Após este comando, o usuário deve ver um mapa da região ao redor das estações e valores para zonas não observadas. O mapa é sempre quadrado e as dimensões devem ser fornecidas na função. Isso será claramente observado no próximo capítulo.

Todo este capítulo e o próximo foram realizados com auxílio do R ([R Development Core Team, 2009](#)).

## 6 Resultados

### 6.1 Aplicação

Para testar a eficácia do pacote e suas funções, foram usados dados de precipitação (mm) de 29 estações pluviométricas no Piauí e Maranhão da Agência Nacional de Águas (ANA). As séries são observações diárias de 01/01/1980 a 31/12/2010.

Inicialmente, foi executada a função principal, `GeoPoisson`.

```
dados <- GeoPoisson(data=DataANA, limuser=48,8, c1=0.001, d1=0.001,
  iter=300000, burn_in=200000, X, loca)
```

Logo nas primeiras linhas de código, há na `GeoPoisson` uma chamada da função auxiliar `matrixtransf` que transforma os dados para uma formatação apropriada para a estimação. Com o `limuser`, checka-se em que dias houve mais chuva que o estabelecido como aceitável. Esta função começa extraindo os dias que ultrapassam o limite. Tomando a estação 2 das observações citadas, a Figura 6.1 mostra como o algoritmo procede para detectar valores além do limite.

Após isso, tendo cada estação como coluna, constrói-se a matriz de dias em que o limite foi ultrapassado. Para completar a matriz, preenche-se com NA as colunas incompletas. Um exemplo genérico pode ser verificado na matriz de estações.

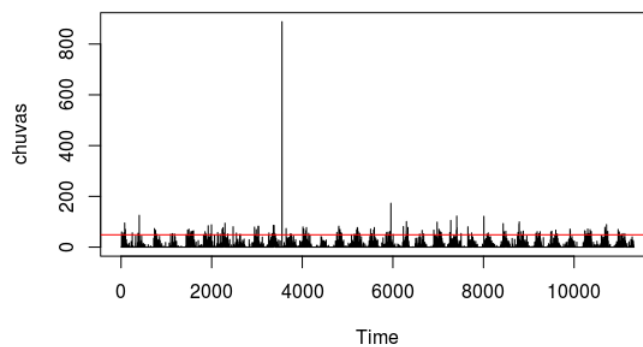


Figura 6.1 – Observações da estação 2 acima do limite no tempo em dias.

	Média	50%	2.5%	97.5%
$\alpha$	0,993	0,993	0,964	1,021
$\beta$	$1,42710^{-5}$	$1,40610^{-5}$	$1,05810^{-5}$	$1,90710^{-5}$
$\phi$	0,186	0,067	0,006	1,047
$\sigma^2$	2,074	0,830	0,070	10,378
$\psi_0$	5,307	5,421	-5,733	16,012
$\psi_1$	-0,037	-0,035	-0,275	0,196
$\psi_2$	0,023	0,025	-0,230	0,269

Tabela 6.1 – Quantis empíricos da simulação das distribuições a posteriori dos parâmetros.

<i>Estação</i>	<i>Estação1</i>	<i>Estação2</i>	<i>Estação3</i>	<i>Estação29</i>
<i>1º Dia</i>	1	3	7	3
<i>2º Dia</i>	4	5	NA	8
<i>3º Dia</i>	6	NA	NA	NA
<i>4º Dia</i>	9	NA	NA	NA

Em seguida, um total de 300000 iterações foram realizadas, com um *burn-in* de 200000 iterações. No algoritmo, cada etapa do amostrador de Gibbs é executada um número de vezes definido em *niter*, e para cada parâmetro, é feita uma chamada para a função auxiliar que gera valores - ou por amostrador de Gibbs (para  $\beta, var$  e  $\Psi$ ) puro, ou com passos de Metropolis-Hastings (demais parâmetros). Durante o *burn-in*, os parâmetros gerados com passo de Metropolis-Hastings tiveram ajustes na taxa de aceitação a cada 50 iterações. O objetivo era alcançar a taxa de aceitação recomendada. O procedimento demorou cerca de meia-hora e o tempo despendido é proporcional ao número de iterações.

A informação obtida de GeoPoisson foi sumarizada na Tabela 6.1.

## 6.2 Diagnóstico das funções do amostrador de Gibbs com passos de Metropolis-Hastings

Ainda na função principal (GeoPoisson), implementou-se uma importante ferramenta de diagnóstico: um contador de aceitação/rejeição dos passos de Metropolis-Hastings. A cada 50 iterações, contava-se o percentual de aceitação. O alvo era de 0,44 para caso univariado de acordo com [Neto \(2012\)](#) e 0,24 para o caso multivariado, como visto em [Roberts et al. \(1997\)](#). Estes valores alvo evitariam uma variação muito grande no passo - o que poderia fornecer estimativas pouco confiáveis, ou muito pequena - demandando um número muito maior de iterações até a convergência. Esta parte, no entanto, não é diretamente acessível. Buscou-se mostrar a taxa de aceitação apropriada, sem que o usuário se preocupasse com o mecanismo.

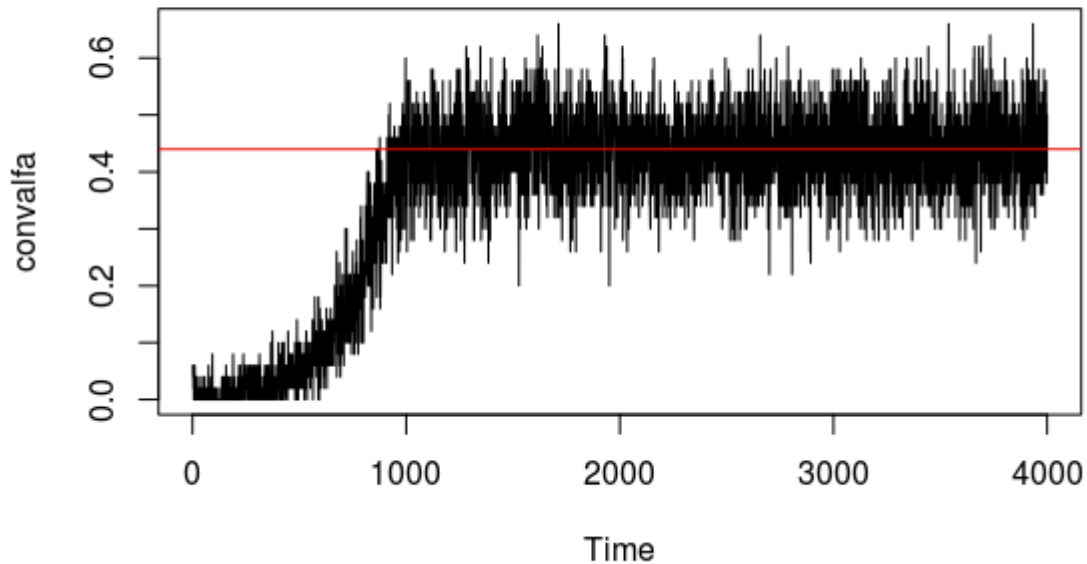


Figura 6.2 – Taxa de aceitação do amostrador de Gibbs para  $\alpha$ .

Para  $\alpha$ , a média de aceitação das últimas 150000 (3000 coletas de 50 repetições) iterações do *burn-in* foi de aproximadamente 0,439, compatível com o esperado de 0,44. O período inicial é desconsiderado, pois o algoritmo ainda está ajustando o passo. A Figura 6.2 apresenta graficamente esse resultado.

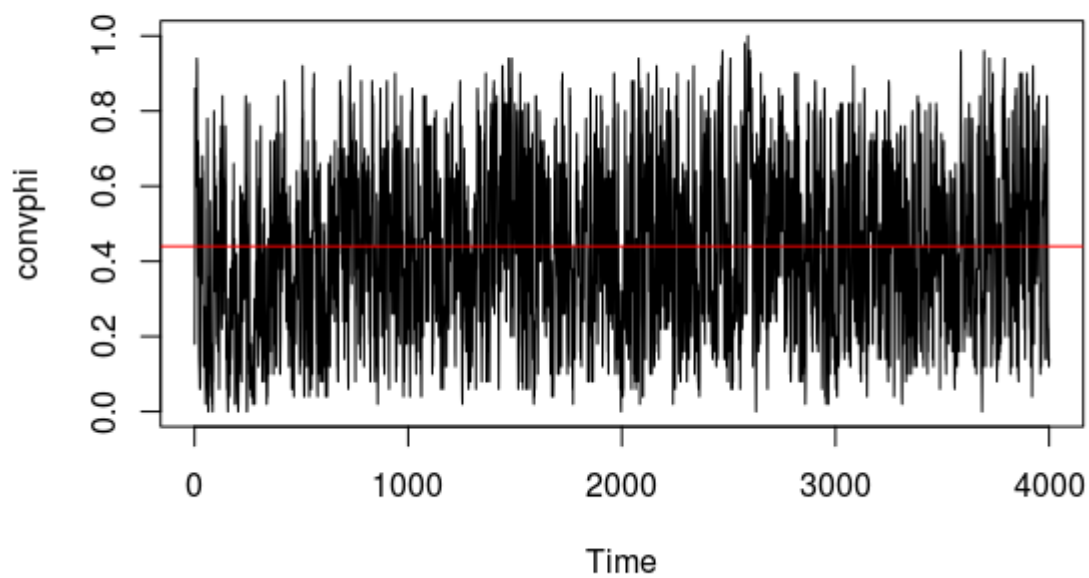
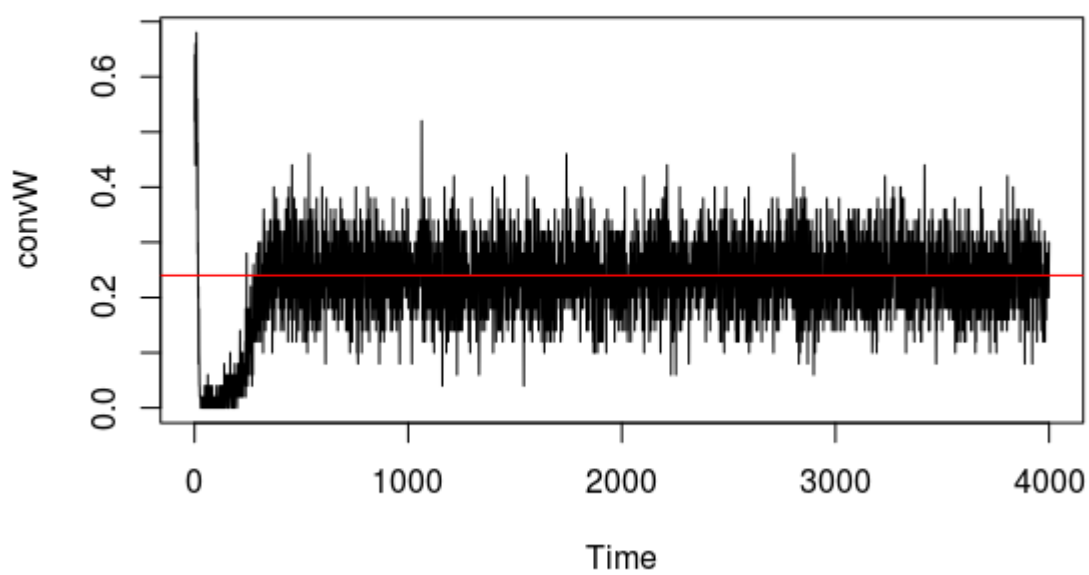
Para  $\phi$ , a média de aceitação nas últimas 150000 iterações do *burn-in* foi de aproximadamente 0.4470, como pode ser visto na Figura 6.3.

Para  $\mathbf{W}$ , a média de aceitação nas últimas 150000 iterações do *burn-in* foi de aproximadamente 0.2399, como se observa na Figura 6.4.

### 6.3 Mapa gerado do Interpolador

Após os resultados obtidos da função principal, executou-se a função de interpolação, prevendo os dias com chuvas acima do limite das regiões não observadas. A previsão com



Figura 6.3 – Taxa de aceitação do amostrador de Gibbs para  $\phi$ .Figura 6.4 – Taxa de aceitação do amostrador de Gibbs para  $W$ .

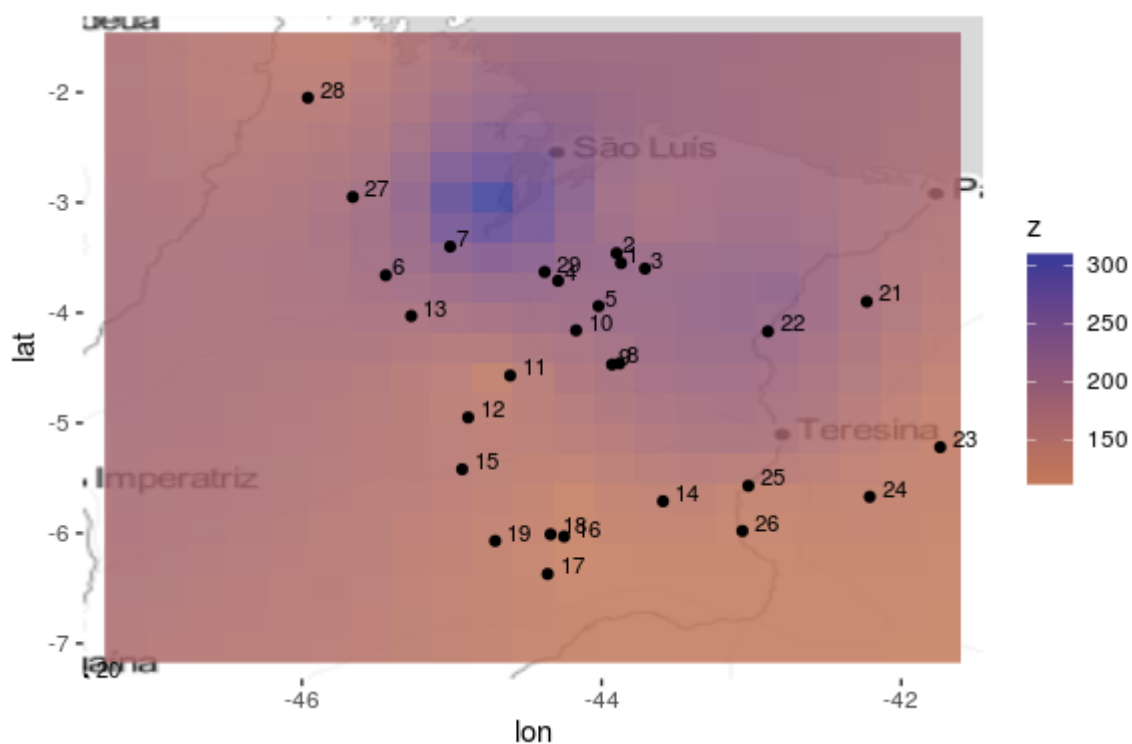


Figura 6.5 – Número esperado de dias com chuvas acima do limite no período de 30 anos.

Nele é possível observar que a região mais próxima a São Luís é também a que se espera maior frequência de chuvas além do limite. No interior, como esperado, a frequência desses dias é menor. A Figura 6.5 é o resultado da função `Interpolationgrid`.

## 7 Considerações Finais

Este trabalho contemplou diversos assuntos de estatística (espacial, inferência bayesiana, simulação estocástica, cadeias de Markov) e explicou o modelo geoespacial com processo de Poisson não homogêneo de [Morales et al. \(2017\)](#), com o objetivo de criar um novo pacote da linguagem R - GeoPoisson - baseado no artigo.

Considerando as diversas aplicações do modelo, as funções foram bem sucedidas em estimar os parâmetros e gerar projeções espaciais para regiões não observadas. Foram necessárias 300.000 iterações, com *burn-in* de 200.000 iterações, para que os parâmetros convergissem e intervalos de credibilidade fossem construídos. Com esses resultados, a função de interpolação espacial retornou um mapa de previsão para regiões não observadas. Os mapas serão melhores à medida em que o número de estações de observação aumentem.

Usuários das ciências ambientais e correlatas podem usufruir do pacote GeoPoisson sem conhecimentos profundos na teoria, apenas tendo noções de R e do funcionamento de algoritmos de simulação MCMC (Simulação de Monte Carlo via Cadeia de Markov). Essa deve ser uma ferramenta poderosa no estudo e prevenção catástrofes (desastres pluviométricos, concentrações letais de gases poluentes, entre outras).

O próximo passo é disponibilizar o algoritmo no *Github* para receber a avaliação de usuários. Com estas críticas, sugestões e melhorias, o pacote deve ser modificado antes de ser lançado ao público no CRAN (ambiente do R em que os pacotes podem ser instalados). Ademais, é possível acrescentar funções importantes, como diagnóstico de convergência por meio do pacote coda [Plummer et al. \(2006\)](#) e geração das cadeias por paralelismo.

## Referências

- BOLSTAD, W. M.; CURRAN, J. M. *Introduction to Bayesian statistics*. USA: John Wiley & Sons, 2016.
- CRESSIE, N. A. C. *Statistics for Spatial Data*. USA: John Wiley & Sons, Inc., 1991.
- GAMERMAN, D.; LOPES, H. F. *Markov chain Monte Carlo: stochastic simulation for bayesian inference*. 2nd. ed. Boca Raton: Chapman & Hall / CRC, 2006.
- GNEITING, T.; GUTTORP, P. *Statistical Methods for Spatial Data Analysis*. Boca Raton: Chapman & Hall / CRC, 2005.
- LEISCH, F. *Compstat 2008-Proceedings in Computational Statistics*. Heidelberg, 2008.
- MORALES, F. E. C. et al. A nonhomogeneous poisson process geostatistical model. *Stochastic environmental research and risk assessment*, Springer, v. 31, n. 2, p. 493–507, 2017.
- NETO, J. H. V. Introduzindo covariáveis na estrutura de covariâncias de processos espaciais. UFRJ, Hoboken, New Jersey, 2012.
- PLUMMER, M. et al. *CODA: Convergence diagnosis and output analysis for mcmc*. USA, 2006. v. 6, n. 1, 7–11 p. Disponível em: <<https://journal.r-project.org/archive/>>.
- R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria, 2009. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>.
- RENCHE, A. C. *Multivariate Statistical Inference and Applications*. New York: John Wiley & Sons Inc., 1998.
- ROBERTS, G. O. et al. Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, Institute of Mathematical Statistics, v. 7, n. 1, p. 110–120, 1997.
- ROSS, S. M. et al. *Stochastic processes*. New York: John Wiley & Sons Inc., 1996. v. 2.
- SCHABENBERGER, O.; GOTWAY, C. A. *Statistical Methods for Spatial Data Analysis*. Boca Raton: Chapman & Hall / CRC, 2005.
- SCHMIDT, A. M.; SANSONO, B. *Modelagem Bayesiana da Estrutura de Covariância de Processos Espaciais e Espaços-Temporais*. São Paulo: ABE - Associação Brasileira de Estatística, 2006.
- TURKMAN, M. A. A.; PAULINO, C. D. *Estatística Bayesiana Computacional: uma introdução*. Lisboa: Sociedade Portuguesa de Estatística, 2015.