

**Flávio Pereira Martins da Silva**

**Modelando dados geostatísticos através da  
função de covariância exponencial sob o  
enfoque Bayesiano**

Niterói - RJ, Brasil

10 de julho de 2019

**Flávio Pereira Martins da Silva**

**Modelando dados geoestatísticos  
através da função de covariância  
exponencial sob o enfoque Bayesiano**

**Trabalho de Conclusão de Curso**

Monografia apresentada para obtenção do grau de Bacharel em  
Estatística pela Universidade Federal Fluminense.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Patrícia Lusié Velozo da Costa

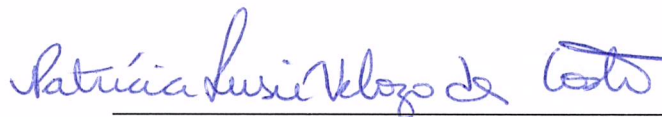
Niterói - RJ, Brasil

10 de julho de 2019

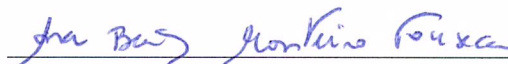
Flávio Pereira Martins da Silva

**Modelando dados geoestatísticos através da  
função de covariância exponencial sob o  
enfoque Bayesiano**

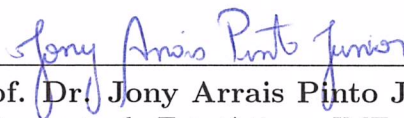
Monografia de Projeto Final de Graduação sob o título  
*“Modelando dados geoestatísticos através da função de  
covariância exponencial sob o enfoque Bayesiano”*, defendida  
por Flávio Pereira Martins da Silva e aprovada em 10 de julho  
de 2019, na cidade de Niterói, no Estado do Rio de Janeiro,  
pela banca examinadora constituída pelos professores:



**Prof<sup>ª</sup>. Dr<sup>ª</sup>. Patrícia Lusié Velozo da Costa**  
Departamento de Estatística – IME – UFF



**Prof<sup>ª</sup>. Dr<sup>ª</sup>. Ana Beatriz Monteiro Fonseca**  
Departamento de Estatística – IME – UFF



**Prof. Dr. Jony Arrais Pinto Junior**  
Departamento de Estatística – IME – UFF

Ficha catalográfica automática - SDC/BIME  
Gerada com informações fornecidas pelo autor

D111m Da silva, Flávio Pereira Martins  
Modelando dados geoestatísticos através da função de  
covariância exponencial sob o enfoque Bayesiano / Flávio  
Pereira Martins Da silva ; Patrícia Velozo, orientadora.  
Niterói, 2019.  
50 f. : il.

Trabalho de Conclusão de Curso (Graduação em  
Estatística)-Universidade Federal Fluminense, Instituto de  
Matemática e Estatística, Niterói, 2019.

1. Estatística Espacial. 2. Inferência Bayesiana. 3. MCMC.  
4. Geoestatística. 5. Produção intelectual. I. Velozo,  
Patrícia, orientadora. II. Universidade Federal Fluminense.  
Instituto de Matemática e Estatística. III. Título.

CDD -

# Resumo

Diversos fenômenos tais como precipitação, ocorrências de doenças, entre outros, são espacialmente correlacionados e, por isso, este trabalho consiste em modelar dados geostatísticos impondo uma estrutura de dependência espacial. Diferentes funções na literatura permitem fazer essa estrutura. Esse trabalho concentra-se na função de covariância exponencial. Os parâmetros desconhecidos são estimados sob o enfoque Bayesiano. Na estatística bayesiana, a amostra e a crença do pesquisador são levadas em consideração na estimação paramétrica através da distribuição *a posteriori*. Porém, a avaliação direta dessa distribuição muitas vezes é custosa ou não é possível. Uma alternativa nesses casos é a utilização dos métodos de **MCMC!** (**MCMC!**) e, em especial, o amostrador de Gibbs e o algoritmo de Metropolis-Hasting. Dados simulados são gerados para analisar a eficiência do modelo proposto de convergir para o valor verdadeiro e analisar a sensibilidade que a estimação possui com diferentes distribuições *a priori*. Finaliza-se o trabalho aplicando um conjunto de dados reais ao modelo proposto.

Palavras-chave: Estatística espacial. Inferência Bayesiana. MCMC. Geoestatística. Alcance.

# Dedicatória

*Este trabalho é dedicado aos jovens negros periféricos,  
que não tiveram o mesmo privilégio que o meu, poder cursar uma faculdade.*

*Orgulho por representar alguns deles.*

# Agradecimentos

Agradeço em primeiro lugar aos meus pais, pela educação que me deram e toda confiança depositada durante essa jornada. À minha mãe, Sônia, que durante esses anos me deu todo suporte e amor que estava ao seu alcance. Ao meu pai, Silvio, que por várias vezes me aconselhou e me engrandeceu. Aos meus avós, Josefa e João, por me abrigarem em sua casa enquanto meus pais trabalhavam. Ao meu tio Roberto, pelo presente inesperado, uma caixa de som, dada na época que eu estudava pela internet para o vestibular e que me ajudou muito na época do vestibular e durante a faculdade, à minha tia Sueli por ser minha principal inspiração em cursar Estatística, à minha madrinha, Luciana, pelas broncas(rs) e pelo seu carinho dado durante toda minha vida, ao meu padrinho, Gilberto, pela sua amizade. Ao restante da minha família, por sempre estar partilhando comigo o amor e a felicidade.

Aos meus melhores amigos que por vezes me incentivaram e me escutaram. Ao meu amigo Pedro Delgado que me indicou o curso de Estatística. À quem sofreu comigo quando os resultados eram ruins e quem sorriu comigo quando os resultados foram bons.

À Escola Técnica Estadual Adolpho Bloch por me apresentar a diversidade do ser humano e ser fundamental na minha formação como pessoa.

Aos meus amigos que a UFF me apresentou: “Bonde do Jubileu” que tornou o tão assustador primeiro período mais leve e engraçado, os amigos que tive o prazer de dividir trabalhos acadêmicos, os amigos que compartilharam horas de estudos pré-prova comigo, os amigos que por diversas vezes me ajudaram e tiveram paciência para me explicar conteúdos que eu estava com dificuldade de aprender e os amigos que na hora do lazer curtiram DCE, Cantareira, choppadas e viagens comigo.

À Associação Atlética Instituto de Matemática e Estatística da UFF(AAIME-UFF) que tive o prazer de ser um dos fundadores e participar da diretoria por quase 3 anos, me fez ter orgulho de representar o IME UFF nos campeonatos dentro e fora da UFF e além disso me apresentou pessoas maravilhosas.

À banda Forfun que por vezes fez a minha trilha sonora durante o caminho de casa para UFF.

À todos os professores que passaram na minha vida. Principalmente a minha orientadora, Patrícia, pela paciência, conselhos e empenho no trabalho de conclusão do curso.



# Sumário

# Lista de Figuras

# Lista de Tabelas

# Lista de Abreviações

**MCMC** Monte Carlo via cadeias de Markov

# 1 Introdução

Com o avanço dos computadores e o aumento dos bancos de dados, a estatística espacial vem se tornando cada vez mais importante dentro da sociedade ao relacionar os fenômenos de interesse com suas localizações espaciais. O estudo sobre fenômenos ocorridos em algum espaço, ajuda a compreender sobre esses fenômenos e auxiliam na prevenção de acidentes, por exemplo. Na área de Segurança, pode-se questionar se existe alguma relação entre a violência e diferentes cidades, por exemplo. Na Geografia, a magnitude de um terremoto pode estar relacionada com questões terrestres que variam conforme a localização espacial. Na área da Saúde, a ocorrência de dengue pode variar para diferentes regiões por existir alguma relação de contágio com as áreas com menos investimento em saneamento, por exemplo. Muitos trabalhos de estatística espacial podem ser encontrados atualmente tais como ??) e ??).

Há 3 tipos de padrões de dados espaciais: geoestatísticos, pontuais e de áreas. O presente trabalho consiste em analisar modelos espaciais para dados geoestatísticos. A geoestatística considera que as coordenadas da localização do fenômeno de interesse podem variar continuamente na região espacial. Essas coordenadas são conhecidas e a incerteza está no fenômeno de interesse. Acredita-se que regiões próximas são altamente correlacionadas e que quanto maior a distância entre 2 regiões, menor será a correlação entre seus fenômenos. Dados geoestatísticos podem ser encontrados nas áreas de minas, petróleo, agricultura, ciência do solo, etc. Exemplos da utilização de modelagem de dados geoestatísticos podem ser encontrados em ??) e ??). No primeiro artigo realiza-se uma análise de casos de hanseníase no Estado de São Paulo e no segundo modela-se a precipitação.

Na modelagem de dados geoestatísticos, a função de covariância mede a variação espacial entre dois fenômenos separados por uma certa distância. Existem diferentes funções de covariância para descrever essa relação, tais como a função exponencial, a Matérn e a esférica. Esse trabalho recorrerá a função de covariância exponencial, inspirado em ??).

Os parâmetros desconhecidos do modelo podem ser estimados através de diferentes abordagens. As principais são através das inferências clássica e Bayesiana. O presente trabalho utiliza a abordagem bayesiana. A estatística Bayesiana trata os parâmetros desconhecidos como variáveis aleatórias, diferentemente da estatística clássica que apresenta os parâmetros desconhecidos como quantidades fixas. Desta forma, pode-se atribuir uma distribuição de probabilidade para esses parâmetros. Além disso, na estatística Bayesiana o processo de inferência combina os dados obtidos na amostra com a informação prévia do pesquisador sobre o vetor paramétrico através da distribuição *a priori*. A combinação da distribuição da amostra com a distribuição *a priori* gera uma distribuição *a posteriori*. A estimação dos parâmetros baseia-se nessa distribuição *a posteriori*. Em ??) também recorre-se a abordagem bayesiana para inferir sobre os parâmetros de um modelo espacial proposto para analisar o índice de abundância de um estoque pesqueiro.

Em muitos casos a distribuição *a posteriori* é muito complexa de ser resolvida analiticamente, fazendo-se necessário recorrer a métodos computacionais tais como os métodos de **MCMC!**. Esses métodos foram utilizados em ??), em ??) e no presente trabalho.

O presente trabalho tem por objetivo principal modelar dados geoestatísticos recorrendo a inferência bayesiana e, em particular, utilizar os métodos de **MCMC!**.

O trabalho está organizado da seguinte maneira: no capítulo ?? há uma breve revisão da literatura sobre estatística espacial e, em particular, sobre Geoestatística e também descreve-se o método Bayesiano para estimação de parâmetros e, em especial, sobre os métodos de **MCMC!**; no Capítulo ??, dados serão simulados visando testar a eficiência do procedimento de estimação dos parâmetros desconhecidos e, com isso, pode-se analisar se os valores verdadeiros dos parâmetros conseguem ser bem estimados, o tempo computacional gasto para isso e a sensibilidade que o modelo tem com a atribuição da distribuição *a priori*. Após o estudo simulado, um conjunto de dados reais será aplicado. **Este conjunto contém a concentração de metais pesados ao redor do rio Meuse.**; o Capítulo ?? contém as conclusões do trabalho.

## 2 Materiais e Métodos

Esse trabalho consiste em propor modelos espaciais para dados geoestatístico e, por isso, há uma revisão sobre modelos espaciais na Seção ?? e, em particular, apresenta-se uma discussão sobre geoestatística na Subseção ?. Nos modelos propostos há parâmetros desconhecidos que são estimados sob a abordagem bayesiana, assim, a Seção ?? contém uma revisão sobre inferência bayesiana. Os métodos de **MCMC!** costumam ser utilizados nessa abordagem, por isso na Subseção ?? é apresentado um resumo desses métodos.

### 2.1 Estatística Espacial

A análise espacial é definida como a compreensão das propriedades e dos relacionamentos dos dados, levando em consideração a sua localização espacial. Ou seja, estuda a influência espacial em um determinado fenômeno.

Relacionar fenômenos de interesse com suas localizações espaciais é um grande desafio. Entender esse relacionamento é útil em diversas áreas tais como as áreas da saúde, geologia, ecologia, agronomia, ambientais, entre outros.

A compreensão visual da distribuição espacial torna-se muito eficiente na hora de verificar se existe algum padrão dessa distribuição. Devido a disponibilidade do **SIG!** (**SIG!**) os estudos espaciais estão se tornando mais comum, possibilitando modelagens mais eficientes dentro desse contexto. O **SIG!** possibilita a integração de mapas e gráficos com um banco de dados.

A estatística espacial é dividida em três áreas: geoestatística, dados pontuais e dados de área. Na geoestatística as coordenadas do fenômeno de interesse podem variar continuamente em uma dada região espacial e a incerteza está no fenômeno de interesse. A geoestatística costuma supor que os dados são estacionários e que a correlação espacial depende apenas da distância e é útil para modelar a quantidade de chuva ou a concentração de ouro, por exemplo. Nos dados pontuais a incerteza está na localização espacial do

evento estudado e útil para modelar a ocorrência de violência pois a incerteza, nesse caso, está em onde ocorrerá o caso de violência. Já nos dados de áreas os fenômenos de interesse estão agrupados por subregiões da região analisada. Por exemplo, o número de indivíduos com dengue pode ser agrupado por cidade. Nesse caso, as coordenadas do fenômeno não variam continuamente na região espacial.

Nesse presente trabalho é dada ênfase na área da geoestatística. Mais informação sobre estatística espacial pode ser obtida em ??) e ??).

### 2.1.1 Geoestatística

A geoestatística admite outras nomeclaturas como superfícies contínuas e superfícies aleatórias e estuda fenômenos que ocorrem em uma determinada região. As coordenadas dessa região podem ser a latitude, a longitude e a altitude, por exemplo. O fenômeno de interesse pode assumir diferentes valores em qualquer ponto dessa região e, por isso, é dito que as coordenadas variam continuamente na região espacial.

A modelagem de dados geoestatísticos pode ser expressa pela soma de duas componentes: uma componente, denotada por  $\mu(\mathbf{s})$ , que descreve a média do processo podendo conter intercepto, covariáveis, componente de tendência ou sazonalidade e uma componente, denotada por  $\epsilon(\mathbf{s})$ , que descreve a estrutura espacial que não foi capturada pela componente anterior. Dessa forma, a variável aleatória de interesse  $Z$ , com vetor de coordenada  $\mathbf{s}$ , pode ser escrita da seguinte forma

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}). \quad (2.1)$$

Costuma-se supor que a variável  $\epsilon(\mathbf{s})$  seja estacionária e, portanto, tem média constante para todas as localizações. E como a componente  $\mu(\mathbf{s})$  pode conter intercepto, costuma-se supor a componente  $\epsilon(\mathbf{s})$  tem média zero para não ter problemas de identificabilidade. Assim, o valor esperado para a variável aleatória  $\epsilon$  nas posições  $\mathbf{s}$  e  $\mathbf{s} + \mathbf{h}$  são iguais a zero para quaisquer localizações com coordenadas  $s$  e  $s + h$  pertencentes à região de analisada. Isto significa que a esperança da diferença dos processos em  $\mathbf{s}$  e  $\mathbf{s} + \mathbf{h}$ , que são separados por um vetor distância  $\mathbf{h}$ , é nula, ou seja,

$$E[\epsilon(\mathbf{s}) - \epsilon(\mathbf{s} + \mathbf{h})] = 0. \quad (2.2)$$

Além disso, devido a suposição de estacionariedade, tem-se que a variância do processo  $\epsilon$  não depende da localização e que a distribuição conjunta de  $(\epsilon(\mathbf{s}), \epsilon(\mathbf{s} + \mathbf{h}))$ , para



quaisquer coordenadas  $\mathbf{s}$  e  $\mathbf{s} + \mathbf{h}$ , na região espacial analisada, depende somente do vetor de separação  $\mathbf{h}$ . Ou seja,

$$\begin{aligned} \text{Var}(\epsilon(\mathbf{s})) &= V, \\ C(\mathbf{h}) &= \text{Cov}[\epsilon(\mathbf{s}), \epsilon(\mathbf{s} + \mathbf{h})] = E[\epsilon(\mathbf{s})\epsilon(\mathbf{s} + \mathbf{h})]. \end{aligned} \quad (2.3)$$

A função  $C(\mathbf{h})$  é chamada de covariograma.

Suponha que o processo é isotrópico, ou seja, que a covariância depende apenas da distância  $\mathbf{h}$  entre dois pontos. Dessa forma, combinando a suposição de estacionariedade com a de isotropia, tem-se que o processo é homogêneo. E, dessa forma, pode-se decompor a função de covariância em um produto de dois fatores, sendo um deles a variância e o outro uma função de correlação, i.e.,

$$C(\mathbf{h}) = V\rho(\mathbf{h}), \quad (2.4)$$

sendo  $\rho(\mathbf{h})$  uma função de correlação válida que depende da distância  $\mathbf{h}$  e pode depender de parâmetros desconhecidos.

Seja  $2\gamma(\mathbf{h}) = \text{Var}(\epsilon(\mathbf{s} + \mathbf{h}) - \epsilon(\mathbf{s}))$ . Essa função é chamada de variograma e com a suposição de homogeneidade, tem-se a seguinte forma

$$\begin{aligned} 2\gamma(\mathbf{h}) &= \text{Var}(\epsilon(\mathbf{s} + \mathbf{h})) + \text{Var}(\epsilon(\mathbf{s})) - 2 \text{cov}(\epsilon(\mathbf{s} + \mathbf{h}), \epsilon(\mathbf{s} + \mathbf{h})) \\ &= 2V(1 - \rho(\mathbf{h})). \end{aligned} \quad (2.5)$$

Através do semivariograma, denotado por  $\gamma(\mathbf{h})$ , é possível detectar características importantes da análise espacial da componente  $\epsilon$  tais como

- Alcance: que é a menor distância na qual não há mais correlação espacial para o fenômeno estudado;
- Patamar: é o valor do semivariograma correspondente ao alcance;
- Efeito Pepita: Por definição de semivariograma tem-se que quando o vetor de separação  $\mathbf{h}$  for nulo, então o semivariograma anula-se pois

$$\gamma(0) = 0,5 \times \text{Var}(\epsilon(\mathbf{s}) - \epsilon(\mathbf{s})) = 0,5 \times \text{Var}(0) = 0.$$

Porém, é comum que haja algum erro amostral ou que haja algum efeito aleatório remanescente no processo, fazendo com que o semivariograma não tenda a zero

quando a distância tender a zero. Nesse caso, diz-se ter um efeito pepita que representa a descontinuidade do semivariograma. Quando há esse efeito, entende-se então que a função de covariância possui a seguinte forma:

$$C(\mathbf{h}) = \sigma^2 I(\mathbf{h} = 0) + V\rho(\mathbf{h}), \quad (2.6)$$

A Figura ?? ilustra um exemplo de semivariograma. Note que para toda distância maior ou igual a  $\frac{2}{3}$  o semivariograma é constante e igual a 1,5. Além disso, para distâncias nulas, o semivariograma assume valor 0,5. Logo, o alcance desse exemplo é  $\frac{2}{3}$ , o patamar igual a 1,5, a variância do efeito pepita é 0,5 e a variância  $V$  é 1.

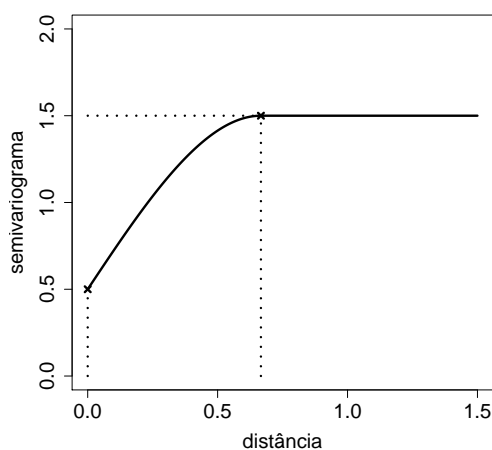


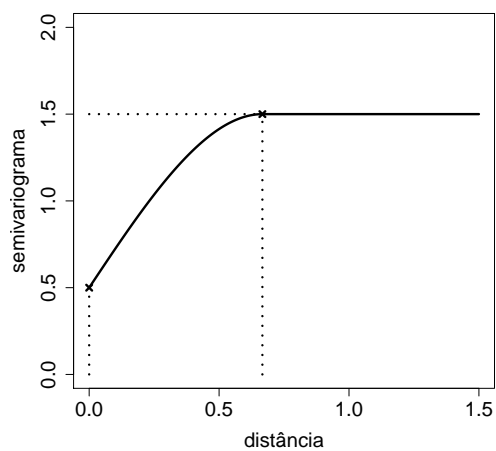
Figura 1: Exemplo de um semivariograma com variância do efeito pepita igual a 0,5, alcance igual  $\frac{2}{3}$  e com o patamar igual a 1,5. Sendo assim, tem-se que a variância  $V$  assume o valor 1.

Existem diferentes funções de covariâncias válidas utilizadas na literatura. A Tabela ?? apresenta alguma dessas funções de covariância e as expressões dos seus respectivos semivariogramas. O parâmetro  $\phi$  do modelo esférico está relacionado com o alcance desse modelo pois a correlação espacial desse modelo é dada por  $\rho(d) = 1 - \frac{2}{3}\phi d + \frac{1}{2}(\phi d)^3$  sendo  $d =$ , a distância entre dois pontos de região em estudo e então o alcance é obtido fazendo  $\rho(d) = 0$  e dessa forma o alcance será uma função de  $\phi$ .

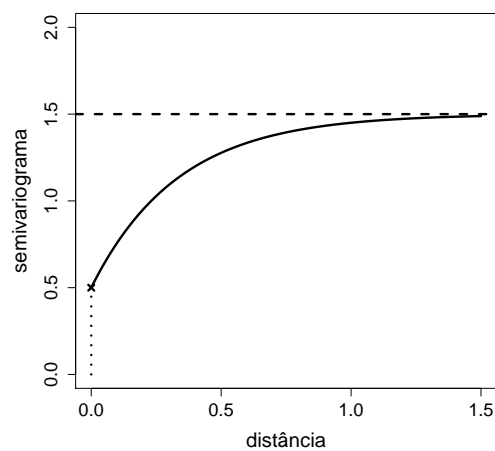
| Modelo      |   |
|-------------|---|
| Esférico    | $C(d) = \begin{cases} 0, & \text{se } d \geq \frac{1}{\phi}, \\ V \left[ 1 - \frac{3}{2}\phi d + \frac{1}{2}(\phi d)^3 \right], & \text{se } 0 < d < \frac{1}{\phi}, \\ \sigma^2 + V, & \text{caso contrário.} \end{cases}$ $\gamma(d) = \begin{cases} \sigma^2 + V, & \text{se } d \geq \frac{1}{\phi}, \\ \sigma^2 + V \left[ \frac{3}{2}\phi d - \frac{1}{2}(\phi d)^3 \right], & \text{se } 0 < d < \frac{1}{\phi}, \\ 0, & \text{caso contrário.} \end{cases}$ |
| Exponencial | $C(d) = \begin{cases} V \exp(-\nu d), & \text{se } d > 0, \\ \sigma^2 + V, & \text{caso contrário.} \end{cases}$ $\gamma(d) = \begin{cases} \sigma^2 + V[1 - \exp(-\nu d)], & \text{se } d > 0, \\ 0, & \text{caso contrário.} \end{cases}$   |
| Gaussiano   | $C(d) = \begin{cases} V \exp(-\nu^2 d^2), & \text{se } d > 0, \\ \sigma^2 + V, & \text{caso contrário.} \end{cases}$ $\gamma(d) = \begin{cases} \sigma^2 + V[1 - \exp(-d^2 \nu^2)], & \text{se } d > 0, \\ 0, & \text{caso contrário.} \end{cases}$   |

Tabela 1: Exemplos de funções de covariâncias válidas e seus respectivos semivariogramas para dados geoestatísticos isotrópicos sendo  $d = \|\mathbf{h}\|$  e  $\sigma^2$  a variância do efeito pepita.

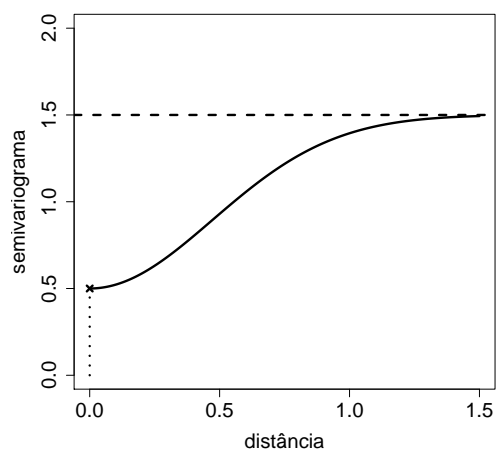
Note que nos modelos exponencial e gaussiano a função de correlação é dada pela função exponencial. Como essa função nunca se anula, então tanto o modelo exponencial quanto gaussiano atingem o patamar apenas assintoticamente. O modelo gaussiano é comumente utilizado para modelar fenômenos contínuos e possui amplitude extensa e patamar parecido ao modelo exponencial. As Figuras ?? e ?? ilustram o comportamento do semivariograma para os diferentes modelos apresentados. Pela Figura ??, é possível notar que o modelo esférico torna-se constante e igual a 1,5 para distâncias maiores ou iguais a  $\frac{2}{3}$ . Já a Figura ?? mostra que a correlação fica nula a partir do ponto  $\frac{2}{3}$ , ou seja o alcance é  $\frac{2}{3}$ . Os demais modelos tem o semivariograma tendendo a 1,5 para distâncias maiores que 1. Como a correlação da função exponencial e gaussiana não se anulam, usa-se valor aproximado para chegar no alcance efetivo, note que o alcance na função exponencial é igual a 1 e o alcance efetivo da função gaussiana é maior. Assim, comparando os alcances, o modelo esférico do exemplo ilustrado tem um alcance menor que os demais modelos.



(a) Esférico



(b) Exponencial



(c) Gaussiano

Figura 2: Exemplos de semivariograma supondo  $V = 1$  (variância do efeito puramente espacial),  $\sigma^2 = 0,5$  (variância do efeito pepita),  $\phi = 1,5$  e  $\nu = 3$ . Note que o patamar no modelo esférico é 1,5 e nos demais tende a 1,5.

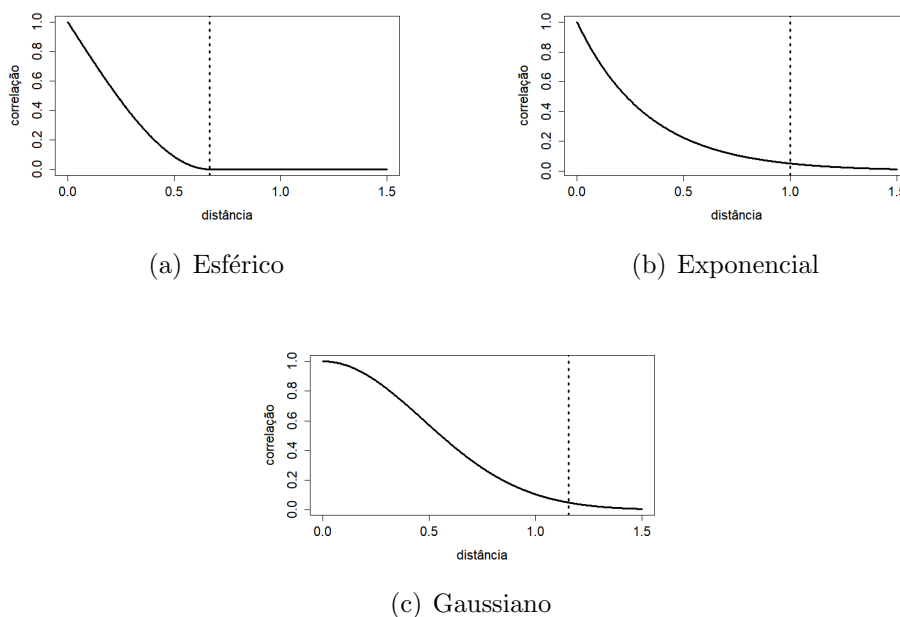


Figura 3: Exemplos de correlação supondo  $\phi = 1, 5$  e  $\nu = 3$ .

## 2.2 Inferência Bayesiana

A inferência estatística consiste em fazer afirmações sobre características desconhecidas da população baseadas em uma amostra coletada dessa população. Há diferentes formas de realizar essa inferência e as mais utilizadas são a clássica e a bayesiana, para mais informação sobre o assunto vide ??) e ??) . Esse trabalho se concentrará na abordagem bayesiana.

Denote o conjunto de parâmetros populacionais desconhecidos por  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)^T$  e a amostra populacional por  $\mathbf{z} = (z_1, z_2, \dots, z_n)^T$ . As afirmações sobre o parâmetro  $\boldsymbol{\theta}$  podem ser realizadas através da distribuição *a posteriori*,  $p(\boldsymbol{\theta}|\mathbf{z})$ , obtida através do Teorema de Bayes, da seguinte forma:

$$p(\boldsymbol{\theta}|\mathbf{z}) = \frac{p(\mathbf{z}, \boldsymbol{\theta})}{p(\mathbf{z})} = \frac{p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{z})}, \quad (2.7)$$

sendo  $p(\boldsymbol{\theta})$  chamada de distribuição *a priori*,  $p(\mathbf{z})$  chamada de distribuição marginal de  $\mathbf{z}$  e  $p(\mathbf{z}|\boldsymbol{\theta})$  a distribuição da variável resposta condicionada no conjunto paramétrico  $\boldsymbol{\theta}$ .

Quando aplica-se os valores amostrados,  $\mathbf{z}$ , em  $p(\mathbf{z}|\boldsymbol{\theta})$ , costuma-se chamar essa função de verossimilhança e denota-se isso por  $l(\boldsymbol{\theta}; \mathbf{z})$ . Note que  $p(\mathbf{z}|\boldsymbol{\theta})$  é uma função da variável aleatória  $\mathbf{Z}$  e por ser uma função de densidade de probabilidade (ou função de distribuição)

precisa atender determinadas características. Quando isso passa a ser denotado por  $l(\boldsymbol{\theta}; \mathbf{z})$  deseja-se ressaltar que agora essa função varia com  $\boldsymbol{\theta}$ .

A grande diferença da inferência clássica para a bayesiana é que a bayesiana permite a inclusão de uma crença probabilística sobre o vetor paramétrico  $\boldsymbol{\theta}$ , anterior a amostragem, através da distribuição  $p(\boldsymbol{\theta})$ . Os clássicos costumam inferir sobre  $\boldsymbol{\theta}$  analisando a função de verossimilhança,  $l(\boldsymbol{\theta}; \mathbf{z})$ , e os bayesianos atualizam a sua crença sobre esse parâmetro após a coleta dos dados analisando a distribuição *a posteriori*,  $p(\boldsymbol{\theta}|\mathbf{z})$ .

Dada a distribuição da variável resposta condicionada no conjunto paramétrico e a distribuição *a priori* do conjunto paramétrico, pode-se obter a distribuição marginal de  $\mathbf{z}$ , através da seguinte integração:

$$\begin{aligned} p(\mathbf{z}) &= \int p(\mathbf{z}, \boldsymbol{\theta}) d\boldsymbol{\theta}, \\ &= \int p(\mathbf{z}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \\ &= E_{\boldsymbol{\theta}} [p(\mathbf{z}|\boldsymbol{\theta})]. \end{aligned} \tag{2.8}$$

A distribuição *a posteriori* de  $\boldsymbol{\theta}$ , dada pela Equação em (??) pode ser reescrita da seguinte forma:

$$p(\boldsymbol{\theta}|\mathbf{z}) = c p(\mathbf{z}|\boldsymbol{\theta}) p(\boldsymbol{\theta}), \tag{2.9}$$

sendo  $c^{-1} = p(\mathbf{z})$ . A constante  $c$  pode ser obtida através das Equações dadas em (??). Além disso, recorrendo a ideia de proporcionalidade, pode-se reescrever a distribuição *a posteriori* colocando as parcelas do produto que não dependem de  $\boldsymbol{\theta}$  na constante de proporcionalidade. Por propriedades de probabilidade essa constante é facilmente obtida integrando-se a distribuição *a posteriori* de  $\boldsymbol{\theta}$  com respeito a  $\boldsymbol{\theta}$  e igualando-se a 1. Sendo assim, pode-se referir a distribuição *a posteriori* de  $\boldsymbol{\theta}$  da seguinte forma:

$$p(\boldsymbol{\theta}|\mathbf{z}) \propto p(\mathbf{z}|\boldsymbol{\theta}) p(\boldsymbol{\theta}). \tag{2.10}$$

Pode-se atribuir uma distribuição *a priori* informativa ou uma não informativa. A distribuição *a priori* não informativa é uma distribuição vaga, que traz pouca ou nenhuma informação probabilística sobre o parâmetro. O pesquisador pode definir essa distribuição recorrendo a um especialista. Porém, muitas vezes a crença inicial sobre o parâmetro resume-se a algumas características e não necessariamente a formulação de uma distribuição. Por exemplo, pode-se haver uma crença apenas no valor médio desse

parâmetro e dessa forma diferentes distribuições podem ser atribuídas atendendo a essa crença.

Outra possível forma de atribuir uma distribuição *a priori* é recorrendo a conjugação, ou seja, escolhe-se uma família de distribuições *a priori* que, após combinada com a função de verossimilhança, implica em uma distribuição *a posteriori* pertencente a mesma família da distribuição *a priori*. Para ilustrar isso, suponha que  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$  tem a seguinte distribuição:

$$Z \sim N(\theta, V), \quad (2.11)$$

com  $V$  conhecido e  $\theta$  desconhecido. A priori, considere que  $\theta \sim N(\mu_0, \tau_0^2)$ . Neste caso a distribuição *a posteriori* de  $\theta$  dado a amostra  $\mathbf{z}$  segue uma distribuição normal com média  $\mu_1$  e variância  $\tau_1^2$ , sendo:

$$\mu_1 = \frac{\tau_0^{-2}\mu_0 + n\sigma^{-2} + z}{\tau_0^{-2} + n\sigma^{-2}} \text{ e } \tau_1^{-2} = \tau_0^{-2} + n\sigma^{-2} \quad (2.12)$$

As contas para chegar nessa distribuição *a posteriori* estão disponível no Apêndice ???. Pode-se atribuir uma distribuição *a priori* informativa tornando a variância  $\tau_0^2$  pequena e quanto maior for essa, menos informativa será a distribuição.

Mesmo utilizando uma distribuição *a priori* conjugada com variância grande, ainda assim pode haver alguma informação sobre o parâmetro. Por exemplo, mesmo supondo que  $\theta \sim N(0, 1000)$ , a probabilidade de  $\theta$  assumir valores maiores que 1000 é praticamente nula. Sendo assim, existem outras formas de atribuir uma distribuição *a priori* não informativa. As mais utilizadas são as de referência e, em particular, a de Jeffreys.

Bayes propôs a seguinte ideia inicial para distribuição *a priori* não informativa: Considerando  $\theta$  um escalar, supor que todos os valores de  $\theta$  são igualmente prováveis, e portanto  $p(\theta) \propto k$  para  $\theta$  variando em subconjunto de reta, ou seja, nenhum valor particular tem preferência.

Um problema dessa formulação é que ela é muitas vezes imprópria, ou seja, a integral de  $p(\theta)$  com respeito a  $\theta$  difere de 1, não respeitando as propriedades básicas de probabilidade. Quando isso ocorre, é necessário a demonstração que a distribuição *a posteriori* será própria.

Sobre a distribuição *a priori* de referência há o desejo de se fazer inferência com o mínimo de influência da informação *a priori*, ou seja, a informação da amostra será dominante. O método de Jeffreys é o método mais utilizado nesse caso, pois esse método

é invariante a transformação paramétricas 1 a 1, assegurando que a inferência não dependa da escolha da parametrização do modelo.

Matematicamente, podemos descrever a regra geral de Jeffrey para o vetor paramétrico  $\boldsymbol{\theta}$  dessa forma:

$$p(\boldsymbol{\theta}) \propto [\det I(\boldsymbol{\theta})]^{\frac{1}{2}} \quad (2.13)$$

onde  $I(\boldsymbol{\theta})$  é a matriz de informação esperada de Fisher para o vetor paramétrico  $\boldsymbol{\theta}$ .

Essa matriz pode ser obtida da seguinte forma:

$$I(\boldsymbol{\theta}) = E \left[ -\frac{\partial^2 \log p(z|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \quad (2.14)$$

Note que essa matriz depende da distribuição dos dados condicionados ao vetor paramétrico  $\boldsymbol{\theta}$  e portanto viola a ideia inicial de que a distribuição a priori é uma crença do parâmetro que independe da distribuição dos dados.

Voltando ao exemplo da equação ??, para o método de Jeffrey, a  $p(\boldsymbol{\theta}) \propto k$ , sendo  $k$  uma constante, as contas estão no apêndice ?. Sendo assim, a figura ?? compara o comportamento das distribuições *a priori* com a variância pequena e grande, e quando a distribuição é uma constante. Nota-se que quando a variância é pequena a distribuição *a priori* fica muito mais informativa, e quando a variância é grande ou a distribuição uma constante, ela não fornece muito informação, assim se caracteriza uma distribuição *a priori* não informativa.



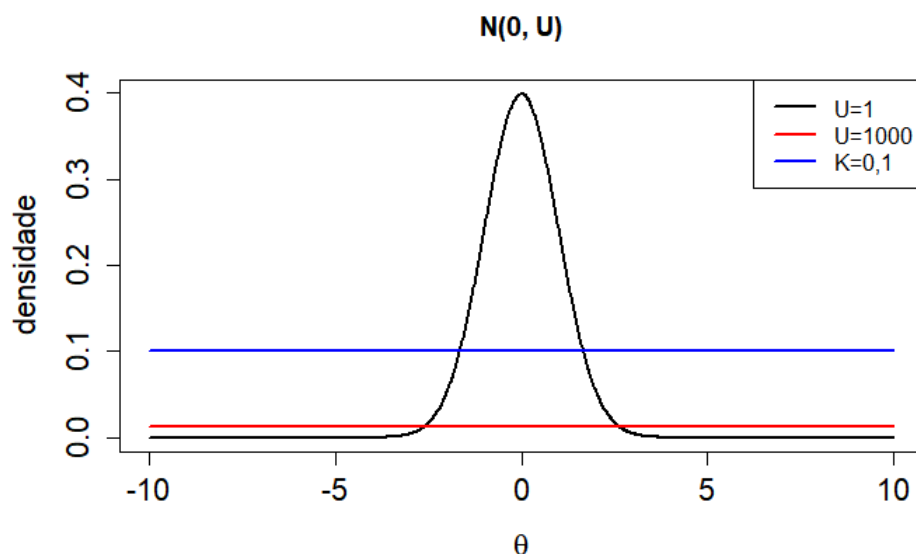


Figura 4: Comportamento das distribuições *a priori* informativa, não informativa e referência.

### 2.2.1 Métodos de Monte Carlos via Cadeia de Markov

Em particular, na inferência bayesiana, é comum obter uma distribuição *a posteriori* com forma analítica desconhecida ou custosa de ser amostrada. Nesses casos, pode-se recorrer aos métodos de **MCMC!** para obter uma amostra da distribuição *a posteriori* e inferir sobre o vetor paramétrico  $\theta$ . Mais detalhes sobre esse método podem ser vistos em ??) e ??).

Para explicar esse método na inferência Bayesiana supinha que o interesse esteja em amostrar da distribuição posteriori  $p(\theta|z)$  e que não seja possível amostrar diretamente uma distribuição. A cadeia de Markov de primeira ordem é um processo estocástico onde a probabilidade de ocorrência no futuro dado o presente estado, independe do passado. Ou seja,

$$P(\Theta_{j+1} = \theta_{j+1} \mid \Theta_j = \theta_j, \Theta_{j-1} = \theta_{j-1}, \dots, \Theta_0 = \theta_0) = \quad (2.15)$$

$$P(\Theta_{j+1} = \theta_{j+1} \mid \Theta_j = \theta_j) \quad (2.16)$$

para todo  $\theta_0, \dots, \theta_{j+1} \in S$ , sendo  $S$  um conjunto de possíveis valores para  $\theta$ .

Os métodos de **MCMC!** exigem que as cadeias sejam homogênea, ergódica e irreduzível. Dizemos que a cadeia é homogênea quando a probabilidade de transição for a mesma em cada passo. Ergódica quando ela é aperiódica e recorrente positiva. A cadeia é recorrente positiva quando o número médio de passos de retornar a qualquer

estado é finito e aperiódica quando não atinge o mesmo ponto com uma regularidade. É irreduzível quando a probabilidade for positiva do estado ser atingido em número finito de iterações.

### Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings foi proposto por (Metropolis e Hastings) e é utilizado quando a distribuição de interesse não possui forma analítica conhecida. Em inferência bayesiana, a distribuição de interesse pode ser a distribuição *a posteriori* e, portanto, a apresentação desse algoritmo será feita com base nessa distribuição de interesse.

A ideia principal desse algoritmo, é gerar valores dada uma distribuição conhecida  $q(\cdot|\theta)$ , nomeada de distribuição proposta, que pode ser escolhida arbitrariamente. Os valores que são gerados por essa distribuição são representados como  $\theta^p$ , e nomeados como valores propostos. O valor gerado  $\theta^p$  é aceito na cadeia com a probabilidade

$$\alpha(\theta, \theta^p) = \min \left( 1, \frac{p(\mathbf{z}|\theta^p)p(\theta^p)q(\theta|\theta^p)}{p(\mathbf{z}|\theta)p(\theta)q(\theta^p|\theta)} \right),$$

Resumidamente, a cadeia de Markov é gerada, e a cada interação um candidato para o próximo estado é gerado,  $\theta^p$ , a partir de uma distribuição proposta  $q(\cdot|\theta)$ . Esse valor dependerá do valor atual da cadeia e será aceito e adicionado à cadeia com a probabilidade de aceitação  $\alpha(\theta, \theta^p)$ .

Portanto, o algoritmo de Metropolis-Hastings possui os seguintes passos:

1. Inicialize  $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})$  e defina o contador de iterações  $j = 1$
2. Obtenha um novo valor  $\theta^p$  da distribuição conhecida  $q(\theta^p|\theta^{(j-1)})$ , chamada de distribuição proposta.
3. Calcule a probabilidade de aceitação  $\alpha(\theta^{(j-1)}, \theta^p)$  e gere  $u \sim U(0, 1)$
4. Caso  $u \leq \alpha$  aceite o valor e faça  $\theta^{(j)} = \theta^p$ , se não faça  $\theta^{(j)} = \theta^{(j-1)}$
5. Atualize o contador de iterações fazendo  $j = j + 1$  e volte ao passo 2 até obter convergência.

A escolha da distribuição proposta é crucial para o algoritmo, a escolha errada de uma distribuição pode ocasionar um número de rejeições muito alto, prejudicando a eficiência do algoritmo.

Existem alguns casos especiais desse algoritmo, quando a distribuição proposta não depende do estado atual da cadeia  $q(\theta^p|\theta) = q(\theta^p)$ , assim  $\theta^p$  é aceito com a probabilidade

$$\alpha(\theta, \theta^p) = \min \left( 1, \frac{p(\mathbf{z}|\theta^p)p(\theta^p)q(\theta)}{p(\mathbf{z}|\theta)p(\theta)q(\theta^p)} \right).$$

Apesar que os valores  $\theta^p$  gerados de forma independente, a probabilidade de aceitação ainda depende de  $\theta$ .

Outra caso especial desse algoritmo ocorre quando considera-se a distribuição proposta simétrica da seguinte forma:  $q(\theta^p|\theta) = q(\theta|\theta^p)$  para todos os valores de  $\theta$  e  $\theta^p$ . Assim a probabilidade de aceitação se simplifica para

$$\alpha(\theta, \theta^p) = \min \left( 1, \frac{p(\mathbf{z}|\theta^p)p(\theta^p)}{p(\mathbf{z}|\theta)p(\theta)} \right).$$

A seguir vamos estudar o Amostrador de Gibbs, que é um outro método de **MCMC!**.

### Amostrador de Gibbs

O algoritmo amostrador de Gibbs é um método de **MCMC!** proposto por ??) se tornando conhecido entre os Estatísticos por ??). Seja  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^T$  um vetor com distribuição  $p(\boldsymbol{\theta}|\mathbf{z})$ . Dado que  $\theta_{-l}$  é formado por todos os  $\theta$  exceto pelo elemento  $\theta_l$ , sendo  $l = 1, \dots, d$ . As distribuições  $p_l(\theta_l) = p(\theta_l|\theta_{-l})$ ,  $l = 1, \dots, d$ , são as distribuições condicionais completas. A utilização do método é comum quando conhecemos as distribuição condicional completa do parâmetro. Assim o Amostrador Gibbs fornece um método alternativo de criação baseado nas gerações sucessivas. Podendo ser escrito da seguinte maneira:

1. Defina um valor inicial  $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})$ .
2. Inicialize o contador de iteração  $j = 1$
3. Obteha um novo valor  $\boldsymbol{\theta}^{(j)} = (\theta_1^{(j)}, \dots, \theta_d^{(j)})^T$  através das sucessivas gerações das distribuições condicionais completas a posteriori

$$\begin{aligned} \theta_1^{(j)} &\sim p(\theta_1|\theta_2^{(j-1)}, \dots, \theta_d^{(j-1)}, \mathbf{z}) \\ \theta_2^{(j)} &\sim p(\theta_2|\theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_d^{(j-1)}, \mathbf{z}) \\ &\vdots \\ \theta_d^{(j)} &\sim p(\theta_d|\theta_1^{(j)}, \dots, \theta_{d-1}^{(j)}, \mathbf{z}) \end{aligned}$$

4. Incremente o contador  $j$  para  $j + 1$  e retorne ao passo 2 até a convergência ser alcançada.

O algoritmo gera uma cadeia de Markov, já que  $\boldsymbol{\theta}^{(j+1)}$  depende somente dos valores obtidos na interação  $j$ . Assim, a amostra gerada através do Amostrador de Gibbs não é independente, cada valor depende do seu antecessor. A forma mais comum de se obter uma amostra aproximadamente independente de tamanho  $c$ , é replicar  $c$  vezes, em cada uma delas fazendo  $m$  iterações. Esta simulação gera  $c \times m$  cadeias. Se os valores estabelecidos para os estados iniciais forem da cadeia  $\boldsymbol{\theta}_t^{(0)}$ ,  $t = 1, \dots, d$ , de forma independente, amostra consistirá em valores independentes de  $p(\boldsymbol{\theta}|\mathbf{z})$ .

Outra forma é considerar apenas uma única replicação do algoritmo, basendo-se na hipótese que após a convergência os valores simulados seguem a distribuição estacionária. Assim, os primeiros valores gerados são considerados como um período de aquecimento do algoritmo (*burn-in*), os valores são descartados, e se utilizam apenas os valores posteriores ao aquecimento como uma amostra da distribuição de interesse.

Normalmente a definição do tamanho de período de aquecimento não pode ser realizado analiticamente. Porém, existem métodos formais e informais para a validação da convergência da cadeia. O método informal mais comum se baseia em gráficos da trajetória da cadeia, para verificar que a partir de qual interação os dados passam a ter o mesmo comportamento. Os métodos formais, usam uma variedade de resultados teóricos e aproximações. ??) apresenta uma revisão dos principais métodos.

Existem algoritmos para verificar convergência, como o CODA ??). Este programa está disponível no software ??).

## 3 Análise dos Resultados

Neste capítulo, dados artificiais são gerados e o modelo espacial descrito em ?? é aplicado nos dados simulados. A inferência bayesiana descrita em ?? é utilizada para estimar os parâmetros desconhecidos. Como os dados são simulados, os valores verdadeiros dos parâmetros são conhecidos e é possível verificar se a inferência adotada recupera esses valores verdadeiros. É feita uma análise de sensibilidade, ou seja, um estudo sobre a influência da distribuição *a priori* na distribuição *a posteriori*. A análise dos dados simulados encontra-se na Seção ?. Em seguida, o modelo é aplicado em um conjunto de dados reais e a análise encontra-se na Seção ?.

### 3.1 Estudo Simulado

Suponha que se deseja modelar um fenômeno contínuo  $Z$  dentro de uma localização  $\mathbf{s}$ . Suponha que haja um intercepto que representa a média comum do processo a todas as localizações e que haja uma covariável para cada localização, isto é,  $\mathbf{X}(\mathbf{s}) = (1, X_1(\mathbf{s}))$ . Sendo assim, utilizando o modelo proposto na Seção ?, tem-se que

$$Z(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + e(\mathbf{s}) + \omega(\mathbf{s}),$$

sendo  $\boldsymbol{\beta} = (\beta_0, \beta_1)$  o vetor de efeitos das covariáveis  $\mathbf{X}(\mathbf{s})$ ,  $e(\mathbf{s})$  um efeito puramente espacial e  $\omega(\mathbf{s})$  um ruído aleatório não correlacionado.

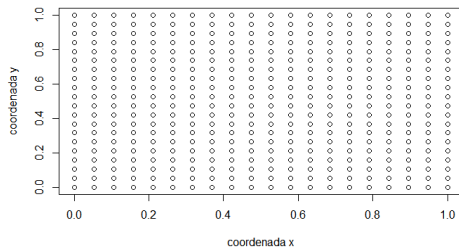
Suponha que o efeito  $e(\mathbf{s})$  tem distribuição normal com média zero e variância  $V$  para todo  $\mathbf{s}$  e que a covariância entre  $e(\mathbf{s})$  e  $e(\mathbf{s} + \mathbf{h})$  seja  $V \times \rho(\mathbf{h})$ , onde  $\rho(\mathbf{h})$  é a função de correlação espacial que dependerá da distância entre as localizações. Usando a distância euclidiana, tem-se que essa distância  $d$  é obtida da seguinte forma:  $d = \sqrt{(h_1^2 + h_2^2)}$  e portanto  $\rho(d) = \exp\{-d\nu\}$ , onde  $\nu > 0$  é um parâmetro a ser estimado. Assim, o vetor paramétrico é formado por  $\theta = (\beta_0, \beta_1, V, \nu)$ .

Considere que  $X_1(\mathbf{s}) \sim N(0, 1)$  e que as localizações forma uma grade regular com 400 localizações. Fixou-se os valores dos parâmetros da seguinte forma  $V = 1$ ,  $\nu = \frac{\max(d)}{4} = 2,83$ ,  $\beta_0 = 3$  e  $\beta_1 = 0,5$ .

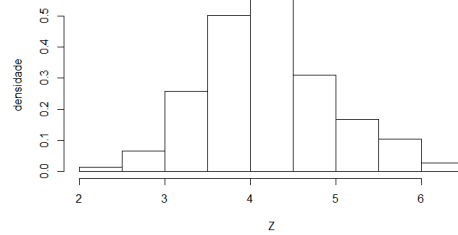
Supondo que os valores dos parâmetros são desconhecidos e deseja-se inferi-los então foram propostas as seguintes distribuições *a priori*:  $V \sim GI(2, 1)$ , o apêndice ?? fala mais sobre essa distribuição,  $(\beta_0, \beta_1)^T \sim N(\mathbf{0}, 1000I_2)$ , onde  $\mathbf{0}$  é um vetor bi-dimensional com todos os elementos iguais a 0 e  $I_2$  é a matriz identidade de ordem 2 e  $\nu \sim G(2, 1)$ . Como visto na Figura ?? , quando a distribuição tem a variância muito grande, ela dá pouca informação sobre o parâmetro, assim a distribuição *a priori* do parâmetro  $\beta$  é pouco informativa. A equação ?? mostra que, quando o parâmetro  $a = 2$ , a  $Var(V) \rightarrow \infty$ , assim teremos uma distribuição *a priori* não informativa. Para o parâmetro  $\nu$  a sua distribuição *a priori* é informativa, ou seja, os valores de  $\nu$  se concentram em uma determinada região. Desta forma, forma-se um conjunto de distribuição *a priori* quase não informativa.

Tabela 2: Estatísticas descritivas dos dados simulados.

|           | Z    |
|-----------|------|
| Média     | 3,04 |
| Variância | 1,77 |
| Mediana   | 3,1  |



(a) Localizações simuladas



(b) Histograma dos dados simulados

Figura 5: Localizações simuladas e o histograma dos dados simulados.

Foram realizadas 200000 interações. Utilizou-se um aquecimento de 1000 interações, isso significa, que foram desconsiderados os 1000 primeiros valores e um espaçamento de 200, retirou-se para compor apenas um a cada 200 iterações da cadeia. O tamanho da amostra final é de 996.

A figura ?? apresenta os gráficos de autocorrelação dos valores estimados para os parâmetros, os valores da autocorrelação decresceram rapidamente para zero, assim não há indício de influências significativas entre os espaçamentos. Assim, parece haver independência entre as iterações.

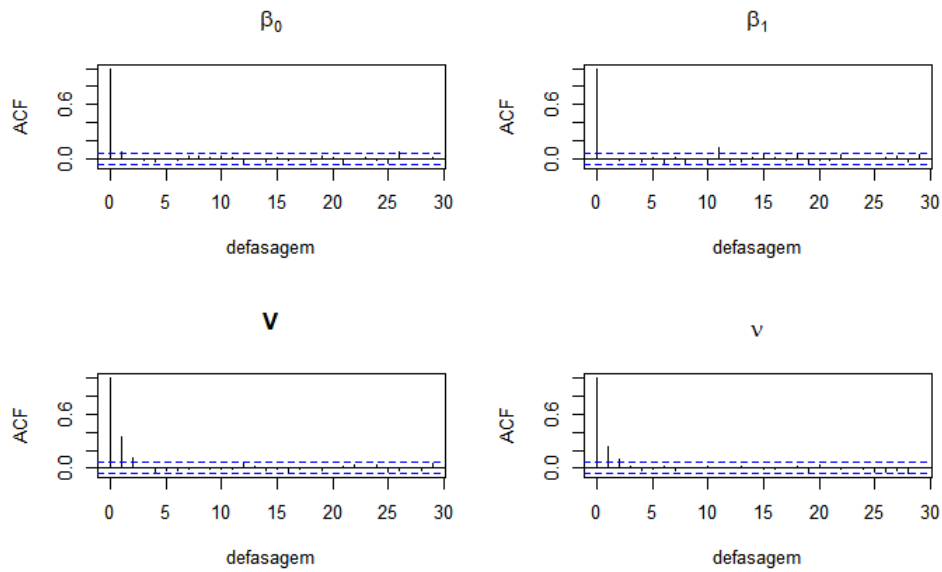


Figura 6: Gráfico de autocorrelação da amostra a posteriori dos parâmetros do modelo espacial.

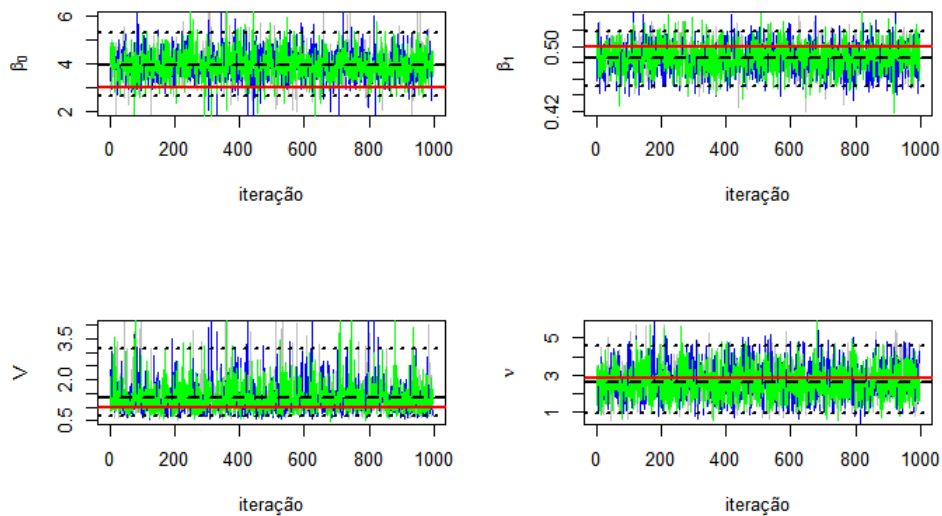


Figura 7: Traços das cadeias da amostra da *posteriori* de  $\beta_0, \beta_1, V$  e  $\nu$ , a linha tracejada representa a média *a posteriori*, as pontilhadas o intervalo de credibilidade de 95% e a linha vermelha cheia os valores verdadeiros.

A figura ?? apresenta os traços das amostras geradas para cada um dos parâmetros do modelo, foram geradas três cadeias obtidas ao se inicializar o amostrador em diferentes pontos. Os pontos iniciais das amostras foram 1, 100 e  $\frac{1}{1000}$ . É possível notar que todas as cadeias simuladas convergem para o mesmo lugar. A linha vermelha e cheia representa

o valor teórico, a linha tracejada a média *a posteriori* e as linhas pontilhadas o intervalo de credibilidade de 95%.

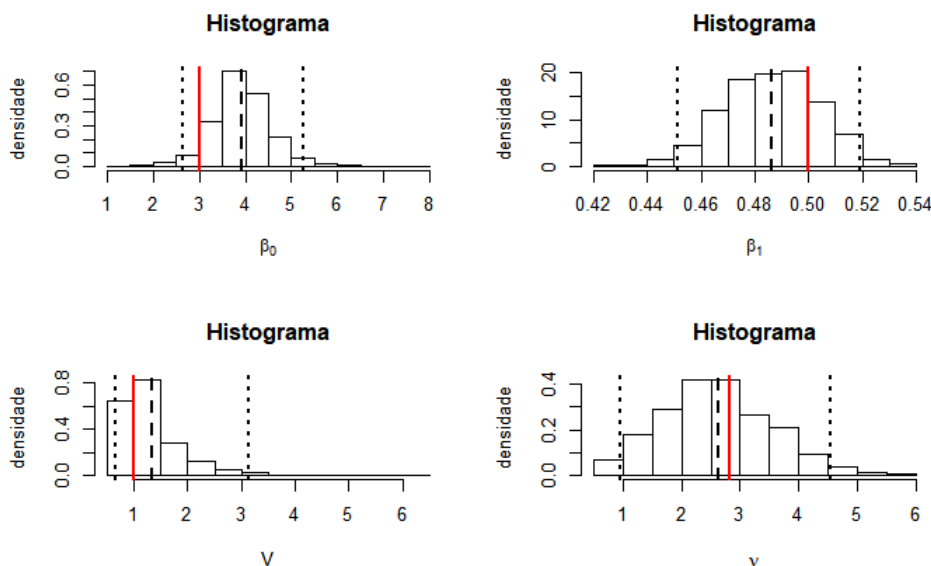


Figura 8: Histograma da distribuição *a posteriori* de  $\beta_0, \beta_1, V$  e  $\nu$ , a linha tracejada representa a média *a posteriori*, as pontilhadas o intervalo de credibilidade de 95% e a linha vermelha cheia os valores verdadeiros.

Na Figura ?? apresenta-se a distribuição *a posteriori* simulada dos parâmetros. A linha vermelha representa os valores verdadeiros dos parâmetros a tracejada a média *a posteriori* e as linhas pontilhadas o intervalo de credibilidade de 95%.

Tabela 3: Estimativa das cadeias

| Parâmetro | valor teórico | média | IC          |
|-----------|---------------|-------|-------------|
| $\beta_0$ | 3             | 3,92  | (2,65;5,26) |
| $\beta_1$ | 0,5           | 0,49  | (0,45;0,52) |
| $\nu$     | 2,63          | 2,62  | (0,96;4,54) |
| V         | 1             | 1,35  | (0,65;3,14) |

A tabela ?? apresenta os valores verdadeiros, as médias da distribuição *a posteriori* e os intervalos de credibilidade da cadeia, assim, pode-se observar que os valores verdadeiros dos parâmetros estão dentro do intervalo de credibilidade. Com isso, podemos concluir com essas evidências que este modelo apresenta boas estimativas para os parâmetros.

### Análise de sensibilidade

A distribuição *a priori* pode impactar na análise final dos resultados. Nesse presente trabalho vamos avaliar a sensibilidade dos hiperparâmetros da distribuição *a priori* nos



resultados finais. Iremos fazer 3 análises, a primeira análise com um conjunto quase não informativo, a segunda com o conjunto não informativo e a terceira com conjunto parcialmente informativo da distribuição *a priori*

Para a primeira simulação da análise de sensibilidade iremos alterar os parâmetros da distribuição *a priori* de  $V$ , com isso as distribuições seguem dessa forma:  $V \sim GI(1/10, 1/10)$ ,  $\beta \sim N(\mathbf{0}, 1000I_k)$ ,  $I$ , onde  $\mathbf{0}$  é um vetor  $K$  dimensional com todos os elementos iguais a 0 e  $I_K$  é a matriz identidade de ordem  $K = 2$  e  $\nu \sim G(2, 1)$ . A Figura ?? mostra que com a mudança dos hiperparâmetros de  $V$ , a distribuição *a priori* do parâmetro fica com menos informação. Como as distribuições *a priori* dos parâmetros  $\beta$  e  $V$  continuam não informativas e a de  $\nu$  informativa. Forma-se assim um conjunto de distribuição *a priori* quase não informativa.

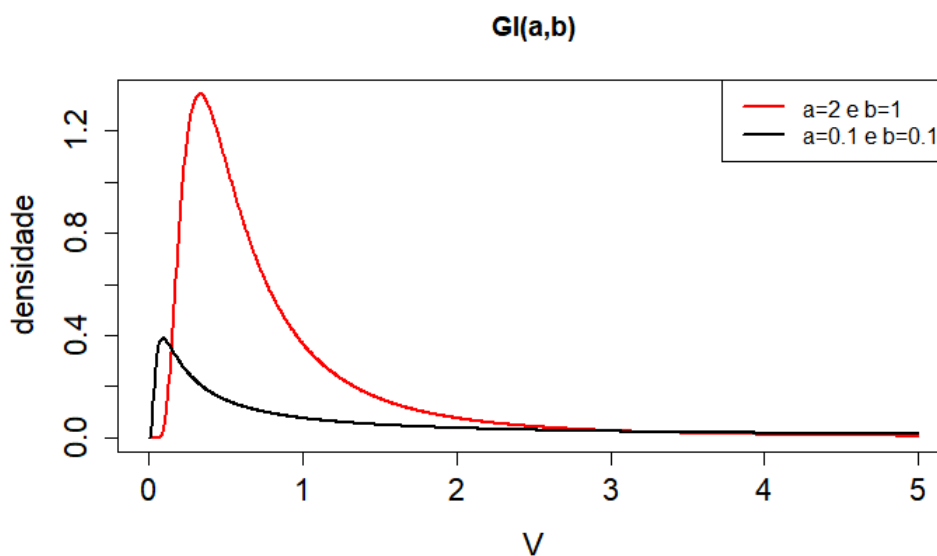


Figura 9: Comportamento das distribuições *a priori* do parâmetro  $V$ .

Para o estudo da sensibilidade do modelo, foram realizadas 200000 iterações para cada análise. Utilizou-se um aquecimento de 1000 interações, isso significa, que foram desconsiderados os 1000 primeiros valores e um espaçamento de 200, retirou-se para compor apenas um a cada 200 iterações da cadeia. O tamanho da amostra final é de 996.

A figura ?? apresenta os gráficos de autocorrelação dos valores estimados para os parâmetros, os valores decresceram rapidamente para 0, dessa forma não há indício de influências significativas entre os espaçamentos. Assim, parece haver independência entre as iterações.

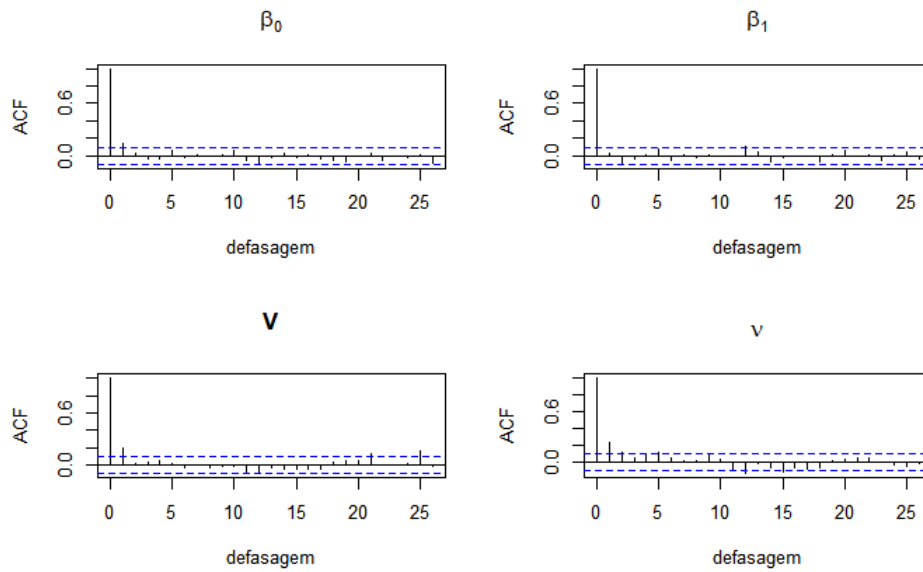


Figura 10: Gráfico de autocorrelação da amostra a posteriori dos parâmetros do modelo espacial.

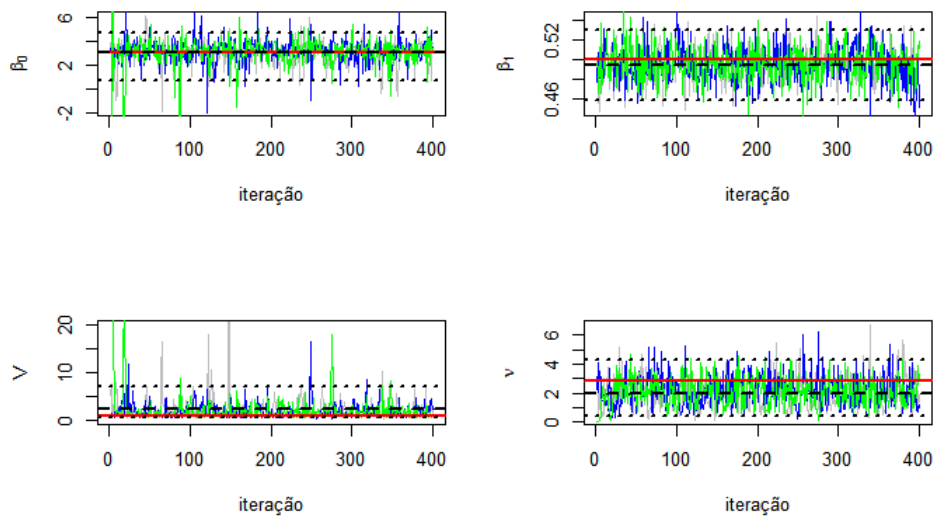


Figura 11: Traços das cadeias da amostra da *posteriori* de  $\beta_0, \beta_1, V$  e  $\nu$ , a linha vermelha representa a média *a posteriori*, as pontilhadas o intervalo de credibilidade de 95% e a linha vermelha cheia os valores verdadeiros.

Foram gerados 3 cadeias, se inicializando de pontos diferentes, a figura ?? mostra os traços das amostras, é possível notar que todas as amostras geradas para distribuições *a posteriori* convergem para a mesma região.

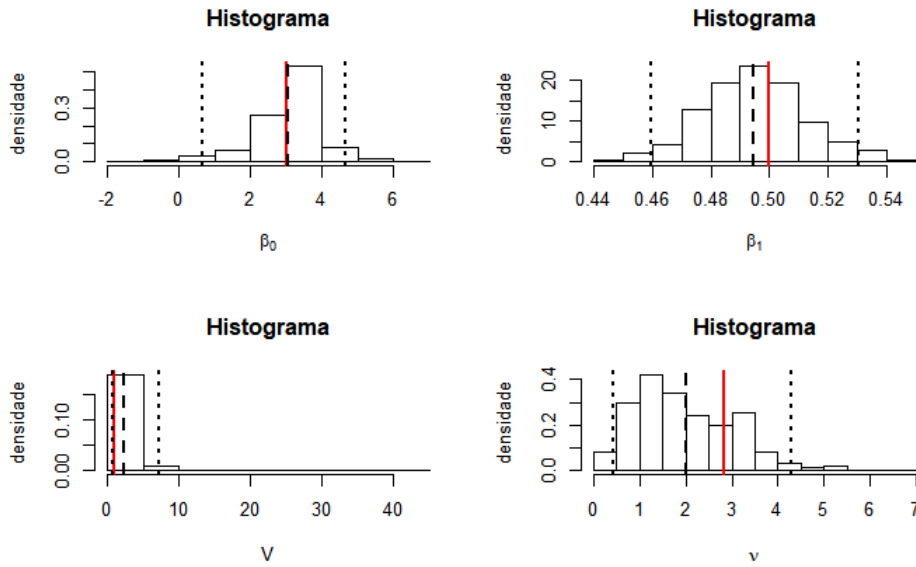


Figura 12: Histograma da distribuição *a posteriori* de  $\beta_0, \beta_1, V$  e  $\nu$ , a linha vermelha representa a média *a posteriori*, as pontilhadas o intervalo de credibilidade de 95% e a linha vermelha cheia os valores verdadeiros.

A figura ?? mostra o comportamento das amostras geradas para as distribuições *a posteriori*, a linha vermelha representa a média *a posteriori*, as pontilhadas o intervalo de credibilidade de 95% e a linha vermelha cheia os valores verdadeiros.

Tabela 4: Estimativa das cadeias do conjunto de distribuições *a priori* quase não informativo

| Parâmetro | valor teórico | média | IC          |
|-----------|---------------|-------|-------------|
| $\beta_0$ | 3             | 3,86  | (1,52;7,29) |
| $\beta_1$ | 0,5           | 0,48  | (0,46;0,52) |
| $\nu$     | 2,63          | 1,78  | (0,28;3,97) |
| $V$       | 1             | 2,66  | (0,81;10,7) |

Pode-se notar, que todos parâmetros pertencem ao intervalo de credibilidade de 95%. Então, mesmo com as mudança na distribuição *a priori* de  $V$  há evidência que este modelo apresenta boas estimativas para os parâmetros.

A segunda análise de sensibilidade contém as distribuições *a priori* seguindo as distribuições  $V \sim GI(1/10, 1/10)$ ,  $\beta \sim N(\mathbf{0}, 1000I_k)$ ,  $I$ , onde  $\mathbf{0}$  é um vetor  $K$  dimensional com todos os elementos iguais a 0 e  $I_K$  é a matriz identidade de ordem  $K = 2$  e  $\nu \sim G(1/10, 1/10)$ , note-se que a alteração foi na distribuição *a priori* de  $\nu$ , a Figura ?? mostra que o parâmetro  $\nu$  perde informação com as alterações dos hiperparâmetros, assim temos que ressaltar que esse conjunto de distribuições *a priori* se caracteriza por ser não informativo.

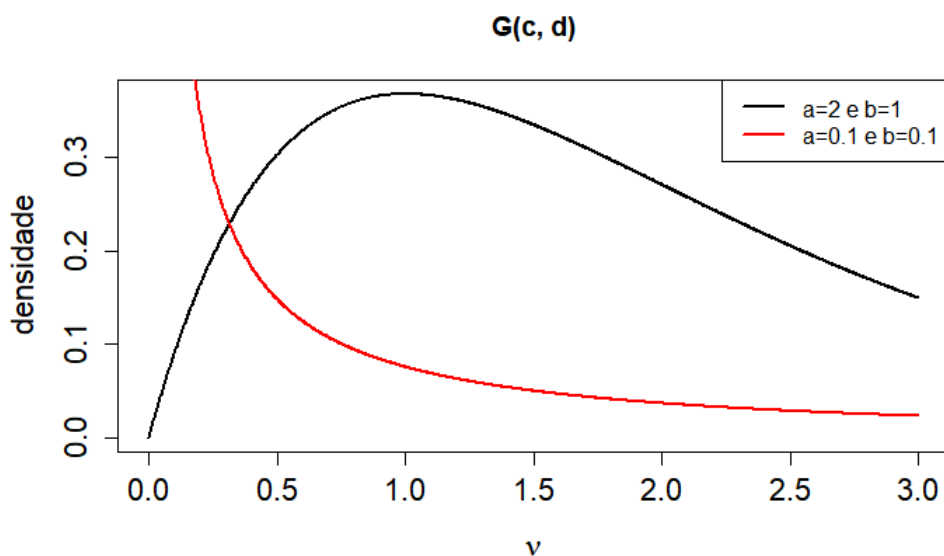


Figura 13: Comportamento das distribuições *a priori* do parâmetro  $\nu$ .

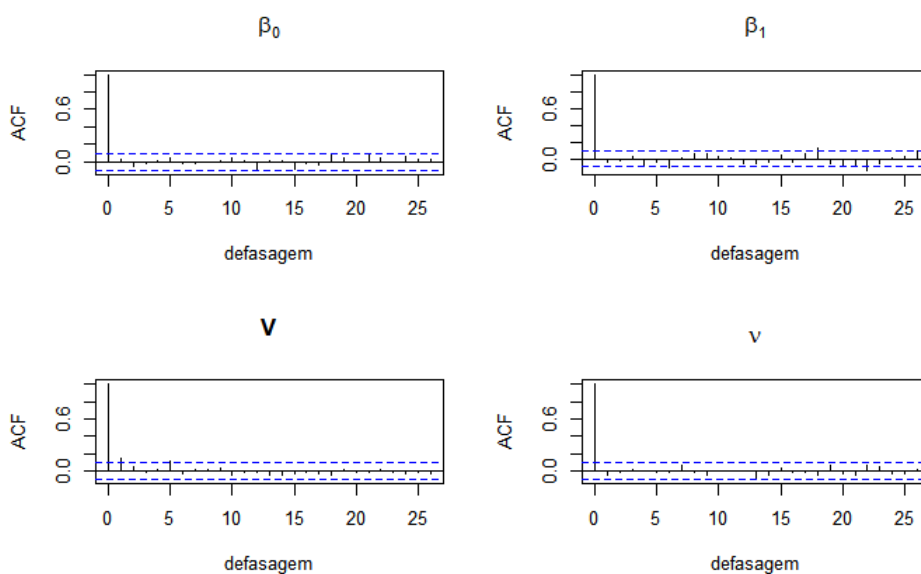


Figura 14: Gráfico de autocorrelação da amostra a posteriori dos parâmetros do modelo espacial.

A Figura ?? mostra que mesmo com a alteração dos parâmetros da distribuição *a priori* de  $\nu$  as cadeias continuam independente entre os espaçamentos.

Foram gerados 3 cadeias, se inicializando de pontos diferentes, a figura ?? mostra os traços das amostras, é possível notar que todas convergem para a mesma região. A linha vermelha representa o valor verdadeiro de cada parâmetro, a pontilha a média *a posteriori* e as linhas tracejadas o intervalo de credibilidade de 95%. Os gráficos mostram que as

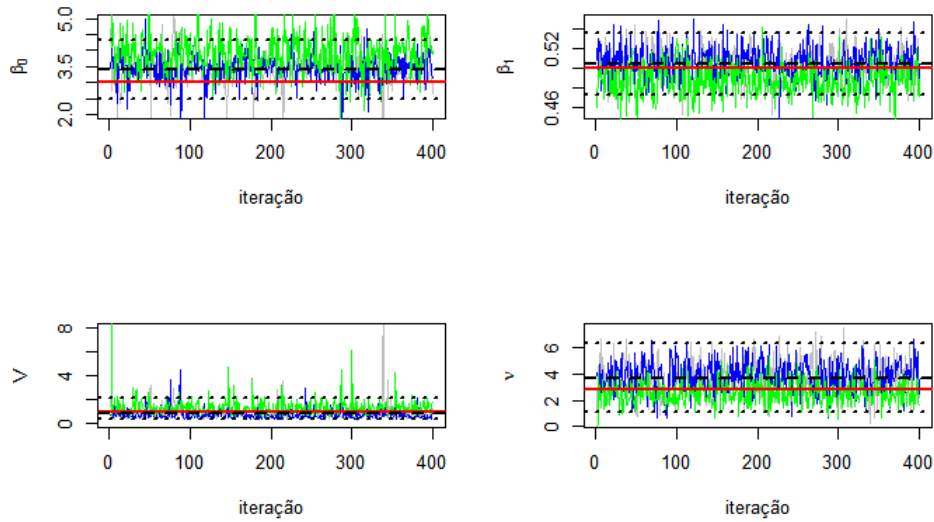


Figura 15: Traços das cadeias da amostra da *posteriori* de  $\beta_0, \beta_1, V$  e  $\nu$ , a linha vermelha representa a média *a posteriori*, as pontilhadas o intervalo de credibilidade de 95% e a linha vermelha cheia os valores verdadeiros.

cadeias continuaram com o mesmo comportamento, os intervalos de credibilidades ainda contém os valores verdadeiros.

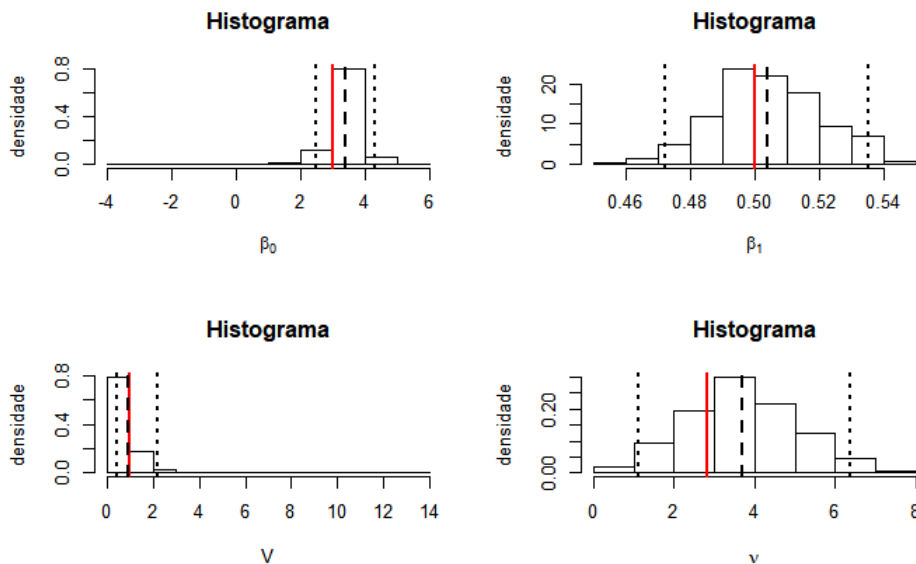


Figura 16: Histograma da distribuição *a posteriori* de  $\beta_0, \beta_1, V$  e  $\nu$ , a linha vermelha representa a média *a posteriori*, as pontilhadas o intervalo de credibilidade de 95% e a linha vermelha cheia os valores verdadeiros.

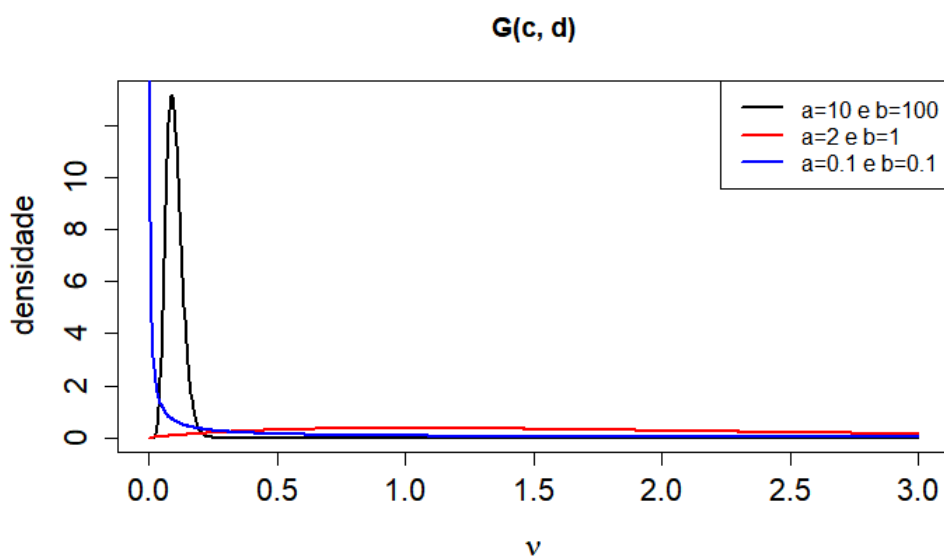
A figura ?? mostra o comportamento das amostras geradas para as distribuições *a posteriori* dos parâmetros.

Tabela 5: Estimativa das cadeias do conjunto de distribuições *a priori* não informativo

| Parâmetro | valor teórico | média | IC          |
|-----------|---------------|-------|-------------|
| $\beta_0$ | 3             | 3,41  | (2,47;4,31) |
| $\beta_1$ | 0,5           | 0,5   | (0,47;0,54) |
| $\nu$     | 2,63          | 3,70  | (1,12;6,37) |
| V         | 1             | 0,88  | (0,43;2,18) |

Nota-se que após as simulações com a distribuição *a priori* não informativa o valor verdadeiro está contido dentro do intervalo de credibilidade de 95%. Com isso, podemos concluir que para conjunto de distribuição *a priori* não informativa o modelo apresenta boas estimativas para os parâmetros.

Para terminar a análise de sensibilidade,  $\nu \sim G(100, 10)$  uma distribuição *a priori* informativa, como a  $E[\nu] = 10$  e  $VAR[\nu] = 1$ , a probabilidade de  $\nu$  estar próximo ao valor verdadeira é baixa, como visto na Figura ???. Logo, as distribuições *a priori* seguem dessa forma:  $V \sim GI(1/10, 1/10)$ ,  $(\beta_0, \beta_1)^T \sim N(\mathbf{0}, 1000I_2)$ , onde  $\mathbf{0}$  é um vetor bi-dimensional com todos os elementos iguais a 0 e  $I_2$  é a matriz identidade de ordem 2 e  $\nu \sim G(100, 10)$ . Assim, o conjunto de distribuição *a priori* é parcialmente conhecida.

Figura 17: Comportamento das distribuições *a priori* do parâmetro  $\nu$ .

A figura ?? mostra que mesmo com a alteração dos parâmetros da distribuição *a priori* de  $\nu$  as cadeias continuam independente entre os espaçamentos.

Foram gerados 2 cadeias, se inicializando de pontos diferentes, a figura ?? mostra os traços das amostras, é possível notar que todas cadeias convergem para a mesma região,

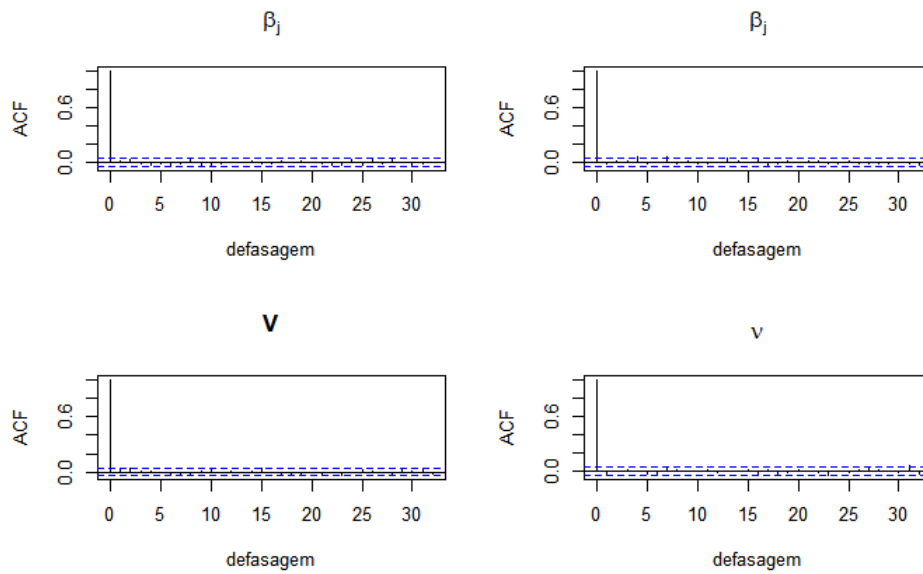


Figura 18: Gráfico de autocorrelação da amostra a posteriori dos parâmetros do modelo espacial.

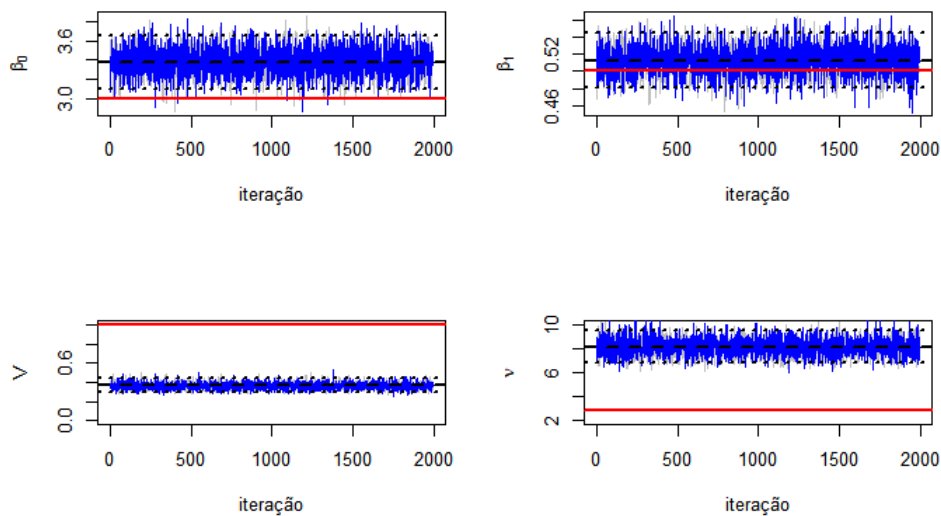


Figura 19: Traços das cadeias da amostra da *posteriori* de  $\beta_0, \beta_1, V$  e  $\nu$ , a linha tracejada representa a média *a posteriori*, as pontilhadas o intervalo de credibilidade de 95% e a linha vermelha cheia os valores verdadeiros.

porém as cadeias dos parâmetros  $\beta_0, \nu$  e  $V$  não convergem para a região do verdadeiro valor.

A Figura ?? mostra o histograma da distribuição *a posteriori* de  $\beta_0, \beta_1, V$  e  $\nu$ .

Na terceira análise de sensibilidade, quando atribuímos um conjunto de distribuição *a priori* observamos que a cadeia converge, porém só o intervalo de credibilidade do

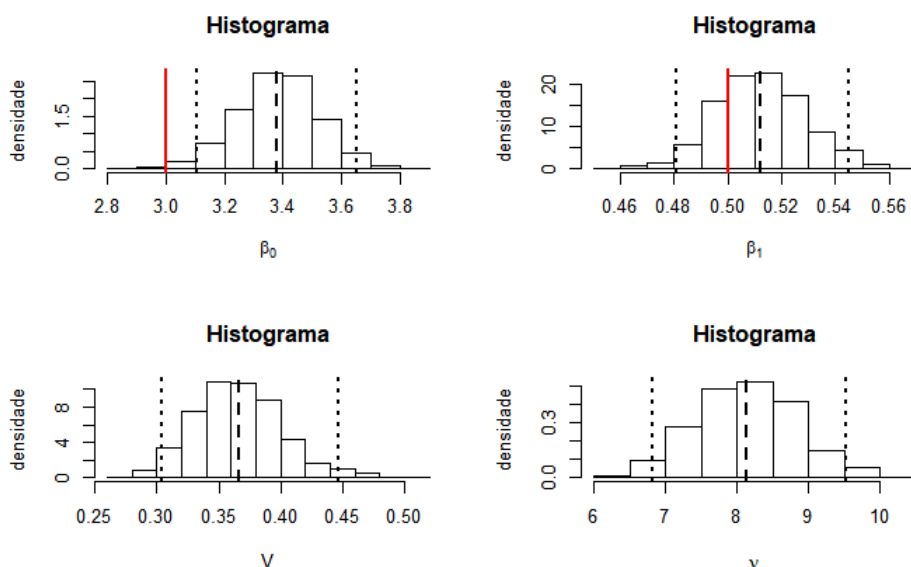


Figura 20: Histograma da distribuição *a posteriori* de  $\beta_0, \beta_1, V$  e  $\nu$ , a linha vermelha representa a média *a posteriori*, as pontilhadas o intervalo de credibilidade de 95% e a linha vermelha cheia os valores verdadeiros.

Tabela 6: Estimativa das cadeias do conjunto de distriuições *a priori* parcialmente informativa

| Parâmetro | valor teórico | média | IC          |
|-----------|---------------|-------|-------------|
| $\beta_0$ | 3             | 3,38  | (3,1;3,65)  |
| $\beta_1$ | 0,5           | 0,51  | (0,48;0,54) |
| $\nu$     | 2,63          | 8,13  | (6,82;9,53) |
| V         | 1             | 0,37  | (0,30;0,45) |

parâmetro  $\beta_1$  contém o verdadeiro valor do parâmetro. Desta forma, para distribuição *a priori* do parâmetro  $\nu$  com mais informação os parâmetros  $\nu$  e  $V$  são bastante afetados, além de mudar as estimativas de  $\beta_0$  significativamente, conclui-se que as estimativas do modelo não são adequada para distribuição *a priori* mais informativa .

## 3.2 Dados reais

Para ilustrar o modelo proposto, vamos aplicar um conjunto de dados reais. O conjunto de dados Meuse é fornecido pelo pacote `sp` no Rstudio.

O rio meuse tem 950 km de comprimento e nasce na França, no planalto de Langres em Bassigny e passa pela Bélgica, Países Baixos e segue em direção ao Mar do Norte.

O solo é um componente muito importante dentro do ecossistema, dentro das suas funções está servir como uma barreira natural à contaminação de águas subterrâneas.



Os metais pesados são considerados uma forte fonte de contaminação nos solos. Nesse contexto, medir os níveis de contaminação de uma determinada região é bastante importante. Para mais informação ??).

O banco de dados Meuse contém informação sobre 155 pontos georeferenciados onde se mediu a concentração de 4 metais pesados(cádmio, cobre, níquel e chumbo). A amostragem dos pontos foi feita num plano bidimensional na parte superficial do solo usando um intervalo de amostragem de 15m x 15m. Estes metais são transportados pelo rio meuse e depositados as suas margens. No presente trabalho foi utilizado o zinco como variável resposta e a distância do rio meuse até a localização como variável explicativa.

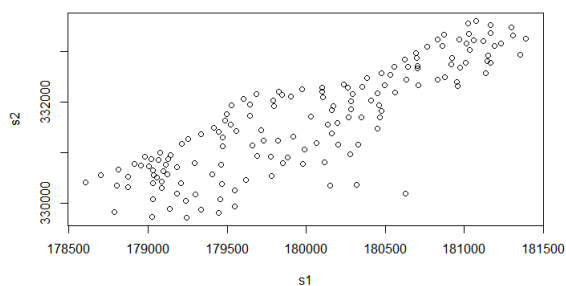
Para analisar este conjunto de dados vamos considerar que o logaritmo do zinco de cada localização é a nossa variável resposta  $Z_i$ , a raiz quadrada da distância de cada localização para o rio meuse será a nossa cováriavel, representada por  $x_i$ , e o efeito espacial é representado,  $V$  é a variância do processo espacial e  $M$  é a matriz da correlação espacial dado pela função exponencial, para facilitar na hora da estimação fazemos  $\phi = \frac{1}{\nu}$ , com isso,  $M = \exp\{-d_{ij} \cdot \nu\}$  e  $d_{ij}$ , onde  $\nu > 0$  é um parâmetro a ser estimado e  $d_{ij}$  é a distância euclidiana entre os pontos . Assim, o vetor paramétrico é formado por  $\theta = (\beta_0, \beta_1, V, \nu)$ .

Tabela 7: Estatísticas descritivas dos dados reais.

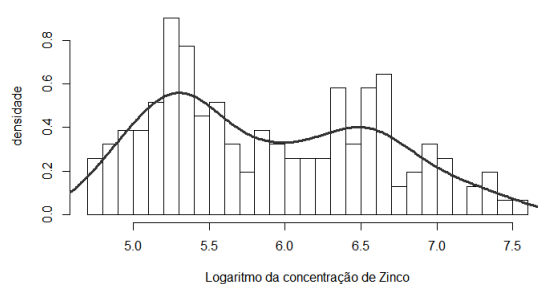
|           | Z    |
|-----------|------|
| Média     | 5,89 |
| Variância | 0,52 |
| Mediana   | 5,79 |

Dado que neste trabalho estamos estimando os parâmetros pela abordagem Bayesiana, é necessário atribuir uma distribuição *a priori* para cada parâmetro. Como não temos nenhuma crenças dos parâmetros que serão estimados, vamos atribuir uma distribuição *a priori* não informativa para eles, assim daremos ênfase a informação da amostra. Sendo assim, as distribuições da amostra seguirá dessa forma :  $V \sim GI(2, 1)$ ,  $\beta \sim N(\mathbf{0}, 1000I_k)$ ,  $I$ , onde  $\mathbf{0}$  é um vetor  $K$  dimensional com todos os elementos iguais a 0 e  $I_K$  é a matriz identidade de ordem  $K = 2$  e  $\nu \sim G(2, 1)$ .

Como feito no estudo simulado, foram realizadas 200000 interações. Utilizou-se um aquecimento de 1000 interações, isso significa, que foram desconsiderados os 1000 primeiros valores e um espaçamento de 200, retirou-se para compor apenas um a cada 200 iterações da cadeia. O tamanho da amostra final é de 996.



(a) Coordenadas meuse



(b) Histograma logaritmo de concentração de zinco

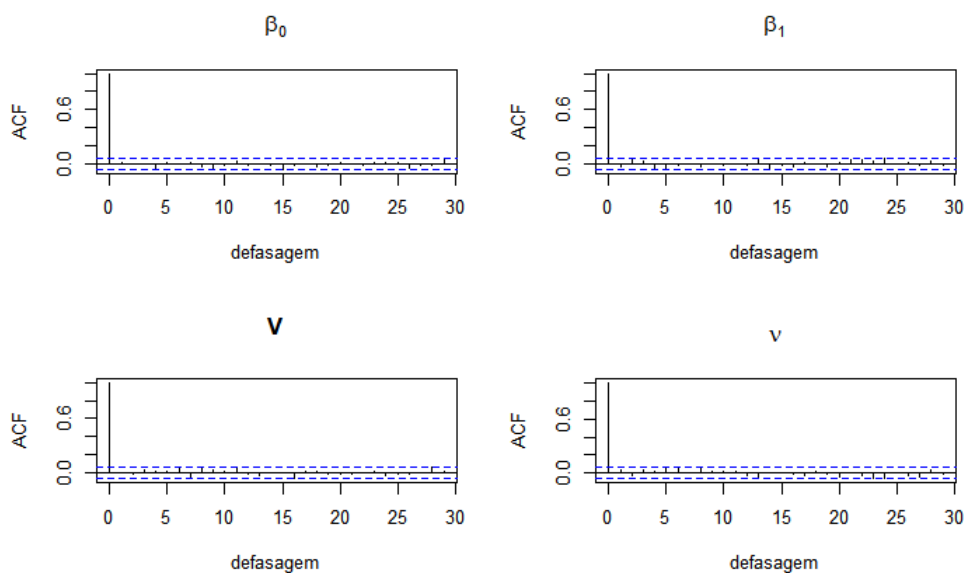


Figura 21: Gráfico de autocorrelação da amostra a posteriori dos parâmetros do modelo espacial.

A figura ?? apresenta os gráficos de autocorrelação das cadeias geradas das distribuições *a posteriori*, como todas as cadeias estão convergindo para zero, podemos considerar as cadeias são independentes entre si.

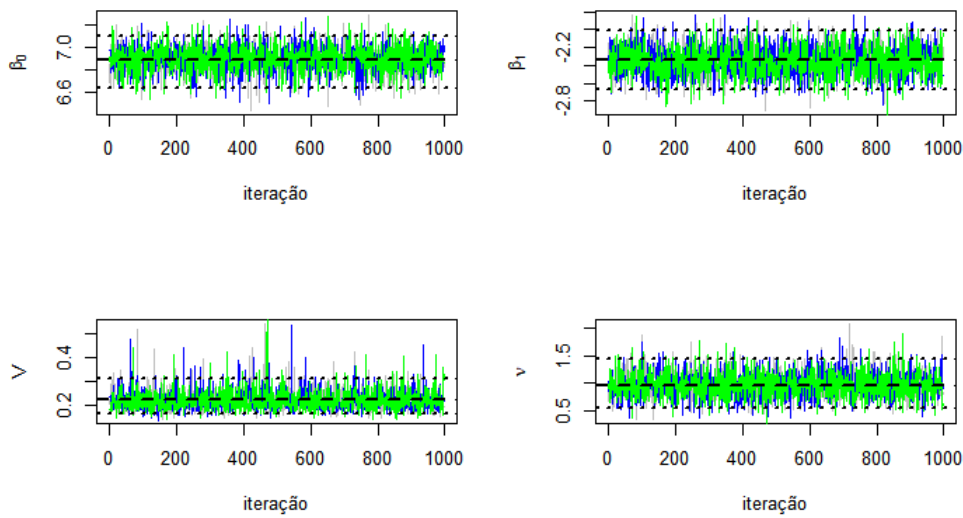


Figura 22: Traços das cadeias da amostra da *posteriori* de  $\beta_0, \beta_1, V$  e  $\nu$ , a linha tracejada representa a média *a posteriori* e as pontilhadas o intervalo de credibilidade de 95%.

Foram geradas 3 cadeias para cada um dos parâmetros obtidas ao se inicializar o amostrador em diferentes pontos, a imagem ?? apresenta o comportamento de cada parâmetro, a linha tracejada é média *a posteriori* dos parâmetros e seus intervalos de credibilidade de 95% são representados pelas linhas pontilhadas. Há indício de convergência pois as 3 cadeias estão convergindo para a mesma região.

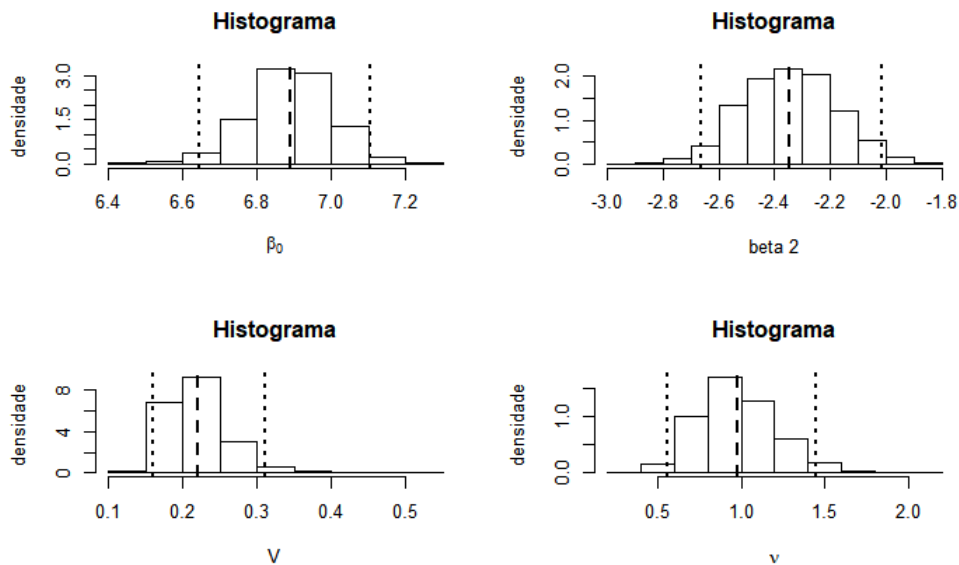


Figura 23: Histograma da distribuição *a posteriori* de  $\beta_0, \beta_1, V$  e  $\nu$ , a linha tracejada representa a média *a posteriori* e as pontilhadas o intervalo de credibilidade de 95%.

A Figura ?? apresenta os histogramas das cadeias da distribuições *a posteriori* dos parâmetros. As linha tracejada indica a média *a posteriori* do parâmetro, as linhas pontilhadas indicam o intervalo de credibilidade de 95%.

Tabela 8: Estimativa das cadeias dos dados reais

| Parâmetro | média | IC            |
|-----------|-------|---------------|
| $\beta_0$ | 6,89  | (6,65;7,11)   |
| $\beta_1$ | -2,35 | (-2,67;-2,01) |
| $\nu$     | 0,97  | (0,56;1,45)   |
| V         | 0,22  | (0,16;0,31)   |

Podemos notar que no intervalo de credibilidade de  $\beta_1$  o zero não está contido, então a raiz da distância é significativa no logaritmo da quantidade de zinco no solo, além disso, enquanto a raiz da distância ao rio meuse aumenta, o logaritmo da quantidade de concentração de zinco decresce.

Em relação ao efeito espacial dos dados, como visto na seção dos dados espaciais, o alcance é a distância que faz o processo de correlação tender para zero.

$$\exp\{-\nu d\} \rightarrow 0$$

Como essa a função do variograma que quantifica a relação espacial do evento é uma função exponencial, essa correlação nunca chegará a zero, portanto usaremos um valor aproximado para chegar no valor do alcance efetivo. Sendo assim:

$$\begin{aligned} \exp\{-\nu d\} &= 0,05 \\ -\nu d &= \ln 0,05 = -3 \\ d &= \frac{3}{\nu} \end{aligned}$$

Para chegar ao valor do alcance vamos usar a média da distribuição *a posteriori* simulada de  $\nu$ , com isso a distância entre as localizações dentro da qual as amostras apresentam-se correlacionadas espacialmente é  $d = \frac{3}{0,97} = 3,1$ .

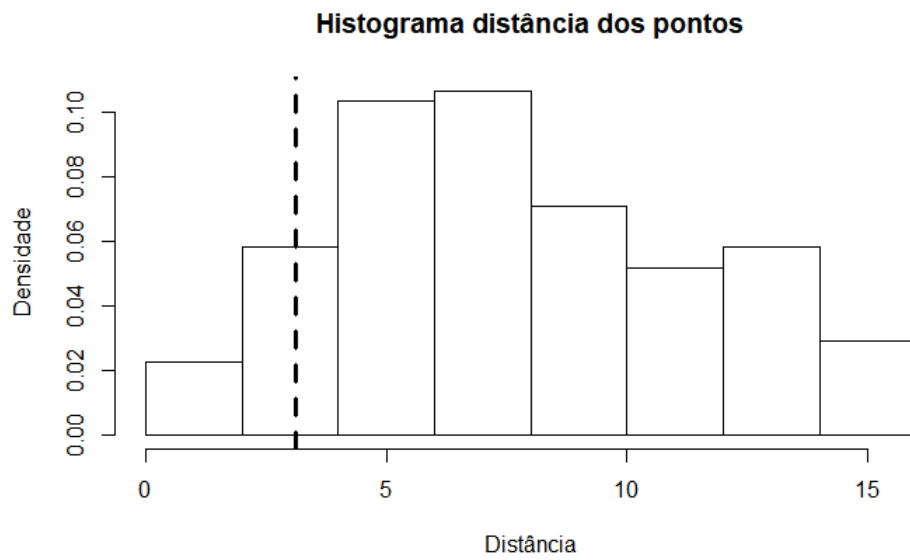


Figura 24: Histograma das distâncias dos pontos georeferenciados. A linha tracejada representa o valor do alcance efetivo 3,1.

## 4 Conclusão

Este trabalho propôs modelar dados geoestatístico através da função de covariância exponencial sob o enfoque Bayesiano.

Afim de testar a eficiência do modelo proposto para estimar dados geoestatístico, foi feito um estudo com dados simulados. Seja  $Z_i \sim N(\beta_0 + \beta_1 x, VM)$ , onde  $V$  é a variância do processo espacial e  $M$  é a matriz da correlação espacial dado pela função exponencial, para facilitar na hora da estimação foi feito  $\phi = \frac{1}{\nu}$ , com isso,  $M = \exp\{-d_{ij} \cdot \nu\}$  e  $d_{ij}$ , onde  $\nu > 0$  é um parâmetro a ser estimado e  $d_{ij}$  é a distância euclidiana entre os pontos, assim, o vetor paramétrico ficou  $\theta = (\beta_0, \beta_1, V, \nu)$ . Para amostrar as distribuições *a posteriori* foram geradas cadeias através do amostrador de Gibbs com passos de Metropolis-Hastings. Foram feitas 3 amostras se inicializado de diferentes pontos, e as amostras convergiram para o mesmo lugar. Todos os intervalos de credibilidades dos parâmetros estavam contido os valores teóricos, assim, mostramos a eficiência do modelo proposto.

Em seguida aplicou-se análise de sensibilidade no modelo proposto, foram testados 3 conjuntos de distribuição *a priori* diferentes, o primeiro conjunto com as distribuições *a priori* quase não informativo os valores teóricos continuaram contemplado no intervalo de credibilidade, o mesmo aconteceu para o conjunto de distribuições *a priori* não informativa, porém na terceira análise a distribuição *a priori* do parâmetro  $\nu$  era conhecida, assim os parâmetros  $\nu$  e  $V$  foram bastante afetados, apenas o intervalo de credibilidade de  $\beta_1$  conteve o valor teórico. Assim  $V, \beta_0$  e  $\nu$  não obtiveram sucesso na estimação. Em todos os estudos de sensibilidade as cadeias apresentaram convergência.

Para ilustrar o modelo proposto, foi aplicado um conjunto de dados reais. O conjunto de dados Meuse é oriundo do pacote `sp` no Rstudio. A amostra com 155 pontos georeferenciados foi feita num plano bidimensional na parte superficial do solo, o fenômeno de interesse ( $Z$ ) foi o logaritmo da concentração de zinco no solo e a covariável ( $X_1$ ) foi a raiz da distância do rio meuse ao ponto amostrado. O modelo capturou que o quanto maior fosse a raiz da distância o logaritmo da concentração de zinco no solo diminuía. Além

de estimar o alcance efetivo que foi 3,1. Cerca de 10% das localizações tem dependência espacial.

## 5 Apêndice

### 5.1 Apêndice A

Se  $\mathbf{Z} = (z_1, z_2, \dots, z_n) \sim N(\theta, \sigma^2)$  sendo  $\sigma^2$  conhecido e  $\theta \sim N(\mu_0, \tau_0^2)$ , temos que:

$$p(\mathbf{z}|\theta) \propto \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^n (z_i - \theta)^2\right\} \text{ e } p(\theta) \propto \exp\left\{\frac{-1}{2\tau_0^2}(\theta - \mu_0)\right\}$$

e portanto

$$\begin{aligned} p(\theta|\mathbf{z}) &\propto \exp\left\{\frac{-1}{2}[\sigma^{-2}(\theta^2 - 2 \sum_{i=1}^n z_i \theta) + \tau_0^{-2}(\theta^2 - 2\mu_0\theta)]\right\} \\ &\propto \exp\left\{\frac{-1}{2}[\theta^2(\sigma^{-2} + \tau_0^{-2}) - 2\theta(\sigma^{-2} \sum_{i=1}^n z_i + \tau_0^{-2}\mu_0)]\right\} \end{aligned}$$

os termos que não dependem de  $\theta$  foram incorporados à constante de proporcionalidade.

Assim, definimos  $\tau_1^{-2} = \sigma^{-2} + \tau_0^{-2}$  e  $\tau_1^{-2}\mu_1 = \sigma^{-2} \sum_{i=1}^n z_i - \tau_0^{-2}\mu_0$  a posteriori fica:

$$p(\theta|\mathbf{z}) \propto \exp\left\{-\frac{\tau_1^{-2}}{2}(\theta^2 - 2\theta\mu_1)\right\} \propto \exp\left\{-\frac{\tau_1^{-2}}{2}(\theta - \mu_1)^2\right\}$$

Portanto, a função densidade *a posteriori* tem os parâmetros  $\mu_1$  para média e  $\tau_1^2$  para variância.



## 5.2 Apêndice B

Se  $\mathbf{Z} = (z_1, z_2, \dots, z_n) \sim N(\theta, \sigma^2)$  sendo  $\sigma^2$  conhecido, temos que:

$$p(\mathbf{z}|\theta) = \left( \frac{1}{\sqrt{2\pi V}} \right)^n \exp\left\{-\frac{1}{2V} \sum_i^n (Z_i - \theta)^2\right\}$$

e portanto

$$\begin{aligned} \log p(\mathbf{z}|\theta) &= -\frac{n}{2} \log(2\pi V) - \frac{1}{2V} \sum_i^n (Z_i^2 - 2\theta z_i + \theta^2) \\ &= -\frac{n}{2} \log(2\pi V) - \frac{1}{2V} \sum_i^n z_i^2 + \frac{1}{V} \theta \sum_i^n z_i - \frac{n}{2} \theta^2 \end{aligned}$$

$$\frac{\partial \log p(\mathbf{z}|\theta)}{\partial \theta} = \frac{\sum_i^n z_i}{V} - \frac{n\theta}{V}$$

$$\frac{\partial^2 \log p(\mathbf{z}|\theta)}{\partial \theta \partial \theta} = -\frac{n}{V}$$

$$I[\theta] = E \left[ -\frac{\partial^2 \log p(\mathbf{z}|\theta)}{\partial \theta \partial \theta} \right] = \frac{n}{V}$$

Assim a distribuição *a priori* fica dessa forma:

$$p(\theta) \propto \left( \frac{n}{V} \right)^{-\frac{1}{2}} \propto k$$

## 5.3 Apêndice C

- **Gama inversa:**

Seja  $V \sim GI(a, b)$ , sua função de densidade é dada por:

$$p(V|a, b) = \frac{b^a}{\Gamma(a)} V^{-(a+1)} \exp\left\{-\frac{b}{V}\right\}, V > 0$$

para  $b > 0$  e  $a > 2$

$$E(V) = \frac{b}{a-1} \text{ e } Var(V) = \frac{b^2}{(a-1)^2(a-2)} \quad (5.1)$$

Também temos que  $\frac{1}{V} \sim G(a, b)$ .

# Referências

- CARVALHO, M. S.; SOUZA-SANTOS, R. Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. *Cadernos de Saúde Pública*, 2005.
- MACHADO, J. P.; OLIVEIRA, R. M. de; SOUZA-SANTOS, R. Análise espacial da ocorrência de dengue e condições de vida na cidade de Nova Iguaçu, Estado do Rio de Janeiro, Brasil. *Cadernos de Saúde Pública*, v. 25, n. 5, p. 1–10, 2009.
- OPROMOLLA, P. A.; DALBEN, I.; CARDIM, M. Análise geoestatística de casos de hanseníase no Estado de São Paulo, 1991-2002. *Revista de Saúde Pública*, 2006.
- ALMEIDA, A. Q. de et al. Geoestatística no estudo de modelagem temporal da precipitação. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 2011.
- COSTA, P. L. V. da. *Modelos espaço-temporais assimétricos para dados contínuos e categóricos*. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2014.
- PEREIRA, J. C. *Estimação do índice de abundância de um estoque pesqueiro com estrutura de correlação espacial: uma abordagem bayesiana*. Tese (Doutorado) — Universidade de São Paulo, 2009.
- CRESSIE, N. A. C. *Statistics for Spatial Data*. [S.l.]: John Wiley & Sons, 1993.
- BANERJEE, S.; GELFAND, A. E.; CARLIN, B. P. *Hierarchical Modeling and Analysis for Spatial Data*. [S.l.]: Chapman & Hall/CRC, 2003.
- LARSON, H. J. *Introduction to Probability Theory and Statistical Inference*. 3. ed. [S.l.]: John Wiley & Sons, 1982.
- CASELLA, G.; BERGER, R. L. *Statistical Inference*. 2. ed. [S.l.]: Duxbury, 2004.
- ROBERT, C. P.; CASELLA, G. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005. ISBN 0387212396.
- ROSS, S. M. *Simulation*. 4. ed. [S.l.]: Academic Press, 2006.
- METROPOLIS, N. et al. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, v. 21, n. 6, p. 1087–1092, 1953.
- HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, v. 57, p. 97–109, 1970.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, v. 6, n. 6, p. 721–741, 1984.

GELFAND, A. E.; SMITH, A. F. M. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, v. 85, p. 398–409, 1990.

COWLES, M. K.; CARLIN, B. P. Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 1996.

PLUMMER, M. et al. Coda: Convergence diagnosis and output analysis for mcmc. *R News*, v. 6, n. 1, p. 7–11, 2006. Disponível em: [i<https://journal.r-project.org/archive/>i](https://journal.r-project.org/archive/).

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2019. Disponível em: [i<http://www.R-project.org/>i](http://www.R-project.org/).

PEBESMA, E. J.; BIVAND, R. S. Classes and methods for spatial data in R. *R News*, v. 5, n. 2, p. 9–13, November 2005. Disponível em: [i<https://CRAN.R-project.org/doc/Rnews/>i](https://CRAN.R-project.org/doc/Rnews/).

PEREIRA, A. et al. Análise e monitoramento de metais pesados no solo. In: \_\_\_\_\_. [S.l.: s.n.], 2015. p. 71–89. ISBN 978-85-89274-22-7.