

Bruno Leonardo dos Santos Nobrega

**Comparando o poder preditivo de processos
de Cox log-Gaussianos: uma abordagem via
WAIC e validação cruzada**

Niterói - RJ, Brasil

20 de Julho de 2016

Bruno Leonardo dos Santos Nobrega

**Comparando o poder preditivo de
processos de Cox log-Gaussianos:
uma abordagem via WAIC e
validação cruzada**

Trabalho de Conclusão de Curso

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

Orientador: Prof. Jony Arrais Pinto Junior

Niterói - RJ, Brasil

20 de Julho de 2016

Bruno Leonardo dos Santos Nobrega

**Comparando o poder preditivo de processos
de Cox log-Gaussianos: uma abordagem via
WAIC e validação cruzada**

Monografia de Projeto Final de Graduação sob o título “*Comparando o poder preditivo de processos de Cox log-Gaussianos: uma abordagem via WAIC e validação cruzada*”, defendida por Bruno Leonardo dos Santos Nobrega e aprovada em 20 de Julho de 2016, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

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

Prof. Dr. Luis Guillermo Coca Velarde
Departamento de Estatística – UFF

Prof. Dr. Wilson Calmon Almeida dos Santos
Departamento de Estatística – UFF

**Ficha catalográfica elaborada pela Biblioteca do Instituto de Matemática e Estatística da
UFF**

N754 Nobrega, Bruno Leonardo dos Santos

Comparando o poder preditivo de processos de cox log –
Gaussianos: uma abordagem via WAIC e validação cruzada / Bruno
Leonardo dos Santos Nobrega. – Niterói, RJ: [s.n.], 2016.

32f.

Orientador: Prof. Dr. Jony Arrais PintoJunior
TCC (Bacharelado em Estatística) – Universidade Federal
Fluminense, 2016.

1.Estatística. 2.Modelo estatístico. 3. Poder preditivo. I. Título.

CDD 519.5

Resumo

Uma fase muito importante em qualquer análise estatística trata-se da escolha do modelo a ser utilizado. Muitos autores têm discutido a difícil tarefa de selecionar modelos tanto do ponto de vista frequentista quanto Bayesiano. Dentre os métodos mais utilizados pode-se citar o fator de Bayes e os critérios AIC, BIC, DIC, entre outros. Este projeto tem por objetivo estudar e comparar critérios para a seleção de modelos de Cox log-Gaussianos para padrões de pontos com base em métodos preditivos, isto é, o modelo M preferível será aquele que consegue prever melhor observações futuras geradas de um mesmo processo que os dados originais. O entendimento destas técnicas no contexto de dados georreferenciados é de extrema importância, uma vez que possibilitarão a escolha de um modelo parcimonioso com alto poder preditivo, permitindo que políticas públicas de saúde e segurança, por exemplo, sejam adotadas, uma vez que padrões de pontos de sinistros de veículos e óbitos por doenças infecciosas, como o HIV, em um município seriam típicos problemas tratados neste contexto. Das técnicas estudadas, se incluem critérios básicos propostos, obtendo-lhes por cálculos sobre a distribuição preditiva e outros critérios existentes na literatura com WAIC, LOO e K-fold. Neste projeto pretende-se utilizar uma abordagem completamente Bayesiana. Toda a metodologia será implementada no software livre R e OpenBUGS, com todos os resultados obtidos via estudos com dados simulados, tornando possível verificar se os critérios recuperam bem o modelo original gerador dos dados.

Palavras-chaves: padrões de pontos, comparação, poder preditivo

Dedicatória

Dedico toda minha família que sempre me apoiou em toda minha trajetória e se mostrava tão ansiosa quanto eu para esse momento. Dedico também a todos que me deram forças e incentivaram a continuar.

À minha mãe, Do Carmo, ao meu pai, Júnior e minha madrinha, Sueli, este trabalho é para vocês também.

Agradecimentos

Agradeço a Deus por sempre se mostrar presente na minha vida, e por me abençoar e me dar forças para concluir mais esta etapa.

Também agradeço à minha família, por todo apoio e incentivo, por cada momento que se mostraram interessados pela faculdade que eu estava fazendo. Obrigado, a minha mãe, meu pai, minhas irmãs, Dayane e Fernanda, meus sobrinhos, Lucas, Rafael e Lara(em memória), a todos os tios e primos dessa grande família.

Gostaria de agradecer a todos os meus professores, desde o Ensino Fundamental até a Faculdade, aqueles que proporcionaram o gosto pela aprendizagem. Em especial, ao meu orientador, Jony Arrais, por me orientar não só neste trabalho, mas também na iniciação científica, pela disposição e vontade de executar este trabalho.

Preciso agradecer também aos meus amigos, que me fizeram companhia nestes 4 anos de UFF, Amanda, Deborah, Hugo, Pedro Henrique, Pedro Romeiro, Rayssa, Thaylla. E claro, aos dois que sempre estiveram ali, estudando juntos, lembrando das listas e das provas e rindo boa parte do tempo, Andrine e Daniel.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 11
2	Objetivos	p. 13
3	Materiais e Métodos	p. 14
3.1	Inferência Bayesiana	p. 14
3.1.1	Estimadores Pontuais	p. 15
3.1.2	Estimadores Intervalares	p. 16
3.1.3	Distribuição Preditiva	p. 16
3.1.4	Método de Monte Carlo via Cadeias de Markov	p. 17
3.2	Estatística Espacial	p. 19
3.2.1	Processos Espaciais	p. 19
3.3	Modelos para Padrões de Pontos	p. 20
3.3.1	Processo de Poisson Homogêneo	p. 21
3.3.2	Processo de Cox Log-Gaussiano	p. 22
3.4	Critérios de Comparação	p. 27
3.4.1	Critérios Básicos	p. 27
3.4.2	Critérios WAIC e validação cruzada	p. 28
4	Análise dos Resultados	p. 34

4.1	Descrição da Simulação	p. 34
4.2	Comparação com critérios básicos	p. 37
4.3	Comparação com WAIC, LOO e K-fold	p. 38
5	Conclusão	p. 41
	Referências	p. 43

Lista de Figuras

1	Exemplo de padrões de pontos observados	p. 27
2	Padrão de Pontos obtido via simulação com $\beta = 1, 3$, $\mu = 0$, $\tau = 1$ e $\phi = 4, 25$	p. 35
3	Resultados estimando o Intercepto	p. 36

Lista de Tabelas

1	Estimativa e intervalo de credibilidade dos parâmetros e hiperparâmetros.	p. 36
2	Resultados dos critérios básicos utilizando a mediana como estimativas.	p. 37
3	Resultados dos critérios WAIC, LOO e K-fold para MC, ME1 e ME2 .	p. 38
4	Números efetivos de parâmetros de WAIC, LOO e K-fold para MC, ME1 e ME2	p. 38
5	Resultados dos critérios WAIC, LOO e K-fold para MC, ME1 e ME2 .	p. 39
6	Números efetivos de parâmetros de WAIC, LOO e K-fold para MC, ME1 e ME2	p. 39

1 Introdução

Uma fase muito importante em qualquer análise estatística trata-se da escolha do modelo a ser utilizado. Muitos autores têm discutido a difícil tarefa de selecionar modelos tanto do ponto de vista frequentista quanto Bayesiano.

Neste trabalho todo enfoque será dado na modelagem de padrões de pontos, que são dados nos quais se conhece o evento de interesse mas se tem as localizações destes eventos aleatória, numa abordagem totalmente bayesiana. Por exemplo, suponha que sejam coletados dados sobre as localizações de infecções por uma doença em alguma cidade. Note que nesse caso, se conhece as ocorrências, mas não suas localizações, que são aleatórias.

Suponha que neste cenário deseja-se entender a incidência das infecções por essa doença nos bairros desta cidade. Além disso, pode-se desejar entender como algumas covariáveis influenciam esta incidência, como por exemplo a qualidade do saneamento básico, o índice de desenvolvimento humano (IDH), o nível de poluição, com estas variáveis referentes aos bairros. Portanto para fazer um estudo desse tipo, é necessária a aplicação de modelos.

Uma possível modelagem nesse caso, é a utilização do modelo de Cox log-Gaussiano Møller *et al.*[1], que será o grande enfoque deste trabalho por ser um dos mais utilizados no contexto de padrões de pontos. Porém, mesmo com processo de Cox log-Gaussiano, por se tratar de uma classe de modelos, existe uma grande variedade de opções de modelagem para o mesmo problema. Assim, ao alterar e combinar diferentes covariáveis espaciais, ou escolher diferentes tipos de classes de função de correlação especifica-se diferentes modelos.

Sendo assim, naturalmente surge-se o seguinte questionamento "Qual é o melhor modelo neste caso?". Então, tendo em vista a difícil tarefa de seleção do modelo a ser utilizado, este projeto terá por objetivo estudar e comparar critérios para a seleção de modelos de Cox log-Gaussianos, devido a falta de estudos nesta área, principalmente se tratando de poder preditivo.

A abordagem mais utilizada na literatura são os critérios como AIC, BIC, DIC entre outros. Porém, estes métodos em geral, avaliam a qualidade do ajuste do modelo, como eles se comportam em relação aos dados observados. Entretanto, essa não é a única abordagem possível. Ao voltar para o exemplo citado acima, suponha nesta cidade, um dos bairros é uma comunidade carente e muito violenta, o que impossibilita a chegada de agentes de saúde na região e não é possível fazer observações nesta região. Neste caso, se gostaria de um modelo que melhor previsse a incidência dessa determinada doença neste bairro. Ou então, caso se queira fazer projeções da doença para os próximos 5 anos, com o intuito da criação de políticas públicas, o melhor modelo seria aquele que fizesse a melhor previsão.

Nestes cenários, o principal interesse não é escolher o modelo que se ajuste melhor aos dados observados e sim aquele que consiga fazer uma melhor previsão. Para isso, foram estudados os critérios C_1 , C_2 , C_3 e C_4 , que serão definidos no decorrer deste trabalho, que se mostraram muito custosos computacionalmente, pois eles eram gerados a partir de cálculos diretamente sobre as distribuições preditivas do modelo para cada região. Posteriormente, foram estudados critérios baseados no poder preditivo, conhecidos na literatura, aplicados em padrão de pontos, como WAIC, LOO e K-fold, que são menos custosos computacionalmente por se tratarem de aproximações das distribuições preditivas.

Com isso, os critérios que serão estudados, ou serão calculados diretamente em cima da distribuição preditiva, ou serão utilizados métodos para de alguma forma estimar como seriam as respectivas distribuições preditivas. Os principais deles serão WAIC, LOO e K-fold, pois fornecem uma forma de fazer essa comparação de forma que não seja custosa computacionalmente.

No Capítulo 2, serão apresentados os objetivos deste trabalho, e como se pretende trabalhar para alcançá-los.

No Capítulo 3, serão apresentados os Materiais e Métodos, que foram essenciais para obter os resultados almejados. Estarão apresentados conceitos sobre Inferência Bayesiana, sobre Estatística Espacial e sobre critérios de comparação baseados no poder preditivo.

No Capítulo 4 deste trabalho, estarão presentes os resultados obtidos pelo estudo de simulação. Tais como, resultados sobre a estimação dos parâmetros e sobre o cálculo dos critérios.

Por fim, no Capítulo 5, estarão a conclusão e os projetos futuros.

2 Objetivos

Este trabalho tem por objetivo entender e comparar critérios para a seleção de modelos para padrões de pontos com base em métodos preditivos, isto é, o modelo M preferível será aquele que consegue prever melhor observações futuras geradas de um mesmo processo que os dados originais. Para isso os pontos a serem observados serão:

- Inicialmente, estudar critérios básicos propostos, observar o comportamento e a decisão de cada um desses critérios, que envolvam cálculos diretos da distribuição preditiva.
- Estudar como se comportam critérios baseados no poder preditivo conhecidos na literatura como WAIC, LOO e K-fold Vehtari e Gelman[2], no contexto de padrões de pontos.
- Realizar um estudo de simulação para verificar o comportamento dos critérios citados no tópicos anteriores.

3 Materiais e Métodos

Neste capítulo serão apresentados os materiais utilizados neste trabalho bem como os métodos de análise e principalmente os métodos de comparação de modelo, que é o foco deste estudo. Para isso, será necessário apresentar algumas definições, passando por Inferência Bayesiana e por Estatística Espacial, para se ter bem definidos os modelos que serão utilizados e o contexto que será aplicado todos os critérios. Em seguida, finalmente, aprofundar nos critérios de comparação com base no poder preditivo.

3.1 Inferência Bayesiana

Quando se obtém os dados, resultantes de um estudo por amostragem, deseja-se utilizar os dados dessa amostra para poder descrever a população total. Quando alguém utiliza deste método, ele está inferindo sobre a população, daí o termo Inferência. Na maioria das vezes, é observado o uso da Inferência Clássica, neste caso o interesse está dado na Inferência Bayesiana

A Inferência Bayesiana é baseada no teorema de Bayes. Aqui o parâmetro θ torna-se, também, uma variável aleatória. Por mais que o parâmetro seja desconhecido é tomada uma informação prévia sobre ele, que estará contida em $p(\theta)$, conhecida com distribuição *a priori*. A partir disso, e da função de verossimilhança dos dados $f(\mathbf{y}, \theta)$ é possível calcular $p(\theta|\mathbf{y})$, a chamada distribuição *a posteriori* de θ . Logo, pelo teorema de Bayes:

$$p(\theta|\mathbf{y}) = \frac{f(\mathbf{y}, \theta)p(\theta)}{p(\mathbf{y})}, \quad (3.1)$$

em que,

$$\begin{aligned} p(\mathbf{y}) &= \int p(\mathbf{y}, \theta)d\theta \\ &= \int f(\mathbf{y}|\theta)p(\theta)d\theta. \end{aligned}$$

Após a definição da distribuição *a posteriori* o mais importante é definir estimadores para o parâmetro θ , a partir de $p(\theta|\mathbf{y})$ é possível construir estimadores pontuais e intervalares que serão apresentados nas subseções a seguir.

3.1.1 Estimadores Pontuais

Um estimador pontual é qualquer função da amostra. Em geral, por meio dela, deseja-se recuperar o valor verdadeiro de um parâmetro θ . Pelo método bayesiano, os estimadores são, em sua maioria, medidas resumo da distribuição *a posteriori*. Esses estimadores são obtidos a partir do estimador de Bayes.

Estimador de Bayes

O estimador de Bayes é aquele escolhido para minimizar a média *a posteriori* de alguma função que meça o quão longe está um estimador do parâmetro de interesse. Então, torna-se necessário definir funções que atendam essa necessidade.

Funções Perda

As funções perda são, justamente, aquelas que controlam a distância do estimador e o parâmetro. A partir delas é que serão definidos quais os estimadores que serão utilizados. Existem 3 funções perdas principais que são utilizadas, são definidas por:

1. Perda Quadrática

$$\ell(\theta, a) = (\theta - a)^2.$$

2. Perda Absoluta

$$\ell(\theta, a) = |\theta - a|.$$

3. Perda 0-1

$$\ell(\theta, a) = \begin{cases} 0, & \text{se } |\theta - a| > \epsilon, \quad \forall \epsilon > 0. \\ 1, & \text{se } |\theta - a| < \epsilon, \quad \forall \epsilon > 0. \end{cases}$$

Além de controlar a distância, deseja-se aquele estimador que tenha a menor distância em média. Para isso, é necessário utilizar a perda esperada.

Perda Esperada

A perda esperada é definida da forma,

$$E[\ell(\theta, a)] = \int_{\Omega} \ell(\theta, a)p(\theta)d\theta,$$

sendo $p(\theta)$ tal como em 3.1, portanto, ao se escolher um estimador escolhe-se aquele que minimize a perda esperada.

Os estimadores pontuais bayesianos são obtidos a partir da minimização da Perda Esperada. Então, utilizando as funções perda: perda quadrática, perda absoluta e perda 0-1, tem-se os estimadores que serão a média, a mediana e a moda, *a posteriori*, respectivamente.

A seguir, será apresentado como obter estimadores intervalares dentro desse contexto. Os estimadores intervalares bayesianos são os chamados intervalos de credibilidade.

3.1.2 Estimadores Intervalares

Pela definição, um intervalo de credibilidade C de $100(1 - \alpha)$ para θ é um intervalo tal que, $\int_C p(\theta|\mathbf{y})d\theta = 1 - \alpha$. Pode ser interpretado como qualquer intervalo da distribuição *a posteriori* da forma $[Q_{\alpha_1}, Q_{\alpha_2}]$, tal que $\alpha_2 - \alpha_1 = 1 - \alpha$. Em geral, toma-se o intervalo central, ou seja, $[Q_{\frac{\alpha}{2}}, Q_{1-\frac{\alpha}{2}}]$, que em distribuições simétricas corresponde ao intervalo de menor amplitude.

Como o enfoque será comparar modelos com base naquele que melhor prevê uma observação, será utilizada, para fazer estas previsões, a chamada distribuição preditiva apresentada a seguir.

3.1.3 Distribuição Preditiva

Na abordagem Bayesiana, existe uma distribuição muito útil para fazer previsões de dados não observados, a chamada distribuição preditiva. Suponha que tenham sido observados y_1, \dots, y_n em determinada amostra e deseja-se prever o valor de y_{n+1} , utilizando, para isso, os dados observados. A distribuição preditiva é manipulada da mesma forma que a distribuição *a posteriori*, estima-se a partir de medidas resumos da distribuição. A distribuição preditiva, denotada por $p(y_{n+1}|\mathbf{y})$ é definida por:

$$p(y_{n+1}|\mathbf{y}) = \int p(\theta|\mathbf{y})p(y_{n+1}|\theta)d\theta,$$

em que, $p(\theta|\mathbf{y})$ é a distribuição *a posteriori* de θ , como em 3.1, e $p(y_{n+1}|\theta)$ é a distribuição suposta para os dados aplicada em y_{n+1} .

Para melhor se trabalhar, com as distribuições *a posteriori* e preditiva, será necessário muitas das vezes um método de simulação, devido a dificuldade de se trabalhar com essas

distribuições. Portanto, a seguir será apresentado o método de Monte Carlo via Cadeia de Markov.

3.1.4 Método de Monte Carlo via Cadeias de Markov

Como dito anteriormente, os estimadores são obtidos a partir de medidas resumo da distribuição *a posteriori* $p(\boldsymbol{\theta}|\mathbf{y})$. Em geral, essa distribuição é complexa, de difícil manipulação, portanto utiliza-se de simulação de amostras dessa distribuição para se obter as estimativas para cada parâmetro. Portanto, um caminho para simular estas amostra é o Método de Monte Carlo via Cadeias de Markov(MCMC), que após a convergência os valores gerados por ele simulam uma amostra de $p(\boldsymbol{\theta}|\mathbf{y})$.

Existem dois métodos de utilização do MCMC, conhecidos como o Amostrador de Gibbs e o algoritmo de Metropolis-Hastings, com aplicações em diferentes circunstâncias que serão explicitadas abaixo.

Amostrador de Gibbs

O Amostrador de Gibbs Gelfand e Smith[3] é utilizado quando a distribuição condicional completa para o parâmetro de interesse pertence a uma família de distribuição conhecida, ou seja, de fácil amostragem direta, porque simula valores obtidos diretos das condicionais completas $p(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{y})$ de cada parâmetro.

Distribuição Condicional Completa: A condicional completa para θ_i é a distribuição deste, condicionado a todos os outros parâmetros de $\boldsymbol{\theta}$ e os dados \mathbf{y} , denotada por $p(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{y})$.

O amostrador de Gibbs consiste num esquema de amostragem baseado em sucessivas gerações dessas distribuições, com descrito abaixo:

1. Determina-se valores iniciais para cada uma das componentes do parâmetro $\boldsymbol{\theta}$, tal como, $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_d^{(0)})'$.
2. Inicializa-se o contador de iterações $j = 1$.

3. Obtem-se $\boldsymbol{\theta}^{(j)}$ a partir das seguintes instruções:

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

4. Acrescente j de 1 e repita os passos 3 e 4 até a convergência.

A medida que o número de iterações cresce e a convergência é alcançada pode se dizer que se obteve uma amostra para $\boldsymbol{\theta}$.

Algoritmo de Metropolis-Hastings

Quando não se tem uma condicional completa conhecida, utiliza-se o Algoritmo de Metropolis-Hastings Metropolis *et al.*[4], Hastings[5], porque este não amostra direto das condicionais completas, mas utiliza uma distribuição de auxílio, chamada distribuição proposta, para fazê-lo. Este método pode ser descrito pelos seguintes passos:

1. Semelhante ao amostrador de Gibbs, determina-se valores iniciais para $\boldsymbol{\theta}$, tal como, $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_d^{(0)})'$.
2. Inicializa-se o contador de iterações $j = 1$.
3. Cria-se $q(\epsilon|\boldsymbol{\theta}^{(j-1)})$ vinda de uma distribuição conhecida, a chamada distribuição proposta. Aceita-se o ponto obtido com probabilidade $\min \left\{ 1, \frac{p(\epsilon) q(\boldsymbol{\theta}^{(j-1)}|\epsilon)}{q(\epsilon|\boldsymbol{\theta}^{(j-1)}) p(\boldsymbol{\theta}^{(j-1)})} \right\}$, sendo $p(\cdot)$ a distribuição de interesse. Se o ponto for aceito $\boldsymbol{\theta}^{(j)} = \epsilon$, caso contrário $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^{(j-1)}$.
4. Acrescenta-se j em 1, e repete-se 3 e 4 até a convergência.

Com esses dois algoritmos é possível obter as amostras das distribuições *a posteriori* necessárias para o andamento do trabalho.

Terminado os conceitos necessários presentes em Inferência Bayesiana, agora serão apresentados os conceitos necessários de Estatística Espacial.

3.2 Estatística Espacial

A chamada Estatística Espacial é utilizada quando, em um determinado estudo, as características variam de acordo com o espaço, tendo este, de fato, interferência na característica de interesse, o que o torna uma variável relevante para as análises. Existem exemplos desse comportamento nas mais diversas áreas de estudo.

Essas variáveis que mudam de comportamento de acordo com o espaço podem ser mensuradas de diferentes maneiras, com isso há uma diferença nos conceitos dentro da Estatística Espacial.

3.2.1 Processos Espaciais

Para descrever e entender melhor os dados descritos anteriormente, são utilizados os chamados processos espaciais. Um processo espacial é aquele que tem seu domínio definido no espaço. Ele pode ser definido por:

$$\{Z(s) : s \in D \subset \mathbb{R}^2\},$$

em que $Z(s)$ é a variável de interesse na região s , e D a região de interesse onde encontra-se s . Os processos espaciais podem ser de três tipos, que serão explicitados abaixo.

- **Dados Geoestatísticos**

Neste tipo de processo, a variável de interesse $Z(s)$ é aleatória e as localizações s são fixas dentro de uma região D contínua. Por exemplo, a medição da umidade relativa do ar em alguns pontos de uma cidade.

- **Dados de área**

Este é um caso semelhante do anterior, a variável $Z(s)$ é aleatória, porém as localizações s são fixas e discretas, formam uma partição da região D . Por exemplo, contabilizando o número de infectados por determinada doença, por cidade, em algum estado.

- **Padrões de Pontos**

Os padrões de pontos se diferenciam dos outros dois, porque se tem $Z(s) = 1$, a ocorrência de interesse conhecida, mas localizações s são aleatórias. Por exemplo, as localizações das ocorrências de assalto numa cidade, se conhece o evento de interesse, mas é necessário coletar os dados sobre as localizações desses assaltos.

Deste último exemplo, a seguir serão mostrados modelos aplicáveis nesta situação.

3.3 Modelos para Padrões de Pontos

Dentro do contexto de padrões de pontos existem alguns tipos de modelos para o ajuste dos dados tais como Processo de Poisson homogêneo, Processo de Poisson não homogêneo, Processo de Cox e Processo de Cox Log-Gaussiano. Este último tendo maior enfoque neste trabalho, devido a sua importância dentro da área, por ser um dos modelos mais utilizados na literatura para a modelagem de padrões de ponto. Inicialmente serão apresentadas algumas definições necessárias para construir os modelos e a seguir a definição de alguns deles.

Processo Gaussiano

Um processo gaussiano $Y(\cdot)$ definido no espaço, ou seja, em $S \subset \mathbb{R}^2$, é tal que, para qualquer $\{\mathbf{s}_1, \dots, \mathbf{s}_n\} \subset S, \forall n > 1$,

$$Y(\mathbf{s}) \sim N_n(\boldsymbol{\mu}, \Sigma).$$

- O processo será dito estacionário se $\boldsymbol{\mu} = \mu \mathbf{1}$ e $\Sigma = \sigma^2 R$, sendo R uma matriz de correlação $R_{ij} = \rho(s_i, s_j)$, ou seja, as médias $\boldsymbol{\mu}$ e as variâncias de $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ são constantes e iguais a μ e σ^2 .
- O processo será dito isotrópico se, além de ser estacionário, a correlação entre as localizações depende apenas da distância $\|s_i - s_j\|$ entre elas.

Neste trabalho, será utilizada a seguinte notação para um processo gaussiano $Y(\cdot) \sim PG(\mu, \tau, \rho_\phi)$. Em que, τ será a precisão do processo gaussiano, ou seja, $\tau = \frac{1}{\sigma^2}$ e ρ_ϕ a função de correlação que dependerá de um parâmetro de alcance ϕ .

Nas situações que serão apresentadas neste trabalho, a matriz de correlação R é definida através de algumas famílias de correlação, que dependem da distância entre as ocorrências. Estas serão apresentadas a seguir.

Funções de Correlação Espacial

As funções de correlação espacial são aquelas que determinam o quanto uma observação em uma determinada localização s_i influencia uma outra localização s_j . Se o arranjo espacial for isotrópico, ou seja, a correlação entre as localizações depende apenas da distância entre elas é preferível que, a medida que a distância cresça a correlação entre

as localizações diminua e que $\rho(s_i, s_j) \rightarrow 0$ se $\|s_i - s_j\| \rightarrow \infty$. Com isso, em cada função de correlação existirá pelo menos um parâmetro ϕ , o parâmetro de alcance, que controlará a velocidade com que a função tende a zero. Para o trabalho a principal família de função de correlação utilizada foi a família exponencial potência.

- **Família Exponencial Potência**

As funções de correlação desta família são da forma:

$$\rho(s_i, s_j; \phi, \kappa) = \exp \left\{ - \left(\frac{\|s_i - s_j\|}{\phi} \right)^\kappa \right\},$$

assim, tem-se que $\phi > 0$ é o parâmetro de alcance e $\kappa \in [0, 2]$. Então, se $\kappa = 1$, a função chama-se Função de Correlação Exponencial e se $\kappa = 2$, Função de Correlação Gaussiana.

Existem outras famílias de funções de correlação espacial como a Família Esférica e a Matérn, mas, aqui, os esforços foram concentrados na família exponencial potência. Maiores informações sobre estas outras funções de correlação podem ser encontradas em Reis[6].

3.3.1 Processo de Poisson Homogêneo

Neste tipo de modelagem, o número de eventos N em uma região $A \subset \mathbb{R}^2$ é uma variável aleatória Poisson com média $\lambda|A|$, sendo $|A|$ a área de A , além disso, as contagens nas regiões disjuntas são independentes. Este processo tem sua intensidade $\lambda(s) = \lambda$ e seu efeito de segunda ordem $\lambda_2(s_i, s_j) = \lambda^2$ que independe das regiões, é estacionário e isotrópico.

Com essa definição, em que a intensidade λ não varia com as diversas localizações $\{s_i\}_{i=1}^n$ tem-se que a função de verossimilhança para este modelo será:

$$L(\lambda; N) \propto (\lambda|A|)^N \exp(-\lambda|A|).$$

Esse modelo foi um dos primeiros a serem utilizados neste tipo de contexto. Este modelo não é aplicável a maioria das situações, porque considera que todas as regiões tem intensidades iguais o que na realidade raramente acontece. Então é utilizado apenas como base para comparações.

3.3.2 Processo de Cox Log-Gaussiano

Para definir o processo de Cox log-Gaussiano que será o objeto de interesse deste trabalho, considere um conjunto de localizações $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$, tal que $\mathbf{s}_i \in S \forall i$, que apontam o local onde o evento i ocorreu na região de interesse S . O modelo é definido, assim como em Benes *et al.*[7] de forma hierárquica por:

$$\begin{aligned} X &\sim PP(\Lambda(\cdot)), \\ \Lambda(\mathbf{s}) &= r(\mathbf{s})\lambda(\mathbf{s}), \forall \mathbf{s} \in S, \\ \log \lambda(\mathbf{s}) &= w(\mathbf{s}) + \boldsymbol{\beta}'\mathbf{z}(\mathbf{s}), \\ w(\cdot) &\sim PG(\mu, \tau, \rho_\phi), \end{aligned}$$

em que X é um processo de Poisson de intensidade $\Lambda(\cdot)$, $r(\mathbf{s})$ é uma constante conhecida, que pode representar a densidade populacional, por exemplo, $\mathbf{z}(\mathbf{s})$ são as covariáveis associadas ao espaço, $\boldsymbol{\beta}$ é o vetor de incrementos das covariáveis na intensidade do processo e $w(\cdot)$ é um processo Gaussiano com média μ , precisão τ e função de correlação espacial ρ_ϕ . Essa decomposição implica num processo de Cox Log-gaussiano, pois X é um processo de Poisson e $\log(\Lambda(\cdot))$ é um processo Gaussiano.

Portanto, nesse caso a função de verossimilhança é da forma:

$$L(\Lambda(\cdot), \{s_j\}_{j=1}^n) \propto \prod_{j=1}^n [r(s_j) \exp\{w(s_j) + \boldsymbol{\beta}'\mathbf{z}(s_j)\}] \exp\left\{-\int_S \Lambda(s) ds\right\}. \quad (3.2)$$

Para fazer inferência para este modelo, ele não poderá ser utilizado como da forma apresentada acima, mas sim da forma discretizada como a seguir.

Discretização do Modelo

Como não é possível tratar o modelo de Cox Log-gaussiano da forma que ele é apresentado, pois ele é contínuo, e seria necessário tratar uma quantidade infinita de pontos, o que se faz é dividir a área total S , em múltiplas sub-regiões, o que é chamado de discretização do modelo.

Ao escrever o modelo da forma discretizada, $S = \bigcup_{k=1}^K S_k$, é possível defini-lo como:

$$\begin{aligned} X &\sim PP(\Lambda(\cdot)), \\ \Lambda(s_j) &= r_k \lambda_k, \forall s_j \in S_k, \\ \log \lambda_k &= w_k + \boldsymbol{\beta}' \mathbf{z}_k, \\ \mathbf{w} &\sim N_p(\mu \mathbf{1}, \tau^{-1} R_\phi), \end{aligned} \quad (3.3)$$

com a nova função de verossimilhança sendo:

$$L(\Lambda(\cdot), \{s_j\}_{j=1}^n) \propto \prod_{j=1}^n [r_{[j]} \exp\{w_{[j]} + \boldsymbol{\beta}' \mathbf{z}_{[j]}\}] \exp \left\{ - \sum_{k=1}^K r_k \exp\{w_k + \boldsymbol{\beta}' \mathbf{z}_k\} |S_k| \right\}, \quad (3.4)$$

em que nesta função de verossimilhança, $[j]$ indicará a qual partição k , pertence a ocorrência s_j , para $\forall j \in \{1, \dots, n\}$.

Inferindo sobre o Modelo

Pelo ponto de vista bayesiano, para estimar os parâmetros e hiperparâmetros desconhecidos \mathbf{w} , $\boldsymbol{\beta}$, μ , τ e ϕ , necessita-se obter suas respectivas distribuições *a posteriori*, para isso determinou-se as distribuições *a priori* para cada parâmetro do modelo:

$$\begin{aligned} \mu &\sim N(a, b), \\ \tau &\sim \text{Gama}(c, d), \\ \phi &\sim \text{Gama}(g, h), \\ \boldsymbol{\beta} &\sim N_q(\mathbf{u}, V), \end{aligned}$$

sendo *a priori* para \mathbf{w} a normal multivariada dita na discretização do modelo. Assumindo independência *a priori* dos parâmetros e hiperparâmetros, a distribuição *a priori* conjunta será:

$$\begin{aligned} p(\mu, \tau, \phi, \mathbf{w}) &= p(\mu)p(\tau)p(\phi)p(\mathbf{w}) \\ &\propto \exp \left\{ -\frac{1}{2b}(\mu - a)^2 \right\} \tau^{c-1} e^{-d\tau} \phi^{g-1} e^{-h\phi} \frac{1}{[\tau^{-K} |R_\phi|]^{\frac{1}{2}}} \times \\ &\exp \left\{ -\frac{1}{2}(\mathbf{w} - \mu \mathbf{1})'(\tau^{-1} R_\phi)^{-1}(\mathbf{w} - \mu \mathbf{1}) \right\} \exp \left\{ -\frac{1}{2}(\boldsymbol{\beta} - \mathbf{u})'(V)^{-1}(\boldsymbol{\beta} - \mathbf{u}) \right\}. \end{aligned}$$

Assim, a distribuição *a posteriori* conjunta do vetor de parâmetros é dada como:

$$\begin{aligned}
p(\mu, \tau, \phi, \mathbf{w} | \{s_j\}_{i=1}^n) &= L(\Lambda(\cdot), \{s_j\}_{i=1}^n) p(\mu, \tau, \phi, \mathbf{w}) \\
&\propto \prod_{j=1}^n [\exp\{w_{[j]} + \boldsymbol{\beta}' \mathbf{z}_{[j]}\}] \exp \left\{ - \sum_{k=1}^K r_k \exp\{w_k + \boldsymbol{\beta}' \mathbf{z}_{[j]}\} |S_k| \right\} \\
&\times \exp \left\{ - \frac{1}{2b} (\mu - a)^2 \right\} \tau^{c-1} e^{-d\tau} \phi^{g-1} e^{-h\phi} \frac{1}{[\tau^{-K} |R_\phi|]^{1/2}} \\
&\times \exp \left\{ \frac{1}{2} (\mathbf{w} - \mu \mathbf{1})' (\tau^{-1} R_\phi)^{-1} (\mathbf{w} - \mu \mathbf{1}) \right\} \exp \left\{ - \frac{1}{2} (\boldsymbol{\beta} - \mathbf{u})' (V)^{-1} (\boldsymbol{\beta} - \mathbf{u}) \right\}.
\end{aligned}$$

Para inferir sobre os parâmetros utilizando o método de Monte Carlo via Cadeias de Markov é necessário conhecer a chamada distribuição condicional completa definida na subseção 3.1.4 e a partir dela, definir se é utilizado o método do Amostrador de Gibbs, ou o Algoritmo de Metropolis-Hastings. Então, a distribuição condicional completa para μ é dada por:

$$\begin{aligned}
p(\mu | \tau, \phi, \mathbf{w}, \boldsymbol{\beta}) &\propto \exp \left\{ - \frac{1}{2b} (\mu - a)^2 - \frac{1}{2} (\mathbf{w} - \mu \mathbf{1})' (\tau^{-1} R_\phi)^{-1} (\mathbf{w} - \mu \mathbf{1}) \right\} \\
&= \exp \left\{ - \frac{1}{2} \left[\frac{1}{b} (\mu - a)^2 + (\mathbf{w} - \mu \mathbf{1})' (\tau^{-1} R_\phi)^{-1} (\mathbf{w} - \mu \mathbf{1}) \right] \right\},
\end{aligned}$$

Portanto analisando a condicional completa para μ pode-se presumir que $\mu | \tau, \phi, \mathbf{w}$ não terá distribuição conhecida. A distribuição condicional completa para τ será:

$$\begin{aligned}
p(\tau | \mu, \phi, \mathbf{w}, \boldsymbol{\beta}) &\propto \tau^{\frac{K}{2} + c - 1} \exp \left\{ -d\tau - \frac{1}{2} (\mathbf{w} - \mu \mathbf{1})' (\tau^{-1} R_\phi)^{-1} (\mathbf{w} - \mu \mathbf{1}) \right\} \\
&= \tau^{\frac{K}{2} + c - 1} \exp \left\{ - \left[d + \frac{1}{2} (\mathbf{w} - \mu \mathbf{1})' (R_\phi)^{-1} (\mathbf{w} - \mu \mathbf{1}) \right] \tau \right\}.
\end{aligned}$$

Para τ , pode-se dizer que esta condicional completa segue uma distribuição conhecida, assim, $\tau | \mu, \phi, \mathbf{w} \sim \text{Gama} \left(\frac{K}{2} + c, \left[d + \frac{1}{2} (\mathbf{w} - \mu \mathbf{1})' (R_\phi)^{-1} (\mathbf{w} - \mu \mathbf{1}) \right] \right)$. Para os outros parâmetros, ϕ , $\boldsymbol{\beta}$ e \mathbf{w} ainda tem-se que a distribuição será:

$$p(\phi | \mu, \tau, \mathbf{w}, \boldsymbol{\beta}) \propto \frac{\phi^{g-1}}{|R_\phi|} \exp \left\{ -h\phi - \frac{1}{2} (\mathbf{w} - \mu \mathbf{1})' (\tau^{-1} R_\phi)^{-1} (\mathbf{w} - \mu \mathbf{1}) \right\}.$$

Para ϕ não é possível, encontrar uma distribuição entre as famílias de distribuições conhecidas, portanto a solução é utilizar o algoritmo de Metropolis-Hastings. Para \mathbf{w} e

β as distribuições serão:

$$p(\mathbf{w}|\mu, \tau, \phi, \beta) \propto \exp \left\{ \sum_{j=1}^n w_{[j]} - \sum_{k=1}^K r_k \exp\{w_k + \beta' \mathbf{z}_k\} |S_k| \right\} \\ \times \exp \left\{ -\frac{1}{2} (\mathbf{w} - \mu \mathbf{1})' (\tau^{-1} R_\phi)^{-1} (\mathbf{w} - \mu \mathbf{1}) \right\},$$

$$p(\beta|\mu, \tau, \phi, \mathbf{w}) \propto \exp \left\{ \sum_{j=1}^n \beta' \mathbf{z}_{[j]} - \sum_{k=1}^K r_k \exp\{w_k + \beta' \mathbf{z}_k\} |S_k| \right\} \\ \times \exp \left\{ -\frac{1}{2} (\beta - \mathbf{u})' (V)^{-1} (\beta - \mathbf{u}) \right\}.$$

Concluindo, para os parâmetros μ , ϕ , β e \mathbf{w} , não é possível determinar nenhuma distribuição de probabilidade conhecida, como ocorreu com τ . Portanto ao utilizar o MCMC, utiliza-se o método de Amostrador de Gibbs para τ e para μ , ϕ , β e \mathbf{w} o método de Metropolis-Hastings.

Agora, é possível obter a distribuição preditiva para os Processos de Cox log-Gaussianos. Como obtê-las é mostrado abaixo.

Distribuição Preditiva para Processos de Cox log-Gaussianos

Após todo o trabalho de cálculos, das distribuições *a posteriori* e condicionais completas, o interesse principal será trabalhar com a distribuição preditiva dos Processos de Cox log-Gaussianos. As distribuições preditivas serão para o número total de ocorrências observadas em determinada região S_k (n_{S_k}). Portanto utilizando a definição de distribuição preditiva apresentada e os cálculos obtidos acima, é possível obter:

$$p(X_k|X_{-k}, M) = \int p(X_k|X_{-k}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|X_{-k}) d\boldsymbol{\theta} \\ = \int [\Lambda_k]^{n_{S_k}} \exp\{-\Lambda_k |S_k|\} p(\boldsymbol{\theta}|X_{-k}) d\boldsymbol{\theta} \\ = E_{\boldsymbol{\theta}|X_{-k}} \left[\Lambda_k^{n_{S_k}} \exp\{-\Lambda_k |S_k|\} \right] \\ = \frac{1}{T} \sum_{t=1}^T \left[(\Lambda_k^{(t)})^{n_{S_k}} \exp\{-\Lambda_k^{(t)} |S_k|\} \right] \\ = \frac{1}{T} \sum_{t=1}^T \left[\left(r_k \exp\{w_k^{(t)} + \beta^{(t)} \mathbf{z}_k\} \right)^{n_{S_k}} \exp \left\{ -r_k \exp\{w_k^{(t)} + \beta^{(t)} \mathbf{z}_k\} |S_k| \right\} \right].$$

Ao observar o resultado, nota-se que $\beta^{(t)}$ é obtido diretamente do modelo, onde (t) refere-se a iteração t da distribuição a posteriori, porém existe um problema para w_k , dado

que a região não foi observada, não se têm estimativas para w_k . Porém, pelas propriedades do Processo Gaussiano e que \mathbf{w} tem como distribuição conjunta uma normal multivariada, pode-se utilizar as propriedades da condicional da normal para isso. Logo:

1. Obtém-se um $\mathbf{w}_{-k}^{(t)}$, valor da *a posteriori* para regiões observadas.
2. Gerar um valor de $w_k | \mathbf{w}_{-k}^{(t)}$.

Com esses valores se torna possível a obtenção da cadeia *a posteriori* da distribuição preditiva.

Exemplo

Na prática, ao observar-se um conjunto de dados como o apresentado na Figura 1, ao considerar um modelo teórico na forma discretizada como:

$$\begin{aligned} X &\sim PP(\Lambda(\cdot)), \\ \Lambda(s_j) &= r_k \lambda_k, \forall s_j \in S_k, \\ \log \lambda_k &= w_k + \beta z_k, \\ \mathbf{w} &\sim N_p(\mu \mathbf{1}, \tau^{-1} R_\phi), \end{aligned}$$

supondo que as observações são localizações de óbitos por câncer de pulmão e z , a variável espacial, é tal que controla o nível de poluição na região k , por exemplo, obtém-se uma verossimilhança da forma:

$$\begin{aligned} L(\Lambda(\cdot), \{s_j\}_{i=1}^n) &\propto [r_{[1]} \exp\{w_{[1]} + \beta z_{[1]}\}]^2 \times [r_{[2]} \exp\{w_{[2]} + \beta z_{[2]}\}]^3 \\ &\times [r_{[3]} \exp\{w_{[3]} + \beta z_{[3]}\}]^3 \times [r_{[4]} \exp\{w_{[4]} + \beta z_{[4]}\}]^1 \\ &\times [r_{[5]} \exp\{w_{[5]} + \beta z_{[5]}\}]^4 \times [r_{[7]} \exp\{w_{[7]} + \beta z_{[7]}\}]^1 \\ &\times [r_{[8]} \exp\{w_{[8]} + \beta z_{[8]}\}]^3 \times [r_{[9]} \exp\{w_{[9]} + \beta z_{[9]}\}]^2 \\ &\times \exp \left\{ - \sum_{k=1}^9 r_k \exp\{w_k + \beta z_k\} |S_k| \right\}, \end{aligned}$$

assim, as informações da região que são as intensidades de cada uma, geradas pelas covariáveis específicas de cada região e o intercepto, são multiplicadas o mesmo número de vezes que de ocorrência, portanto, cada intensidade é elevada ao número de ocorrências observadas na região.

A seção seguinte, finalmente, apresentará os critérios de comparação de modelos baseados no poder preditivo, que são estudados neste trabalho.

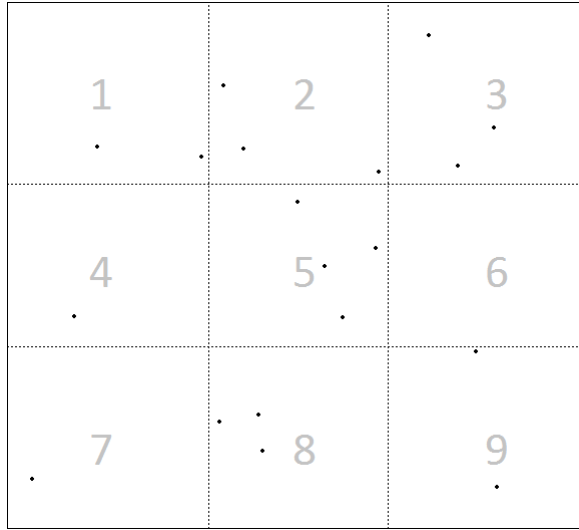


Figura 1: Exemplo de padrões de pontos observados

3.4 Critérios de Comparação

Dado que os modelos e os métodos de estimação foram definidos, o foco é fazer, de fato, a comparação dos modelos. Portanto, agora, serão apresentados os critérios de comparação estudados. Inicialmente foram propostos alguns critérios básicos, posteriormente a partir de uma revisão bibliográfica foram encontrados alguns critérios de comparação com base no poder preditivo já conhecidos na literatura, porém sem uma utilização neste contexto de padrões de pontos.

3.4.1 Critérios Básicos

Os primeiros critérios avaliados foram o desvio quadrático médio das previsões com as reais contagens (C_1), a quantidade de contagens contidas nos seus respectivos intervalos de credibilidade (C_2), a amplitude dos intervalos de credibilidade (C_3) e os desvios quadráticos ponderados pela amplitude dos intervalos (C_4). Os quatro critérios citados acima são definidos por:

- $C_1 = \frac{\sum_{k=1}^K (\hat{n}_{S_k} - n_{S_k})^2}{N}$,
- $C_2 = \#(n_{S_k} \subset IC_{n_{S_k}})$,
- $C_3 = \frac{\sum_{k=1}^K \text{amplitude do } IC_{n_{S_k}}}{N}$,
- $C_4 = \frac{\sum_{k=1}^K \text{amplitude do } IC_{n_{S_k}} (\hat{n}_{S_k} - n_{S_k})^2}{N}$,

em que \hat{n}_{S_k} é a estimativa para as contagens na região k , n_{S_k} é a contagem observada na região k , $IC_{n_{S_k}}$ é o intervalo de credibilidade para n_{S_k} e N é o número total de regiões. Pela definição dos critérios, para C_1, C_3, C_4 o modelo escolhido será aquele com o menor valor obtido com o cálculo dos critérios, ao contrário de C_2 , que deseja-se o maior valor.

Esses critérios levam em conta cálculos vindos diretamente da distribuição preditiva, por isso, tornam-se custosos computacionalmente, devido a necessidade de gerar amostras para distribuição preditiva para cada região. Então, foram buscados outros critérios que contornassem esses problemas.

3.4.2 Critérios WAIC e validação cruzada

Após uma revisão bibliográfica, os critérios, WAIC, LOO e K-fold, apareceram como possíveis alternativas. Com a particularidade de não ser necessário gerar amostras de todas as distribuições preditivas, porque trabalham com estimativas das distribuições preditivas, utilizando as amostras das distribuições *a posteriori* de cada parâmetro. A seguir serão apresentadas as definições de cada critério, como pode-se encontrar em Vehtari e Gelman[2].

WAIC (Watanabe-Akaike)

Suponha y_1, \dots, y_n variáveis independentes que são modeladas por um vetor de parâmetros θ e com sua distribuição *a posteriori* dada por $p(\theta|\mathbf{y})$, sua distribuição preditiva $p(y_i|\mathbf{y})$ e função de verossimilhança $p(y_i|\theta)$. Para calcular o WAIC, primeiramente, precisa-se fazer o processo de estimação dos parâmetros de forma bayesiana, obtendo assim a amostra da distribuição *a posteriori* para os parâmetros. Sendo assim o WAIC será dado por:

$$\text{WAIC} = -2\widehat{\text{elpd}}_{\text{waic}},$$

em que o elpd é uma medida para avaliar a precisão preditiva, sendo assim, para calcular o $\widehat{\text{elpd}}_{\text{waic}}$ utiliza-se a expressão:

$$\widehat{\text{elpd}}_{\text{waic}} = \widehat{\text{lpd}} - \widehat{p}_{\text{waic}}, \quad (3.5)$$

Para $\widehat{\text{lpd}}$ e $\widehat{p}_{\text{waic}}$, os valores exatos a serem calculados são:

$$\begin{aligned}\text{lpd} &= \sum_{i=1}^n \log p(y_i|\mathbf{y}), \\ p_{\text{waic}} &= \sum_{i=1}^n \text{var}(\log p(y_i|\mathbf{y})).\end{aligned}\tag{3.6}$$

Com isso, nota-se que para esse critério escolhe-se aquele modelo que obteve o menor valor de critério.

Na maioria das vezes, o cálculo da distribuição preditiva será inviável analiticamente, como acontece nas distribuições *a posteriori*. Com isso, a partir da definição de distribuição preditiva, nota-se:

$$\begin{aligned}p(y_i|\mathbf{y}) &= \int p(y_i|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \\ &= E_{\boldsymbol{\theta}|\mathbf{y}}[p(y_i|\boldsymbol{\theta})]\end{aligned}$$

Sendo assim, como melhor estimador para esperança de determinada variável, podemos utilizar a média das verossimilhanças obtidas com os valores da cadeia *a posteriori* como estimador para densidade preditiva. O que é de fato utilizado nos cálculos, como é mostrado abaixo:

$$\begin{aligned}p(y_i|\mathbf{y}) &= \frac{1}{T} \sum_{t=1}^T p(y_i|\boldsymbol{\theta}^t), \\ \widehat{\text{lpd}} &= \sum_{i=1}^n \log \left(\frac{1}{T} \sum_{t=1}^T p(y_i|\boldsymbol{\theta}^t) \right), \\ \widehat{p}_{\text{waic}} &= \sum_{i=1}^n V_{t=1}^T \log p(y_i|\boldsymbol{\theta}^t),\end{aligned}$$

em que T é o número total de amostras da distribuição *a posteriori* $p(\boldsymbol{\theta}|\mathbf{y})$ obtidas, e $\boldsymbol{\theta}^t$ é o t -ésimo valor dessa amostra.

Para os modelos Cox Log-Gaussiano, após o processo de estimação dos parâmetros e eles terem sido bem recuperados, pode-se obter o critério de comparação, por:

$$\begin{aligned}\widehat{\text{lpd}} &= \sum_{k=1}^K \log \left(\frac{1}{T} \sum_{t=1}^T p(s_k|\mu^t, \tau^t, \phi^t, \mathbf{w}^t) \right), \\ \widehat{p}_{\text{waic}} &= \sum_{k=1}^K V_{t=1}^T \log p(s_k|\mu^t, \tau^t, \phi^t, \mathbf{w}^t),\end{aligned}$$

com $V(\cdot)$ sendo a variância amostral.

LOO (leave-one-out)

Diferentemente do primeiro critério, neste a avaliação será feita de modo que, ao se estudar o resultado da previsão para uma variável será suposto que ela não participou do processo de estimação. Porém, o modelo será estimado apenas uma vez e para usar essa suposição cada região terá um peso. Outra diferença para o modelo anterior é o fato de que neste critério escolhe-se o modelo que obteve o maior valor de critério. Para obter esse critério, calcula-se:

$$\widehat{\text{elpd}}_{\text{is-loo}} = \sum_{i=1}^n \widehat{\text{elpd}}_{\text{is-loo}_i} = \sum_{i=1}^n \log \left(\frac{\sum_{t=1}^T p(y_i | \boldsymbol{\theta}^t) \tilde{w}_t}{\sum_{t=1}^T \tilde{w}_t} \right),$$

em que o peso \tilde{w} para a amostra t da distribuição *a posteriori*, será calculado como:

$$w_t = \frac{1}{p(y_i | \boldsymbol{\theta}^t)},$$

contudo com os pesos sendo definidos dessa forma, eles podem causar instabilidades, uma vez que estes pesos podem assumir valores muito grandes ou até infinitos, por isso os pesos serão dados por:

$$\tilde{w}_t = \min(w_t, \sqrt{T\bar{w}}).$$

Outra medida importante é o número efetivo de parâmetros que pela relação utilizada acima será escrito como:

$$\hat{p}_{\text{is-loo}} = \widehat{\text{lpd}} - \widehat{\text{elpd}}_{\text{is-loo}},$$

com $\widehat{\text{lpd}}$ podendo ser calculado como em 3.6. Utilizando o critério no contexto de padrões de pontos, especialmente para os modelos Cox Log-Gaussiano utilizados:

$$\widehat{\text{elpd}}_{\text{is-loo}} = \sum_{k=1}^K \log \left(\frac{\sum_{t=1}^T p(s_k | \mu^t, \tau^t, \phi^t, \mathbf{w}^t) \tilde{w}_t}{\sum_{t=1}^T \tilde{w}_t} \right).$$

K-fold

Para este caso, os dados serão divididos em partições e semelhante ao LOO, quando for estudada determinada observação da amostra, será suposto que ela não participou do processo de estimação, aqui, porém, toda a partição a qual ela pertence não participará do processo de estimação. Com a recomendação de que se use poucas partições, sendo 5 ou 10 os valores mais comuns. Diferentemente dos outros dois critérios, neste caso o

processo de estimação ocorrerá tantas vezes quanto o número de partições. Dado todos esses fatores o critério é calculado por:

$$\widehat{\text{elpd}}_{\text{xval}} = \sum_{i=1}^n \widehat{\text{ldp}}_i,$$

$$\widehat{\text{ldp}}_i = \log \left(\frac{1}{T} \sum_{t=1}^T p(y_i | \boldsymbol{\theta}^{m,t}) \right),$$

em que, $\boldsymbol{\theta}^{m,t}$ representa o t -ésimo valor da amostra *a posteriori* para o parâmetro, estimando sem utilizar a partição m a qual i pertence. Como o critério se baseia em somar os valores do log da densidade preditiva, também se escolherá aquele modelo que tiver o maior valor de critério.

Nesta situação, para calcular o número efetivo de parâmetros haverá a mesma relação do critério anterior:

$$\widehat{p}_{\text{xval}} = \widehat{\text{lpd}} - \widehat{\text{elpd}}_{\text{xval}},$$

com $\widehat{\text{lpd}}$ definido em 3.6.

Novamente ao utilizar o critério no contexto observado neste trabalho, para os modelos Cox Log-Gaussiano utilizados:

$$\widehat{\text{elpd}}_{\text{xval}} = \sum_{k=1}^K \log \left(\frac{1}{T} \sum_{t=1}^T p(s_k | \mu^{m,t}, \tau^{m,t}, \phi^{m,t}, \mathbf{w}^{m,t}) \right).$$

Note que, para o cálculo dos critérios definidos anteriormente, será utilizado $p(y_i | \boldsymbol{\theta})$. Neste trabalho, foi utilizado:

$$L(\Lambda_k | \{S_k\}) = \exp \left\{ -r_k \exp\{w_k + \boldsymbol{\beta}' \mathbf{z}_k\} | S_k \right\} \frac{[r_k \exp\{w_k + \boldsymbol{\beta}' \mathbf{z}_k\} | S_k]^{n_{S_k}}}{n_{S_k}!}, \quad (3.7)$$

obtido a partir de,

$$L(\Lambda(\cdot) | \{s_k\}_{k=1}^K) = \exp \left\{ - \sum_{k=1}^K r_k \exp\{w_k + \boldsymbol{\beta}' \mathbf{z}_k\} | S_k \right\} \\ \times \prod_{k=1}^K \frac{[r_k \exp\{w_k + \boldsymbol{\beta}' \mathbf{z}_k\} | S_k]^{n_{S_k}}}{n_{S_k}!}, \quad (3.8)$$

que é gerada pela distribuição de Poisson. Pode-se computar dessa forma pois 3.4, a função de verossimilhança do processo de Cox log-Gaussiano, é proporcional à distribuição de Poisson.

Exemplo de cálculo exato dos critérios

Considere $\{y_1, \dots, y_n\}$ uma amostra de uma variável $Y \sim N(\theta, 1)$, com θ desconhecido. Suponha que deseja-se estimar θ utilizando inferência Bayesiana. Será utilizada uma distribuição *a priori* não-informativa para θ . Neste cenário a função de verossimilhança é dada por:

$$p(\mathbf{y}|\theta) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (y_i - \theta)^2 \right\}, \quad (3.9)$$

e a distribuição *a posteriori* de θ seria dada por:

$$\begin{aligned} p(\theta|\mathbf{y}) &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (y_i - \theta)^2 \right\} \\ &= \exp \left\{ -\frac{1}{2} (n\theta^2 - 2\theta n\bar{y}) \right\} \\ &= \exp \left\{ -\frac{n}{2} (\theta - \bar{y})^2 \right\}. \end{aligned} \quad (3.10)$$

Portanto, a distribuição *a posteriori* de $\theta \sim N(\bar{y}, \frac{1}{n})$. Com base em 3.9 e 3.10, pode-se obter a distribuição preditiva:

$$\begin{aligned} p(y_{n+1}|\mathbf{y}) &= \int p(y_{n+1}|\theta)p(\theta|\mathbf{y})d\theta \\ &\propto \int \exp \left\{ -\frac{1}{2} (y_{n+1} - \theta)^2 \right\} \exp \left\{ -\frac{n}{2} (\theta - \bar{y}) \right\} d\theta \\ &\propto \int \exp \left\{ -\frac{1}{2} (y_{n+1}^2 - 2y_{n+1}\theta + \theta^2 + n(\theta^2 - 2\theta\bar{y} + \bar{y}^2)) \right\} d\theta \\ &\propto \exp \left\{ -\frac{1}{2} y_{n+1}^2 \right\} \int \exp \left\{ -\frac{1}{2} \left(-2\theta[y_{n+1} + \sum_{i=1}^n y_i] + (n+1)\theta^2 \right) \right\} d\theta \\ &\propto \exp \left\{ -\frac{1}{2} y_{n+1}^2 \right\} \int \exp \left\{ -\frac{1}{2} (-2\theta(n+1)\bar{y}_{n+1} + (n+1)\theta^2) \right\} d\theta \\ &\propto \exp \left\{ -\frac{1}{2} (y_{n+1}^2 - (n+1)\bar{y}_{n+1}^2) \right\} \int \exp \left\{ -\frac{(n+1)}{2} (\theta - \bar{y}_{n+1})^2 \right\} d\theta \\ &\propto \exp \left\{ -\frac{1}{2} \left(y_{n+1}^2 - (n+1) \left(\frac{y_{n+1} + n\bar{y}}{n+1} \right)^2 \right) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left(y_{n+1}^2 - (n+1) \left(\frac{y_{n+1}^2 + 2y_{n+1}n\bar{y} + n^2\bar{y}^2}{(n+1)^2} \right) \right) \right\} \end{aligned}$$

$$\begin{aligned} &\propto \exp \left\{ -\frac{1}{2} \left(\frac{n}{n+1} y_{n+1}^2 - \frac{n}{n+1} 2y_{n+1}\bar{y} \right) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \frac{n}{n+1} (y_{n+1} - \bar{y})^2 \right\}, \end{aligned}$$

assim, se conclui que a distribuição preditiva para um novo dado não observado y_{n+1} é normal, $y_{n+1} \sim N(\bar{y}, \frac{n+1}{n})$. Assim, se torna possível calcular as quantidades lpd e p_{WAIC} para se obter, por exemplo, o primeiro critério.

$$\begin{aligned} \text{lpd} &= \sum_{i=1}^n \log p(y_i | \mathbf{y}) \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log\left(\frac{n+1}{n}\right) - \frac{1}{2} \frac{n+1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \\ p_{WAIC} &= \sum_{i=1}^n \text{var}_{post}[\log(p(y_i | \theta))] \\ &= \sum_{i=1}^n \text{var}_{post} \left[-\frac{1}{2} (y_i - \theta)^2 \right] \\ &= \frac{(n-1)}{n} s_y^2 + \frac{1}{2n} \end{aligned}$$

Assim, para o critério WAIC para esse exemplo, bastará fazer $WAIC = -2(\text{lpd} - p_{WAIC})$. Para calcular os outros critérios LOO e K-fold bastará calcular de maneira semelhante.

4 Análise dos Resultados

Neste capítulo, serão apresentados os resultados obtidos no estudo com dados simulados, entre eles os resultados da simulação dos dados, estimação e cálculos de critérios.

4.1 Descrição da Simulação

No estudo de simulação, o objetivo inicial era gerar os dados a partir de um processo conhecido e verificar se o modelo conseguia recuperar bem os valores dos parâmetros após o processo de estimação, para posteriormente utilizar os critérios de comparação nesse modelo.

Inicialmente foi estudado o modelo de Cox log-Gaussiano incluindo uma covariável espacial e o intercepto que varia no espaço, dado por:

$$\begin{aligned} X &\sim PP(\Lambda(\cdot)), \\ \Lambda(s) &= r(s)\lambda(s), \forall s \in S, \\ \log \lambda(s) &= \beta z(s) + w(s), \\ w(\cdot) &\sim PG(\mu, \tau, \rho_\phi). \end{aligned}$$

Entre as inúmeras possibilidades de modelos definidos dessa forma, para realizar as comparações foram definidos três modelos como base, são eles:

- Modelo Correto (MC): Modelo com 1 covariável ($z(\mathbf{s})$), em que $z(\mathbf{s})$ é a covariável utilizada na geração dos dados.
- Modelo Errado 1 (ME1): Modelo com 1 covariável ($u(\mathbf{s})$), uma covariável qualquer associada ao espaço, independente a $z(s)$ e gerada a partir de uma Gama(9,3).
- Modelo Errado 2 (ME2): Modelo com 2 covariáveis ($z(\mathbf{s})$) e ($u(\mathbf{s})$).

Para simular os dados e avaliar o ajuste do mesmo, bem como os critérios apresentados na seção anterior, foi considerada uma discretização do espaço por meio de uma grade regular, considerando 100 regiões. Além disso, inicialmente será considerado que o parâmetro de alcance ϕ é conhecido. Tal parâmetro é de difícil estimação, fato conhecido e relatado na literatura de estatística espacial Liang *et al.*[8], Junior *et al.*[9]. A seguir, notou-se também uma certa instabilidade com o parâmetro de média μ , devido a maneira como é tratado no software utilizado, por isso, a providência tomada foi a mesma do parâmetro de alcance, este parâmetro ficou fixo na hora das estimativas.

O estudo consistiu em estudar duas situações diferentes, uma com poucas observações e outro com mais. Assim, seria possível notar como os critérios se comportariam nas diferentes situações. A primeira simulação gerou dados com aproximadamente 1.000 observações e a segunda dados com 4.000.

Num primeiro momento era interessante descobrir, se por meio dos métodos de estimação os parâmetros estavam sendo bem recuperados, isto é, se as estimativas estavam próximas dos valores verdadeiros dos mesmos. Para o primeiro conjunto de dados, com exatamente 1.466 observações, os valores fixados para os parâmetros foram $\beta = 1,3$, $\mu = 0$, $\tau = 1$ e $\phi = 4,25$ o que gerava uma correlação de 0,05 para a maior distância observada entre as regiões do espaço discretizado. Além disso, $r(\mathbf{s})$ foi gerada de uma Beta(2,3) e $z(\mathbf{s})$ de uma Normal(2, 0,25). O resultado obtido nesta simulação está apresentado abaixo.

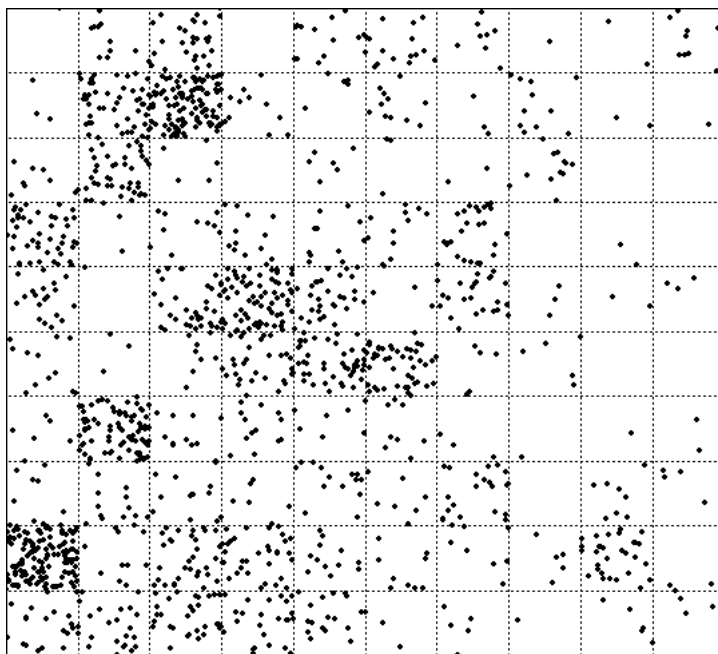


Figura 2: Padrão de Pontos obtido via simulação com $\beta = 1,3$, $\mu = 0$, $\tau = 1$ e $\phi = 4,25$.

Para a segunda ocasião foi utilizado um método semelhante, dessa vez, porém, para alcançar um número de observações de 4.107, os valores fixados para os parâmetros foram $\beta = 1,7$, $\mu = 0$, $\tau = 1$ e $\phi = 4,25$, com as mesmas covariáveis utilizadas na situação anterior.

Ajuste do Modelo

Abaixo são apresentadas as estimativas e os intervalos de credibilidade dos parâmetros e hiperparâmetros do modelo:

	n = 1.466				n = 4.107			
	2,5%	50%	97,5%	real	2,5%	50%	97,5%	real
τ	0,87	1,44	2,31	1	0,71	1,05	1,53	1
β	1,00	1,19	1,38	1,3	1,42	1,59	1,79	1,7

Tabela 1: Estimativa e intervalo de credibilidade dos parâmetros e hiperparâmetros.

A seguir são apresentadas as estimativas versus os valores reais dos interceptos que variam espacialmente, bem como os intervalos de credibilidade tracejados.

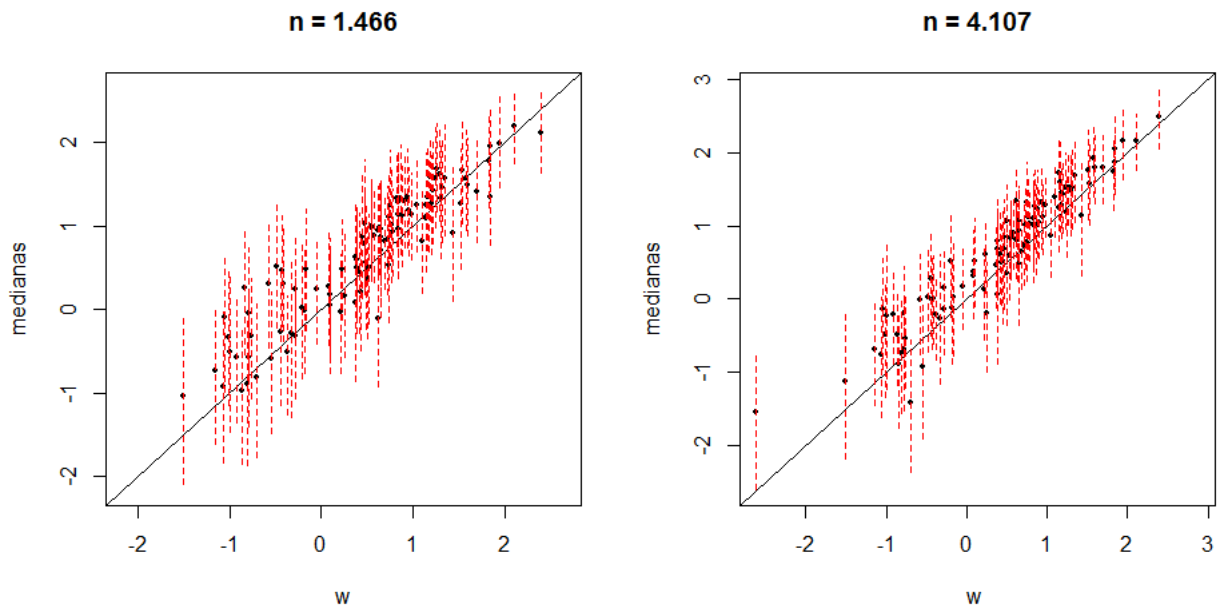


Figura 3: Resultados estimando o Intercepto

As estimativas para os parâmetros e hiperparâmetros estavam próximas dos valores verdadeiros, para as duas situações, e os mesmos sempre estavam contidos nos intervalos de credibilidade. O mesmo pode se afirmar para os interceptos, visto que todos se encontram em volta da reta identidade e nenhum intervalo de credibilidade apresenta muita discrepância. Com os resultados apresentados é possível perceber que o modelo consegue recuperar bem seus parâmetros e hiperparâmetros.

Após checar o ajuste do modelo, partiu-se para a etapa de comparação dos modelos, primeiramente passando pelos critérios mais básicos que lidam diretamente com a distribuição preditiva.

4.2 Comparação com critérios básicos

A seguir são apresentados os resultados obtidos para os critérios C_1 a C_4 considerando os três modelos descritos anteriormente, nas duas situações citadas com 1.466 e 4.107 observações, sendo a mediana *a posteriori* utilizada com estimativa para os parâmetros desejados:

Critérios	Situação					
	n = 1.466			n = 4.107		
	MC	ME1	ME2	MC	ME1	ME2
C_1	51,16	3.035,52	365,36	752,87	73.898,93	4.598,22
C_2	85	81	87	89	90	90
C_3	30,70	167,20	32,34	86,61	902,78	85,94
C_4	4.670,21	4.218.369,15	1.691,15	360.302,1	444.584.422	30.907,94

Tabela 2: Resultados dos critérios básicos utilizando a mediana como estimativas.

Note pela Tabela 2 que o modelo MC mostrou ser o preferido em dois dos quatro critérios estudados para a situação com 1.611 observações, sendo eles C_1 e C_3 . Para a situação com 4.107 observações o resultado foi diferente, por conter a variável certa e o número elevado de observações, o ME2 conseguiu melhorar sua performance e ser escolhido pela maioria dos critérios, exceto por C_1 , o erro quadrático médio.

Com os resultados obtidos pela Tabela 2, pode-se notar que o critério C_1 foi o único que observou o modelo que de fato tinha gerado os dados com precisão. Os outros são mais instáveis, pois não lidam apenas com o valor estimado pela distribuição preditiva, mas também com outras características dessa distribuição.

Após avaliar os critérios mais básicos, utilizou-se os critérios baseados no poder pre-

ditivo: WAIC, LOO e K-fold.

4.3 Comparação com WAIC, LOO e K-fold

Os critérios foram avaliados em cima do conjunto de dados avaliado anteriormente e utilizando os três modelos apresentados. A seguir, então os critérios para cada um desses modelos e os números efetivos de parâmetros são fornecidos nas tabela 3 e 4.

Critérios	Situação					
	n = 1.466			n = 4.107		
	MC	ME1	ME2	MC	ME1	ME2
WAIC	534,04	547,38	536,47	618,75	619,34	619,26
LOO	-274,67	-288,11	-275,91	-321,65	-325,38	-321,75
K-fold	-287,01	-338,86	-287,83	-360,68	-432,57	-360,33

Tabela 3: Resultados dos critérios WAIC, LOO e K-fold para MC, ME1 e ME2

Número efetivo de parâmetros	Situação					
	n = 1.466			n = 4.107		
	MC	ME1	ME2	MC	ME1	ME2
WAIC	38,38	49,85	38,82	44,15	47,54	44,46
LOO	46,03	64,27	46,49	56,43	63,25	56,57
K-fold	58,38	115,01	58,42	95,46	170,44	95,16

Tabela 4: Números efetivos de parâmetros de WAIC, LOO e K-fold para MC, ME1 e ME2

Como se esperava, os critérios apontaram para o modelo certo como sendo o melhor ajuste, para as duas situações. Com exceção do K-fold, para a situação com mais observações, mas a diferença entre o modelo certo e modelo escolhido neste critério ficou pequena.

É interessante observar também, que o modelo ME2 é preferível ao ME1 em todas as situações, a penalização pelo número efetivo de parâmetros não fez que essa ordem se invertesse, o que é bom pois ME2 contém a covariável espacial correta.

Outro ponto interessante para teste desses critérios, é a verificação da sensibilidade deles, quanto a mudanças na estrutura de correlação dos dados. Portanto a partir de

agora, serão incluídos mais dois modelos, ambos contendo apenas a covariável correta $z(s)$:

- Modelo Errado 3 (ME3): Um modelo cuja função de correlação pertence a família exponencial potência e tem parâmetro $\kappa = 0, 5$.
- Modelo Errado 4 (ME4): Um modelo cuja função de correlação pertence a família exponencial potência e tem parâmetro $\kappa = 1, 6$.

Então, ao verificar a sensibilidade dos critérios quando se faz pequenas mudanças nos modelos obteve-se os seguintes resultados:

Critérios	Situação					
	n = 1.466			n = 4.107		
	MC	ME3	ME4	MC	ME3	ME4
WAIC	534,04	529,78	545,69	618,75	5.469,67	5.328,16
LOO	-274,67	-274,20	-279,08	-321,65	-2.125,15	-2.040,58
K-fold	-287,01	-292,51	-287,61	-360,68	-379,70	-450,16

Tabela 5: Resultados dos critérios WAIC, LOO e K-fold para MC, ME1 e ME2

Número efetivo de parâmetros	Situação					
	n = 1.466			n = 4.107		
	MC	ME3	ME4	MC	ME3	ME4
WAIC	38,38	39,47	38,15	44,15	1.704,28	1.616,58
LOO	46,03	48,78	44,39	56,43	1.094,59	993,07
K-fold	58,38	67,09	52,93	95,46	-650,86	-597,35

Tabela 6: Números efetivos de parâmetros de WAIC, LOO e K-fold para MC, ME1 e ME2

Nota-se que quando o número de observações é pequeno, os critérios, em geral, tem dificuldade de escolher o modelo MC como preferível, devido ao baixo número de observações, que de fato, torna mais difícil perceber essa diferença. Quando se olha para o caso com 4.107 observações, observa-se que o MC foi escolhido por todos os critérios, o que comprova que com mais observações é mais fácil de notar qualquer diferença com relação a estruturas de correlação.

Por fim, nota-se que a diferença no número de observações, para efeitos de comparação, foi muito relevante no caso em que não havia grandes diferenças entre os modelos, quando o número de observações aumentou, o modelo correto se diferenciou bastante dos outros dois. Quando os modelos tinham uma diferença maior, nos casos de ME1 e ME2, os resultados foram como esperado tanto para a situação com 1.466 observações, quanto para com 4.107.

5 Conclusão

Inicialmente o objetivo do trabalho era estudar e entender como se comportam critérios baseados no poder preditivo no contexto de padrões de pontos. Portanto, foram estudados critérios que eram calculados diretamente da distribuição preditiva, como C_1 , C_2 , C_3 e C_4 . Posteriormente, outros critérios existentes na literatura foram estudados, sendo estes WAIC, LOO e K-fold, baseados no poder preditivo, porém com cálculos feitos, não necessariamente, em cima da distribuição preditiva.

A partir do estudo de simulação, pode-se observar que os parâmetros do modelo correto estavam sendo bem recuperados, com suas estimativas próximas dos valores originais e com os intervalos de credibilidade relativamente pequenos, com seus tamanhos variando de parâmetro em parâmetro de acordo com sua característica e dificuldade de estimação. Com isso, foi possível a sequência do trabalho, tendo os resultados das comparações mais confiáveis, não sendo influenciados pela qualidade do ajuste.

Em seguida, com o estudo sobre os critérios mais básicos, percebe-se um comportamento irregular, exceto pelo erro quadrático médio, muito utilizado em outros contexto de avaliação, aqui aplicado diretamente na distribuição preditiva. Os outros critérios, dependem da dispersão da distribuição preditiva, dependendo do modelo e das distribuições *a priori*, podem tornar mais dispersas ou mais concentradas. Por exemplo, para o C_2 amplitudes de intervalos de credibilidades exageradamente grandes, podem tornar mais provável de conter o valor verdadeiro das contagens.

Contudo, mesmo para C_1 que teve resultados bons, o custo para o cálculo deste critério é muito grande. Como é necessário gerar as distribuições preditivas para todas as 100 regiões leva tempo. A geração dessas cadeias preditivas se torna mais demorado ainda quando estão presentes no modelo os parâmetros de alcance e a média, então na prática é quase impraticável utilizar qualquer desses critérios.

Quando se trata dos outros critérios, WAIC, LOO e K-fold, os resultados são consideravelmente melhores. Em sua maioria escolheram o MC como melhor modelo, quando

não a diferença foi mínima, tornando os modelos quase equivalentes. E se tratando da pequena diferença entre os modelos, no caso na estrutura de correlação os resultados não foram muito satisfatório para um baixo número de observações, mas quando esse número cresceu o resultado voltou a ser eficiente.

Além de escolher MC como melhor modelo, os critérios conseguiram também escolher ME2, que continha o processo gerador dos dados, como melhor que o ME1 que continha uma covariável espacial totalmente diferente. Ao se fazer uma escala dos modelos escolhidos, ficariam MC, em seguida ME2 e por último ME1. Quando se fala da mudança da estrutura de correlação é complicado, pois depende da característica da função que se supõe, mais suave ou mais aguda.

Dos problemas trabalhados, seria ainda interessante estudar o comportamento dos critérios quando estima-se os parâmetros de média e de alcance, porque estes dificultam a estimação do modelo. Outro ponto a se trabalhar, seria a utilização do programa Stan Gelman *et al.*[10], que facilitaria o emprego da inferência Bayesiana e contém métodos implementados para o cálculo dos critérios WAIC, LOO e K-fold. Por último, empregar as técnicas de comparação estudadas para decidir um melhor modelo para se ajustar a um banco de dados real.

Referências

- [1] MØLLER, J.; SYVERSVEEN, A. R.; WAAGEPETERSEN, R. P. Log gaussian cox processes. *Scandinavian journal of statistics*, Wiley Online Library, v. 25, n. 3, p. 451–482, 1998.
- [2] VEHTARI, A.; GELMAN, A. Waic and cross-validation in stan. *Submitted*. http://www.stat.columbia.edu/~gelman/research/unpublished/waic_stan.pdf Accessed, Citeseer, v. 27, n. 2015, p. 5, 2014.
- [3] GELFAND, A. E.; SMITH, A. F. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, Taylor & Francis Group, v. 85, n. 410, p. 398–409, 1990.
- [4] METROPOLIS, N. et al. Equation of state calculations by fast computing machines. *The journal of chemical physics*, AIP Publishing, v. 21, n. 6, p. 1087–1092, 1953.
- [5] HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, Biometrika Trust, v. 57, n. 1, p. 97–109, 1970.
- [6] REIS, E. *Modelos dinâmicos Bayesianos para processos pontuais espaço-temporais*. Tese (Doutorado) — Ph. D. thesis, Universidade Federal do Rio de Janeiro, 2008.
- [7] BENES, V. et al. *Bayesian analysis of log Gaussian processes for disease mapping*. [S.l.], 2002.
- [8] LIANG, S.; CARLIN, B. P.; GELFAND, A. E. Analysis of minnesota colon and rectum cancer point patterns with spatial and nonspatial covariate information. *The annals of applied statistics*, NIH Public Access, v. 3, n. 3, p. 943, 2008.
- [9] JUNIOR, P. et al. Point pattern analysis with spatially varying covariate effects, applied to the study of cerebrovascular deaths. *Statistics in medicine*, Wiley Online Library, v. 34, n. 7, p. 1214–1226, 2015.
- [10] GELMAN, A.; LEE, D.; GUO, J. Stan a probabilistic programming language for bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, SAGE Publications, p. 1076998615606113, 2015.
- [11] THOMAS, A.; SPIEGELHALTER, D. J.; GILKS, W. Bugs: A program to perform bayesian inference using gibbs sampling. *Bayesian statistics*, v. 4, n. 9, p. 837–842, 1992.