

Renan Xavier Cortes

**Estimando modelos dinâmicos utilizando o  
INLA para Campos Aleatórios Markovianos Não  
Gaussianos**

Brasil

Janeiro de 2014

Renan Xavier Cortes

# **Estimando modelos dinâmicos utilizando o INLA para Campos Aleatórios Markovianos Não Gaussianos**

Dissertação apresentada ao curso de Mestrado da Universidade Federal de Minas Gerais - UFMG, como requisito para a obtenção do título de Mestre em Estatística.

Orientador: Marcos Oliveira Prates

Universidade Federal de Minas Gerais

Instituto de Ciências Exatas

Programa de Pós-Graduação em Estatística

Brasil

Janeiro de 2014

---

Renan Xavier Cortes

Estimando modelos dinâmicos utilizando o INLA para Campos Aleatórios Markovianos Não Gaussianos/ Renan Xavier Cortes. – Brasil, Janeiro de 2014-  
68 p. : il. (algumas color.) ; 30 cm.

Dissertação (Mestrado) – Orientador: Marcos Oliveira Prates  
Universidade Federal de Minas Gerais  
Instituto de Ciências Exatas  
Programa de Pós-Graduação em Estatística, Janeiro de 2014.

1. Modelos Dinâmicos. 2. INLA. 3. Inferência Robusta. I. Orientador: Marcos Oliveira Prates. II. Universidade Federal de Minas Gerais. III. Programa de Pós-Graduação em Estatística. IV. Estimando modelos dinâmicos utilizando o INLA para Campos Aleatórios Markovianos Não Gaussianos

---

Renan Xavier Cortes

## **Estimando modelos dinâmicos utilizando o INLA para Campos Aleatórios Markovianos Não Gaussianos**

Dissertação apresentada ao curso de Mestrado da Universidade Federal de Minas Gerais - UFMG, como requisito para a obtenção do título de Mestre em Estatística.

Trabalho aprovado. Brasil, 13 de janeiro de 2014:

---

**Orientador**

Ph. D. Marcos Oliveira Prates

---

**Professor**

Ph. D. Dani Gamerman

---

**Professor**

Ph. D. Glaura da Conceição Franco

---

**Professor**

Ph. D. Bráulio Figueiredo Alves da Silva

Brasil

Janeiro de 2014

*Este trabalho é dedicado à minha família,  
que desde pequeno me deram todo o suporte necessário  
para realizar meus objetivos.*

# Agradecimentos

Em primeiro lugar, o meu maior agradecimento vai para a minha família, em especial aos meus pais, Maria Helena e Luiz Plínio, e ao meu irmão Marcelo por sempre me apoiarem na minha difícil decisão de abrir mão de diversas coisas em Porto Alegre para realizar a prova de admissão do mestrado da UFMG e ir para Belo Horizonte. Sem vocês eu não seria nada e sou eternamente grato por fazer parte desta família. Amo vocês!

Gostaria de deixar registrado também o meu enorme agradecimento ao meu orientador, Marcos Oliveira Prates, por ter se mostrado sempre uma excelente pessoa, inteligente e prestativa. Com certeza você não deixou nada a desejar como orientador, assim como profissional da área de estatística onde tive a oportunidade de trabalhar e acompanhar em projetos paralelos a dissertação. Valeu!

Agradeço muito também a Thiago Guerrera Martins, o qual tive a oportunidade de conhecer e que teve grande contribuição neste trabalho. Obrigado pelos esclarecimentos dos modelos envolvidos na dissertação sejam por e-mails ou por reuniões via skype. Infelizmente, não estive apto para incluir seu nome neste trabalho.

Agradeço aos membros da banca examinadora. Agradeço ao Dani Gamerman por ter aceitado o convite da minha defesa, à Glaura Franco por ter aceitado duplamente o convite de qualificação e defesa e ao Bráulio Figueiredo por ter aceitado o convite e contribuído bastante com seu conhecimento criminalista na aplicação dos dados de homicídio.

Por fim, não poderia deixar de citar no agradecimentos as diversas pessoas que passaram pelo processo do meu mestrado incluindo professores do Departamento de Estatística da UFMG e colegas de aula e de laboratório que compartilharam os conhecimentos acadêmicos, assim como as ansiedades acadêmicas. Obrigado Raquel, Calisto, Fernanda, Luís, Rodrigo, Maurício, Lívia, entre outros...! Guardarei somente boas recordações!

# Resumo

Modelos de Espaço de Estados, também referidos como Modelos Dinâmicos, constituem uma maneira útil de descrever a evolução de variáveis temporais através de um sistema de evolução latente estruturado. *Integrated Nested Laplace Approximation* (INLA) é uma recente abordagem proposta para realizar inferência Bayesiana rápida em Modelos Gaussianos Latentes que englobam Modelos Dinâmicos. A princípio, a abordagem INLA é limitada à estimações onde o campo latente é assumido Gaussiano o que inviabiliza a estimação de Modelos Dinâmicos onde o ruído do sistema latente seja não-Gaussiano. O objetivo deste trabalho é descrever a metodologia INLA, Modelos Dinâmicos, e como contornar o problema descrito apresentando uma maneira de estimação utilizando o INLA para Modelos Dinâmicos Robustos quando o campo latente segue uma distribuição t-Student. Estudos de simulação para diversos cenários foram conduzidos enaltecendo a importância desta abordagem robusta quando a série temporal apresenta o que é conhecido na literatura como *Outliers Inovativos*. Por fim, duas aplicações foram conduzidas ilustrando os modelos apresentados; a primeira aplicação em dados anuais de homicídio de municípios brasileiros e a segunda em dados mensais de casos de dengue no estado de Minas Gerais.

**Palavras-chaves:** Inferência Bayesiana. INLA. Modelos Dinâmicos. Outliers Inovativos

# Abstract

State-space models, also referred as Dynamic Models, is a useful way to describe the evolution of a time series variable through a structured latent evolution system. Integrated Nested Laplace Approximation (INLA) is a recent approach proposed to perform fast Bayesian inference in Latent Gaussian Models which naturally comprises Dynamic Models. Originally, the INLA approach is restricted to perform estimates where the Latent Field is assumed to be Gaussian distributed, which make not possible the estimation of Dynamic Models assuming a non-Gaussian distribution for the latent system noise. The objective of this work is to describe the INLA methodology, Dynamic Models, and how to overcome this issue presenting a way to use INLA for Robust Dynamic Models assuming a Student-t Random Field. Simulations under several scenarios were conducted highlighting the importance of this robust approach when time series present what is called in the literature as *Innovative Outliers*. At last, two application were conducted exemplifying the presented models; the first application is on annual homicide data of brazilian cities and the second on monthly data of dengue fever of the brazilian state of Minas Gerais.

**Key-words:** Bayesian Inference. INLA. Dynamic Models. Innovative Outliers



# Lista de ilustrações

Figura 1 – Exemplo de exploração da conjunta dos hiperparâmetros . . . . .	20
Figura 2 – Simulação de um passeio aleatório . . . . .	29
Figura 3 – Estimções do Modelo Dinâmico Gaussiano . . . . .	29
Figura 4 – Estimções do Modelo Dinâmico Poisson . . . . .	30
Figura 5 – Exemplo de Outlier Inovativo . . . . .	38
Figura 6 – Mediana das eficiências para magnitude $f = 2$ (esquerda), $f = 4$ (centro) and $f = 8$ (direita), $n = 100$ (linha sólida), $n = 250$ (linha traçada) e $n = 500$ (linha pontilhada). O eixo x representa a percentagem de contaminação esperada e a mediana da eficiência no eixo y. . . . .	43
Figura 7 – Topo: Mediana de RDIC, no eixo y, para todos cenários; Abaixo: Mediana de RPsBF, no eixo y, para todos cenários. O percentual de contaminação esperada se encontra no eixo x e todos os cenários são: $f = 2$ (esquerda), $f = 4$ (centro) and $f = 8$ (direita), $n = 100$ (linha sólida), $n = 250$ (linha traçada) e $n = 500$ (linha pontilhada). . . . .	45
Figura 8 – Tendência das capitais e vizinhos das regiões Brasileiras . . . . .	49
Figura 9 – Tendências da primeira divisão para cada grupo . . . . .	49
Figura 10 – Tendências da segunda divisão para cada grupo . . . . .	51
Figura 11 – Número de casos de todas as cidades de Minas Gerais de Janeiro de 2008 à Dezembro de 2012 . . . . .	53
Figura 12 – Topo: Ajuste de M3 para Belo Horizonte (Gaussiano e Gaussiano), no eixo y, número de casos; Centro: Preditor Linear com 95% de intervalo de credibilidade; Abaixo: anos de 2011 e 2012 com os dados reais e ajustados. . . . .	55
Figura 13 – Estimativa do modelo Gaussiano conjunto para as 10 cidades mais populosas de Minas Gerais. A linha traçada vermelha representa o exponencial do preditor linear. . . . .	56
Figura 14 – Estimativas Gaussianas de C1, sem efeito espacial, e C2 para as 10 cidades mais populosas de Minas Gerais. As linhas vermelhas e verdes representam as estimativas para C1, sem efeito espacial, e C2, respectivamente. A linha vertical separa os dados de 2012. . . . .	59

# Lista de tabelas

Tabela 1 – Medidas de qualidade da primeira divisão. G1: São Paulo e Rio de Janeiro; G2: Belo Horizonte, Recife, Vitória e Porto Alegre; G3: Demais 21 capitais . . . . .	51
Tabela 2 – Critério do Pseudo Bayes Factor . . . . .	52
Tabela 3 – Medidas a posteriori do $\nu$ . G1: São Paulo e Rio de Janeiro; G2: Belo Horizonte, Recife, Vitória e Porto Alegre; G3: Demais 21 capitais . . . . .	52
Tabela 4 – Medidas de ajuste das 10 cidades mais populosas . . . . .	57
Tabela 5 – Medidas de ajuste dos modelos conjuntos estimados com as 10 cidades mais populosas . . . . .	58
Tabela 6 – Medidas de ajuste e predição dos modelos estimados com base em todos municípios . . . . .	59
Tabela 7 – Número de cidades presentes em cada modelo. As linhas representam o critério de qualidade de ajuste (DIC); as colunas representam o critério de predição (MSE). . . . .	60
Tabela 8 – Tempo (em minutos) para cada modelo . . . . .	61

# Sumário

<b>1</b>	<b>Introdução</b>	<b>11</b>
<b>2</b>	<b>Abordagem INLA</b>	<b>13</b>
2.1	Introdução	13
2.2	Modelos Gaussianos Latentes	13
2.3	Campos Aleatórios Markovianos Gaussianos	15
2.4	Metodologia INLA	18
2.4.0.1	Aproximação Gaussiana	20
2.4.0.2	Aproximação de Laplace	21
2.4.0.3	Aproximação de Laplace Simplificada	21
2.5	INLA para Campos Aleatórios Aproximadamente Gaussianos	23
<b>3</b>	<b>Modelos Dinâmicos</b>	<b>26</b>
3.1	Introdução	26
3.2	Modelos Dinâmicos Não-Gaussianos	27
3.3	Modelos Dinâmicos Utilizando o R-INLA	28
<b>4</b>	<b>Contribuição</b>	<b>31</b>
4.1	Estimação de Modelos Dinâmicos no R-INLA: Framework	31
4.2	Utilizando o R-INLA para Modelos Dinâmicos Não-Gaussianos	36
4.3	Escolha da priori para $\nu$	39
<b>5</b>	<b>Simulações</b>	<b>42</b>
<b>6</b>	<b>Aplicação</b>	<b>46</b>
6.1	Estudo de Homicídios	46
6.2	Estudo de Casos de Dengue	52
<b>7</b>	<b>Discussões</b>	<b>62</b>
	<b>Referências</b>	<b>64</b>

# 1 Introdução

Modelos de espaço de estados, também conhecidos como modelos dinâmicos, representam uma classe de modelos que permite a variação temporal tanto dos parâmetros quanto dos dados observados através de um sistema não-observado (latente). Eles encontram aplicações no ajuste e previsão de modelos de regressão e séries temporais (para maiores detalhes ver [West e Harrison \(1997\)](#)). *Integrated Nested Laplace Approximation* (INLA) é uma recente abordagem proposta por [Rue, Martino e Chopin \(2009\)](#) para inferência Bayesiana em modelos Gaussianos latentes onde o foco é nas distribuições marginais a posteriori. Uma descrição detalhada do método INLA e uma comparação com os algoritmos *Markov Chain Monte Carlo* (MCMC) podem ser encontrados em [Rue, Martino e Chopin \(2009\)](#).

A principal vantagem do método INLA é computacional. Enquanto um algoritmo de simulação estocástica das distribuições a posteriori, como MCMC, pode levar horas ou até mesmo dias, a metodologia do INLA pode reproduzir um resultado muito aproximado em apenas alguns segundos ou minutos. Isto se deve ao fato de que a metodologia substitui a simulação por um procedimento de aproximação Gaussiana das densidades condicionais a posteriori usufruindo das propriedades markovianas do campo latente Gaussiano. No entanto, a principal limitação desta metodologia é da exigência de que as variáveis do campo latente sejam todas Gaussianas. Recentemente, [Martins e Rue \(2012\)](#) propuseram uma alternativa para esta limitação, realizando uma aproximação Gaussiana para a parte não-Gaussiana e corrigindo-a com um termo de correção na verossimilhança do modelo.

Modelos dinâmicos lineares e modelos dinâmicos lineares generalizados são uma ampla classe de modelos que generaliza modelos de regressão e de séries temporais permitindo a flexibilidade de que tanto os parâmetros latentes envolvidos quanto os dados observados são descritos de maneira estruturada evolutiva ([MIGON et al., 2005](#)). Modelos dinâmicos são compostos por uma equação observacional e um, ou mais, sistema de equações no qual usualmente os termos de erros são escolhidos seguindo uma distribuição Gaussiana. No entanto, é sabido que a distribuição Gaussiana é sensível a outliers, o que pode produzir degradação na performance das estimações ([FOX, 1972](#)). Portanto, podemos estar interessados em construir um modelo mais flexível, como por exemplo, baseado em distribuições com caudas mais pesadas ao invés da usual distribuição Gaussiana. Tais modelos caem na classe de Modelos Dinâmicos Não-Gaussianos e uma descrição detalhada, assim como aplicações, pode ser encontrada em [Kitagawa \(1987\)](#) e [Durbin e Koopman \(2000\)](#).

---

O objetivo desta dissertação é avaliar o desempenho da metodologia INLA na estimação de modelos dinâmicos quando alguma parte das variáveis que compõem o campo latente não são necessariamente Gaussianas. O capítulo 2 versa sobre a metodologia INLA e a extensão para campos latentes não-Gaussianos. No capítulo 3, é apresentada a classe de modelos dinâmicos de maneira abrangente. No capítulo 4, está presente a contribuição metodológica em como realizar inferência Bayesiana aproximada usando o INLA para modelos dinâmicos não-Gaussianos. O capítulo 5 apresenta os estudos de simulação. O penúltimo capítulo, 6, é composto pelos estudos de aplicação a dados reais. Por fim, o último capítulo, 7, finaliza este trabalho com as discussões.

## 2 Abordagem INLA

### 2.1 Introdução

INLA (do inglês, *Integrated Nested Laplace Approximation*) é uma recente abordagem computacional introduzida por [Rue, Martino e Chopin \(2009\)](#) para realizar inferência Bayesiana em campos latentes Gaussianos, mais precisamente nos casos em que é assumido independência condicional, em uma subclasse de modelos estruturados aditivos chamados de Campos Aleatórios Markovianos Gaussianos (do inglês *Gaussian Markov Random Fields*), doravante referenciado como CAMG. A principal diferença desta abordagem com relação às abordagens Bayesianas tradicionais é a de que esta metodologia não necessita de simulações estocásticas das distribuições marginais a posteriori como os métodos MCMC. Esta abordagem substitui as simulações com aproximações determinísticas, precisas e, sobretudo, de maneira rápida na classe de modelos dos CAMG. Este capítulo tem como objetivo apresentar este método determinístico de realizar inferências Bayesianas através de uma discussão objetiva e exemplificação desta abordagem.

A qualidade do ajuste com esta abordagem é extremamente alta, tal que mesmo para, muitos casos, longas simulações de MCMC não apresentaram diferenças significativas do INLA. A principal vantagem é a agilidade computacional que é muito superior quando comparado com os métodos tradicionais mesmo para um campo latente com alta dimensão. Outra vantagem é a de que, devido a sua natureza de aproximação analítica, o INLA não sofre dos conhecidos problemas de convergência dos métodos MCMC discutidos, por exemplo, em [Gelman e Rubin \(1992\)](#), [Raftery e Lewis \(1992\)](#), [Gamerman e Lopes \(2006\)](#).

Antes de discutir sobre a metodologia INLA, apresentaremos a classe de Modelos Gaussianos Latentes e CAMG em que ela desempenha as inferências.

### 2.2 Modelos Gaussianos Latentes

Formalmente, modelos Gaussianos latentes são uma subclasse de um modelo de regressão aditiva estruturada muito flexível e extensivamente utilizada. Nesta classe, a variável dependente  $y_i$  é assumida pertencer a uma família de distribuições (não necessariamente da família exponencial) onde a média  $\mu_i$  é ligada a um preditor estruturado aditivo  $\eta_i$  através de uma função de ligação  $g(\cdot)$ , assim  $g(\mu_i) = \eta_i$ . O preditor  $\eta_i$  pode se

relacionar com efeitos de covariáveis da seguinte maneira aditiva:

$$\eta_i = \alpha + \sum_{j=1}^{n_f} f^{(j)}(u_{ji}) + \sum_{k=1}^{\eta_\beta} \beta_k z_{ki} + \epsilon_i, \quad (2.1)$$

onde  $\{f^{(j)}(\cdot)\}$ 's são funções desconhecidas das covariáveis  $\mathbf{u}$ , os  $\{\beta_k\}$ 's representam o efeito linear das covariáveis  $\mathbf{z}$  e os  $\{\epsilon_i\}$ 's são os termos não estruturados. Observe que, nesta representação, as covariáveis  $\mathbf{z}$  representam um caso particular das funções desconhecidas das covariáveis  $\mathbf{u}$ . Neste modelo, todas as variáveis latentes, isto é  $\alpha$ ,  $\{f^{(j)}(\cdot)\}$ ,  $\{\beta_k\}$  e  $\{\epsilon_i\}$  são assumidas Gaussianas.

Modelos Latentes Gaussianos possuem uma numerosa e ampla lista de aplicações. Em modelos lineares generalizados Bayesianos o modelo corresponde ao preditor linear  $\eta_i = \alpha + \sum_{k=1}^{\eta_\beta} \beta_k z_{ki}$  em que os termos  $f^{(j)}(\cdot)$  da equação (2.1) podem ser utilizados para modelar algum efeito não-linear das covariáveis ou introduzir efeitos aleatórios ou ambos. Em modelos dinâmicos, a dependência temporal pode ser introduzida fazendo o indexador  $i$  na equação (2.1) como indexador de tempo  $t$  e definindo  $f^{(j)}(\cdot)$  e uma covariável  $\mathbf{u}$  como  $f^{(j)}(u_t) = f_t^{(j)}$ , modelando assim um modelo auto-regressivo a tempo discreto ou contínuo ou até mesmo algum processo latente de séries temporais estruturado. Em modelos espaciais a dependência espacial pode ser modelada usando a covariável espacial  $\mathbf{u}$  como sendo  $f^{(j)}(u_s) = f_s^{(j)}$ , onde  $s$  representa a localização espacial ou região  $s$ .

Para facilitar a discussão posterior, é conveniente reescrever o modelo (2.1) como uma estrutura hierárquica. De uma maneira simplificada, o primeiro nível é formado pela função de verossimilhança com as propriedades de independência condicional dado o campo latente  $\mathbf{x} = (\boldsymbol{\eta}, \alpha, \mathbf{f}, \boldsymbol{\beta})$  e possíveis hiperparâmetros  $\boldsymbol{\theta}_1$ , onde cada observação  $\{y_i, i = 1, \dots, n_d\}$  é conectada com um elemento do campo latente  $x_i$ . Assumindo que os elementos do campo latente conectados às observações estão nos primeiros  $n_d$  elementos de  $\mathbf{x}$ , temos

$$\text{Nível 1 : } \mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}_1 \sim \pi(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}_1) = \prod_{i=1}^{n_d} \pi(y_i \mid x_i, \boldsymbol{\theta}_1). \quad (2.2)$$

No segundo nível, a distribuição das variáveis do campo latente condicionada a outros hiperparâmetros  $\boldsymbol{\theta}_2$  é assumida ter distribuição Gaussiana multivariada:

$$\text{Nível 2 : } \mathbf{x} \mid \boldsymbol{\theta}_2 \sim \pi(\mathbf{x} \mid \boldsymbol{\theta}_2) = \mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta}_2)), \quad (2.3)$$

onde  $\mathcal{N}(\cdot; \boldsymbol{\mu}, \mathbf{Q}^{-1})$  representa uma distribuição Gaussiana multivariada com vetor de médias  $\boldsymbol{\mu}$  e matriz de precisão  $\mathbf{Q}$ . O campo latente  $\mathbf{x}$  pode possuir algumas restrições do tipo  $\mathbf{Ax} = \mathbf{e}$ , para uma matriz  $\mathbf{A}$  com  $k$  linhas e  $n$  colunas com rank  $k$ , onde  $k$  é o número de restrições e  $n$  é o tamanho do campo latente. O último nível da estrutura hierárquica é a designação de uma priori para os hiperparâmetros  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ :

$$\text{Nível 3 : } \boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}).$$

## 2.3 Campos Aleatórios Markovianos Gaussianos

Resumidamente, um CAMG é um vetor aleatório de dimensão finita que segue uma distribuição Gaussiana multivariada. No entanto, um CAMG satisfaz uma suposição de *independência condicional*, o que justifica o termo *Markoviano*. De maneira prática, um CAMG é o segundo estágio da estrutura da modelagem hierárquica de Modelo Gaussiano Latente em que a matriz de precisão  $\mathbf{Q}(\boldsymbol{\theta}_2)$  possui uma estrutura específica. Antes de definirmos um CAMG, precisamos conhecer o conceito de grafos não-direcionados para representação da estrutura de dependência de um CAMG, assim como algumas propriedades de uma distribuição Gaussiana multivariada.

**Definição 1.** (*Grafos não direcionados*) Um grafo não direcionado  $G$  é uma dupla  $G = (V, E)$ , onde  $V$  é conjunto de nós (ou elementos) em um grafo, e  $E$  é o conjunto de ligações  $\{i, j\}$ , onde  $i, j \in V$ , então existe uma ligação não direcionada entre os nós  $i$  e  $j$ . Caso contrário, não há ligação entre os elementos  $i$  e  $j$ . Um grafo é completamente conectado se  $\{i, j\} \in E$  para todo  $i, j \in V$  com  $i \neq j$ .

**Definição 2.** Um vetor aleatório  $\mathbf{x} = (x_1, \dots, x_n)^T$  tem distribuição Gaussiana multivariada com média  $\boldsymbol{\mu}$  e matriz de covariância  $\boldsymbol{\Sigma}$ , denotado por  $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , se a sua densidade é do tipo

$$\pi(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad \mathbf{x} \in \mathbb{R}^n$$

Se o vetor  $\mathbf{x}$  é dividido em duas partes, isto é,  $\mathbf{x} = (\mathbf{x}_A^T, \mathbf{x}_B^T)^T$  e dividirmos seu vetor de médias e matriz de covariância da forma

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix} \quad e \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{pmatrix}$$

isto implica em algumas propriedades como:

P1.  $\mathbf{x}_A \sim N(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA})$ .

P2.  $\boldsymbol{\Sigma}_{AB} = 0$ , se e somente se  $\mathbf{x}_A$  e  $\mathbf{x}_B$  são independentes.

P3. A distribuição condicional de  $(\mathbf{x}_A | \mathbf{x}_B)$  é  $N(\boldsymbol{\mu}_{A|B}, \boldsymbol{\Sigma}_{A|B})$  onde

$$\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB} \boldsymbol{\Sigma}_{BB}^{-1} (\mathbf{x}_B - \boldsymbol{\mu}_B)$$

$$\boldsymbol{\Sigma}_{A|B} = \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB} \boldsymbol{\Sigma}_{BB}^{-1} \boldsymbol{\Sigma}_{BA}$$

**Notação 1.** Para  $C \in I = \{1, \dots, n\}$ , defina  $\mathbf{x}_C = \{x_i : i \in C\}$ . Com  $-C$  denota-se o conjunto  $I - C$ , tal que  $\mathbf{x}_{-C} = \{x_i : i \in -C\}$ . Para dois conjuntos  $A$  e  $B$ , então  $A \setminus B = \{i : i \in A \text{ e } i \notin B\}$ .



Podemos agora definir formalmente um CAMG:

**Definição 3.** (*Campos Aleatórios Markovianos Gaussianos*) O vetor aleatório  $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n$  é um CAMG com relação ao grafo  $G = (V, E)$  com média  $\boldsymbol{\mu}$  e matriz de precisão  $\mathbf{Q} = \boldsymbol{\Sigma}^{-1} > 0$ , se e somente se sua densidade tem a forma

$$\pi(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} |\mathbf{Q}|^{1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{Q}(\mathbf{x} - \boldsymbol{\mu})\right)$$

e

$$\mathbf{Q}_{ij} \neq 0 \Leftrightarrow \{i, j\} \in E \forall i \neq j.$$

Olhando para esta definição a conclusão que chegamos é de que uma distribuição Gaussiana multivariada é um CAMG e vice-versa. No entanto, estamos interessados no particular caso em que esta distribuição possui propriedades de independência condicional de algum conjunto de variáveis implicando, assim, em uma matriz  $\mathbf{Q}$  esparsa, de modo que os algoritmos utilizados para aproximação de CAMG são beneficiados por esta propriedade e, conseqüentemente, aumentando a velocidade da metodologia INLA. O conceito de independência é um poderoso conceito a ser entendido, pois ele é a chave para o desempenho da metodologia INLA. Seja  $\mathbf{x} = (x_1, x_2, x_3)^T$  um vetor aleatório, então  $x_1$  e  $x_2$  são condicionalmente independentes, se para um dado  $x_3$ , observar  $x_2$  não fornece nenhuma informação adicional sobre  $x_1$ . Desta forma, a representação da distribuição conjunta do vetor  $\mathbf{x}$  pode ser escrita como

$$\pi(\mathbf{x}) = \pi(x_1 | x_3)\pi(x_2 | x_3)\pi(x_3),$$

que, por sua vez, é uma simplificação da representação geral

$$\pi(\mathbf{x}) = \pi(x_1 | x_2, x_3)\pi(x_2 | x_3)\pi(x_3).$$

Isto é, esta propriedade implica  $\pi(x_1 | x_2, x_3) = \pi(x_1 | x_3)$  o que facilita a interpretação e representação.

Como exemplo de um CAMG, considere um processo auto-regressivo de primeira ordem com erro Gaussiano

$$x_t = \phi x_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} N(0, 1), \quad |\phi| < 1. \quad (2.4)$$

As suposições de independência condicional não estão dispostas explicitamente, mas ficam mais explícitas se expressarmos na forma condicional

$$x_t | x_1, \dots, x_{t-1} \sim N(\phi x_{t-1}, 1),$$

para  $t = 2, \dots, n$ . Assumindo que a distribuição de  $x_1$  é Gaussiana de média zero e variância  $1/(1 - \phi^2)$ , então a densidade conjunta de  $\mathbf{x}$  é dada por

$$\pi(\mathbf{x}) = \pi(x_1) \pi(x_2 | x_1) \dots \pi(x_n | x_{n-1}) = \frac{1}{(2\pi)^{n/2}} |\mathbf{Q}|^{1/2} \exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{Q} \mathbf{x}\right),$$

onde a *matriz de precisão*  $\mathbf{Q}$  é uma matriz esparsa da forma

$$\mathbf{Q} = \begin{pmatrix} 1 & -\phi & & & & & \\ -\phi & 1 + \phi^2 & -\phi & & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & & -\phi & 1 + \phi^2 & -\phi \\ & & & & & -\phi & 1 \end{pmatrix}.$$

A relação entre independência condicional e a estrutura da matriz de precisão  $\mathbf{Q}$  não é tão clara na estrutura da matriz de covariância  $\mathbf{\Sigma} = \mathbf{Q}^{-1}$ , que é uma matriz densa. Portanto, é difícil de derivar as propriedades de independência condicional a partir da matriz de covariâncias, pois ela fornece apenas a informação sobre dependência *marginal*, e não dependências condicionais. Por exemplo, para  $n = 7$ , a matriz  $\mathbf{\Sigma}$  do modelo (2.4) será

$$\mathbf{\Sigma} = \frac{1}{1 - \phi^2} \begin{pmatrix} 1 & \phi & \phi^2 & \phi^3 & \phi^4 & \phi^5 & \phi^6 \\ \phi & 1 & \phi & \phi^2 & \phi^3 & \phi^4 & \phi^5 \\ \phi^2 & \phi & 1 & \phi & \phi^2 & \phi^3 & \phi^4 \\ \phi^3 & \phi^2 & \phi & 1 & \phi & \phi^2 & \phi^3 \\ \phi^4 & \phi^3 & \phi^2 & \phi & 1 & \phi & \phi^2 \\ \phi^5 & \phi^4 & \phi^3 & \phi^2 & \phi & 1 & \phi \\ \phi^6 & \phi^5 & \phi^4 & \phi^3 & \phi^2 & \phi & 1 \end{pmatrix}.$$

Em um contexto mais geral, em um CAMG estaremos interessados na distribuição de um subconjunto  $\mathbf{x}_A$  de  $\mathbf{x}$  condicionado a  $\mathbf{x}_{-A}$ , onde  $A$  representa um subgrafo de  $\mathbf{x}$ . Neste contexto a parametrização canônica utilizada será útil e um importante resultado é apresentado como teorema.

**Teorema 1.** *Seja  $\mathbf{x}$  um CAMG com relação ao grafo  $G = (V, E)$  com média  $\boldsymbol{\mu}$  e matriz de precisão  $\mathbf{Q} > 0$ . Seja  $A \subset V$  e  $B = V \setminus A$  onde  $A, B \neq \emptyset$ . A distribuição condicional de  $\mathbf{x}_A \mid \mathbf{x}_B$  é também um CAMG com respeito ao subgrafo  $G^A$  com média  $\boldsymbol{\mu}_{A|B}$  e matriz de precisão  $\mathbf{Q}_{A|B} > 0$ , onde*

$$\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A - \mathbf{Q}_{AA}^{-1} \mathbf{Q}_{AB} (\mathbf{x}_B - \boldsymbol{\mu}_B)$$

$$\mathbf{Q}_{A|B} = \mathbf{Q}_{AA}.$$

Este resultado é poderoso por dois motivos. Primeiro, temos conhecimento explícito de  $\mathbf{Q}_{A|B}$  através da matriz principal  $\mathbf{Q}_{AA}$ , então nenhum cálculo é necessário para obtenção da matriz de precisão condicional. Construir o subgrafo  $G^A$  não altera a estrutura, ele apenas remove os nós nas fronteiras do subgrafo  $A$ . Em segundo, uma vez que

$Q_{ij} = 0$  se  $j$  não é vizinho de  $i$ , a média condicional depende somente dos valores de  $\boldsymbol{\mu}$  e  $Q$  que estão em  $A$  ou em seus vizinhos (RUE; HELD, 2005).

No contexto da estimação de modelos dinâmicos, que é o escopo deste trabalho, segundo Migon et al. (2005), os CAMG podem ser vistos como uma generalização de um modelo linear dinâmico para duas ou mais dimensões. Foi mostrado em Lavine (1999) que uma priori de um Campo Aleatório Markoviano em uma grade regular é idêntica a distribuição a posteriori dos parâmetros dos estados de um particular modelo linear dinâmico atualizado com um específico conjunto de dados. Uma das principais áreas em que os modelos CAMG são utilizados é em modelos hierárquicos descrita posteriormente. Neste caso, CAMG formulam a dependência estocástica, temporal, espacial ou espaço-temporal, entre os parâmetros e, desta maneira, implicitamente entre os dados observados. Especificamente quando as variáveis são Gaussianas, a inferência é obtida de maneira bastante direta. No domínio do tempo, esta classe de modelos corresponde portanto aos modelos de espaços de estados. Para uma discussão mais detalhada dos CAMG ver Rue e Held (2005).

## 2.4 Metodologia INLA

O método INLA possui aplicações em modelos hierárquicos Bayesianos em que um vetor  $\boldsymbol{x}$  Gaussiano multivariado de dimensão  $n$  é latente (isto é, não-observado) e possui sua densidade multivariada controlada por um vetor de hiperparâmetros de dimensão  $m$ , onde  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$ . Alguns elementos do vetor aleatório  $\boldsymbol{x}$  são indiretamente observados através dos dados  $\boldsymbol{y}$ , de dimensão  $n_d$ . Os dados são assumidos condicionalmente independentes dado o campo latente  $\boldsymbol{x}$ . No caso mais simples em que cada observação está conectada com apenas um elemento do campo latente, a independência condicional implica  $\pi(\boldsymbol{y} | \boldsymbol{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n_d} \pi(y_i | x_i, \boldsymbol{\theta})$ .

O objetivo principal da abordagem INLA é a aproximação analítica das marginais a posteriori tanto das variáveis latentes quanto dos hiperparâmetros do modelo Gaussiano latente condicionada aos dados observados. Tais marginais podem ser escritas como

$$\pi(x_i | \boldsymbol{y}) = \int \pi(x_i | \boldsymbol{\theta}, \boldsymbol{y}) \pi(\boldsymbol{\theta} | \boldsymbol{y}) d\boldsymbol{\theta}$$

$$\pi(\theta_j | \boldsymbol{y}) = \int \pi(\boldsymbol{\theta} | \boldsymbol{y}) d\boldsymbol{\theta}_{-j}.$$

Sendo  $\tilde{\pi}(\cdot | \cdot)$  a densidade condicional aproximada dos seus argumentos, o objetivo da abordagem é construir aproximações aninhadas da forma

$$\tilde{\pi}(x_i | \boldsymbol{y}) = \int \tilde{\pi}(x_i | \boldsymbol{\theta}, \boldsymbol{y}) \tilde{\pi}(\boldsymbol{\theta} | \boldsymbol{y}) d\boldsymbol{\theta} \quad (2.5)$$

$$\tilde{\pi}(\theta_j | \mathbf{y}) = \int \tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_{-j}.$$

Uma vez que não dispomos da expressão analítica da condicional completa  $\pi(\boldsymbol{\theta} | \mathbf{y})$  e  $\pi(x_i | \boldsymbol{\theta}, \mathbf{y})$  para o cálculo das densidades marginais de interesse, a abordagem INLA faz uso de aproximações analíticas e de integrações numéricas em duas principais etapas, antes das integrações numéricas, que serão descritas a seguir.

Etapa 1 - Aproximando  $\pi(\boldsymbol{\theta} | \mathbf{y})$

Pela teoria da probabilidade, temos a seguinte identidade para distribuições conjuntas

$$\pi(\boldsymbol{\theta} | \mathbf{y}) = \frac{\pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y})}{\pi(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})}.$$

A aproximação da densidade conjunta a posteriori de  $\boldsymbol{\theta}$  é proposta como sendo

$$\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})} \quad (2.6)$$

onde  $\tilde{\pi}_G(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})$  é a aproximação Gaussiana da condicional completa de  $\mathbf{x}$  da forma

$$\pi(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta}) \mathbf{x} + \sum_{i \in \mathcal{I}} g_i(x_i, \boldsymbol{\theta}, y_i) \right\} \quad (2.7)$$

e  $\mathbf{x}^*(\boldsymbol{\theta})$  é a moda da condicional completa de  $\mathbf{x}$ , para um dado  $\boldsymbol{\theta}$ . Neste caso,  $g_i(x_i, \boldsymbol{\theta}, y_i) = \log \{ \pi(y_i | x_i, \boldsymbol{\theta}) \}$  e a moda  $\mathbf{x}^*(\boldsymbol{\theta})$  é obtida através de algum método de otimização como Newton-Raphson. Sabe-se que quanto mais próximo da configuração modal é realizada esta aproximação Gaussiana, melhor é a qualidade deste ajuste na posteriori dos hiperparâmetros. Para uma discussão mais detalhada, ver [Martins \(2010\)](#).

Para isso, não é necessária a representação paramétrica de  $\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y})$ , apenas é realizada uma exploração suficiente em pontos de alta probabilidade para a integração posterior, começando pela configuração modal, para futuras integrações numéricas, isto é, somas finitas. Em [Barron, Schervish e Wasserman \(1999\)](#) o passo-a-passo é descrito em detalhes que resumo a seguir.

Assuma que  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m) \in \mathbb{R}^m$ , isto é, o espaço paramétrico de cada hiperparâmetro está definido na reta real. O que pode ser obtido via reparametrização.

- Passo 1: Localize a moda de  $\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y})$ , otimizando  $\log \{ \tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) \}$ . Seja  $\boldsymbol{\theta}^*$  a configuração modal.
- Passo 2: em  $\boldsymbol{\theta}^*$  compute a matriz Hessiana negativa  $\mathbf{H} > 0$ . Seja  $\boldsymbol{\Sigma} = \mathbf{H}^{-1}$  que seria a matriz de covariância para  $\boldsymbol{\theta}$  se a densidade fosse Gaussiana. Seja  $\boldsymbol{\Sigma} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T$  uma autodecomposição, defina

$$\boldsymbol{\theta}(\mathbf{z}) = \boldsymbol{\theta}^* + \mathbf{V} \mathbf{\Lambda}^{1/2} \mathbf{z}$$

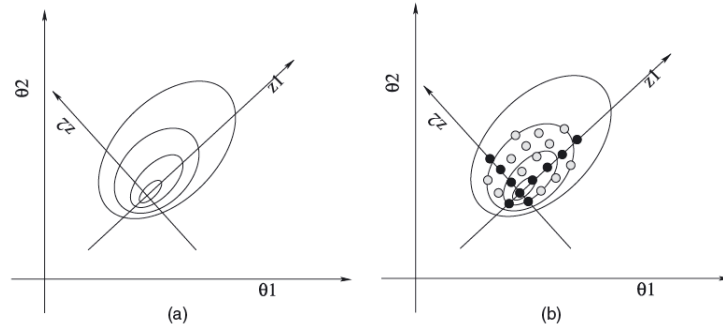


Figura 1 – Exemplo de exploração da conjunta dos hiperparâmetros

- Passo 3: explore  $\log \{\tilde{\pi}(\boldsymbol{\theta} \mid \mathbf{y})\}$  usando a parametrização  $\mathbf{z}$  comece pela moda ( $\mathbf{z} = \mathbf{0}$ ) e ir na direção positiva de  $z_1$  com espaçamento  $\delta_z$  até que

$$\log [\tilde{\pi} \{\boldsymbol{\theta}(\mathbf{0}) \mid \mathbf{y}\}] - \log [\tilde{\pi} \{\boldsymbol{\theta}(\mathbf{z}) \mid \mathbf{y}\}] < \delta_\pi$$

então faça similarmente na direção oposta, gerando os pontos pretos no gráfico, e verificar a condição na combinação destes pontos, gerando os pontos cinzas na Figura 2.1(b).

Etapa 2 - Aproximando  $\pi(x_i \mid \boldsymbol{\theta}, \mathbf{y})$

No primeiro passo, obtivemos um conjunto de pontos ponderados  $\{\theta_k\}$  para serem integrados. Nesta segunda etapa, é proposta uma aproximação da densidade marginal de cada variável do campo latente  $\tilde{\pi}(x_i \mid \theta_k, \mathbf{y})$  fazendo uso da malha obtida na primeira etapa. Tal aproximação é utilizada para a integração numérica que aproxima (2.5):

$$\tilde{\pi}(x_i \mid \mathbf{y}) = \sum_k \tilde{\pi}(x_i \mid \theta_k, \mathbf{y}) \tilde{\pi}(\theta_k \mid \mathbf{y}) \Delta_k. \quad (2.8)$$

Em [Rue, Martino e Chopin \(2009\)](#), três opções são sugeridas para estimação de  $\pi(x_i \mid \theta_k, \mathbf{y})$ . Elas variam em termos computacionais e/ou precisão.

#### 2.4.0.1 Aproximação Gaussiana

A aproximação mais simples e mais rápida é a aproximação Gaussiana  $\tilde{\pi}_G(x_i \mid \theta_k, \mathbf{y})$ , isto é, esta opção consiste apenas no uso da marginal da aproximação Gaussiana do denominador de (2.6). Neste caso, o único custo extra é computar as variâncias marginais da matriz de precisão de  $\tilde{\pi}_G(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})$ . Com esta opção consideramos que

$$\tilde{\pi}_G(x_i \mid \boldsymbol{\theta}, \mathbf{y}) \equiv N(\mu_i(\boldsymbol{\theta}), \sigma_i^2(\boldsymbol{\theta})).$$

A aproximação Gaussiana muitas vezes fornece resultados satisfatórios, no entanto podem existir erros de locação e erros devido à falta de assimetria ([RUE; MARTINO, 2007](#)).

### 2.4.0.2 Aproximação de Laplace

A segunda opção, mais custosa, porém mais precisa, é a aproximação de Laplace

$$\tilde{\pi}_{LA}(x_i | \boldsymbol{\theta}, \mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_{GG}(\mathbf{x}_{-i} | x_i, \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}_{-i} = \mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta})} \quad (2.9)$$

onde  $\mathbf{x}_{-i}$  representa o vetor  $\mathbf{x}$  sem o  $i$ -ésimo elemento,  $\tilde{\pi}_{GG}(\mathbf{x}_{-i} | x_i, \boldsymbol{\theta}, \mathbf{y})$  é a aproximação Gaussiana de  $\mathbf{x}_{-i} | x_i, \boldsymbol{\theta}, \mathbf{y}$  e  $\mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta})$  é a configuração modal de  $\pi(\mathbf{x}_{-i} | x_i, \boldsymbol{\theta}, \mathbf{y})$ . Esta opção é muito custosa, pois é necessário recalcular cada aproximação Gaussiana de  $\tilde{\pi}_{GG}(\mathbf{x}_{-i} | x_i, \boldsymbol{\theta}, \mathbf{y})$  para cada valor de  $x_i$  e de  $\boldsymbol{\theta}$ , uma vez que esta opção requer  $n$  fatorizações da matriz de precisão e achar a sua configuração modal em cada passo. Uma das modificações que o método propõe é aproximar a configuração modal, pelo valor médio

$$\mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta}) \approx E_{\tilde{\pi}_G}(\mathbf{x}_{-i} | x_i).$$

Outra modificação intuitiva construída reside no fato de que somente alguns  $x_j$  que estão “próximos” ao  $x_i$  devem possuir efeito marginal no  $x_i$ . Denotando a região de interesse com relação a distribuição marginal de  $x_i$  por  $R_i(\boldsymbol{\theta})$ . Temos que

$$\frac{E_{\tilde{\pi}_G(x_j|x_i)} - \mu_j(\boldsymbol{\theta})}{\sigma_j(\boldsymbol{\theta})} = a_{ij}(\boldsymbol{\theta}) \frac{x_i - \mu_i(\boldsymbol{\theta})}{\sigma_i(\boldsymbol{\theta})}$$

para algum  $a_{ij}(\boldsymbol{\theta})$ ,  $i \neq j$ . Podendo se considerar uma simples regra para construir  $R_i(\boldsymbol{\theta})$  tal como

$$R_i(\boldsymbol{\theta}) = \{j : |a_{ij}(\boldsymbol{\theta})| > 0,001\}.$$

Com isso, usando a região de interesse  $R_i(\boldsymbol{\theta})$ , apenas será necessário fatorizar matrizes esparsas de dimensão  $|R_i(\boldsymbol{\theta})| \times |R_i(\boldsymbol{\theta})|$ .

### 2.4.0.3 Aproximação de Laplace Simplificada

Uma terceira opção, mais rápida que a de Laplace, porém mais imprecisa, de aproximação de  $\pi(x_i | \boldsymbol{\theta}_k, \mathbf{y})$  denominada Laplace simplificada também é sugerida. Ela é obtida realizando expansões de Taylor de terceira ordem no numerador e primeira ordem no denominador da equação (2.9). Conseguindo assim, contornar os problemas de locação e assimetria da aproximação Gaussiana e ainda tendo um desempenho computacional superior à aproximação de Laplace. Seja

$$x_i^s = \frac{x_i - \mu_i(\boldsymbol{\theta})}{\sigma_i(\boldsymbol{\theta})}$$

defina

$$d_j^{(3)}(x_i, \boldsymbol{\theta}) = \frac{\partial^3}{\partial x_j^3} \log \{ \pi(y_j | x_j, \boldsymbol{\theta}) \} \Big|_{x_j = E_{\tilde{\pi}_G}(x_j | x_i)}$$

assumindo a sua existência.

Expandindo o numerador e denominador da expressão (2.9) em torno de  $x_i = \mu_i(\boldsymbol{\theta})$  até a terceira ordem obtém-se

$$\log \{ \pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y}) \} |_{\mathbf{x}_{-i} = E_{\tilde{\pi}_G(\mathbf{x}_{-i}|x_i)}} = -\frac{1}{2} \left( x_i^{(s)} \right)^2 + \frac{1}{6} \left( x_i^{(s)} \right)^3 \sum_{j \in \mathcal{I} \setminus i} d_j^{(3)} \{ \mu_i(\boldsymbol{\theta}), \boldsymbol{\theta} \} \{ \sigma_j(\boldsymbol{\theta}) a_{ij}(\boldsymbol{\theta}) \}^3 + \dots$$

O primeiro e segundo termo fornece a aproximação Gaussiana, enquanto que o terceiro fornece uma correção para assimetria. Desta forma o denominador da expressão (2.9) se reduz a

$$\log \{ \tilde{\pi}_{GG}(\mathbf{x}_{-i} | x_i, \boldsymbol{\theta}, \mathbf{y}) \} |_{\mathbf{x}_{-i} = E_{\tilde{\pi}_G(\mathbf{x}_{-i}|x_i)}} = \text{constante} + \frac{1}{2} \log |\mathbf{H} + \text{diag} \{ \mathbf{c}(x_i, \boldsymbol{\theta}) \}|$$

onde  $\mathbf{H}$  é a matriz de precisão a priori do CAMG com a  $i$ -ésima linha e coluna deletada e  $\mathbf{c}(x_i, \boldsymbol{\theta})$  é o vetor de menos a segunda derivada do termo da verossimilhança avaliada em  $x_j = E_{\tilde{\pi}_G}(x_j | x_i)$ .

Assim obtemos,

$$\begin{aligned} \log \{ \tilde{\pi}_{GG}(\mathbf{x}_{-i} | x_i, \boldsymbol{\theta}, \mathbf{y}) \} |_{\mathbf{x}_{-i} = E_{\tilde{\pi}_G(\mathbf{x}_{-i}|x_i)}} &= \text{constante} \\ &- \frac{1}{2} x_i^{(s)} \sum_{j \in \mathcal{I} \setminus i} \text{var}_{\tilde{\pi}_G}(x_j | x_i) d_j^{(3)} \{ \mu_i(\boldsymbol{\theta}), \boldsymbol{\theta} \} \sigma_j(\boldsymbol{\theta}) a_{ij}(\boldsymbol{\theta}) + \dots \end{aligned}$$

Note que

$$\text{var}_{\tilde{\pi}_G}(x_j | x_i) = \sigma_j^2(\boldsymbol{\theta}) \{ 1 - \text{corr}_{\tilde{\pi}_G}(x_i, x_j)^2 \}$$

sendo  $\text{corr}_{\tilde{\pi}_G}(x_i, x_j)$  a correlação de  $x_i$  e  $x_j$  sob  $\tilde{\pi}_G$ . Definindo

$$\gamma_i^{(1)}(\boldsymbol{\theta}) = \frac{1}{2} \sum_{j \in \mathcal{I} \setminus i} \sigma_j^{(2)}(\boldsymbol{\theta}) \{ 1 - \text{corr}_{\tilde{\pi}_G}(x_i, x_j)^2 \} d_j^{(3)} \{ \mu_i(\boldsymbol{\theta}), \boldsymbol{\theta} \} \sigma_j(\boldsymbol{\theta}) a_{ij}(\boldsymbol{\theta})$$

e

$$\gamma_i^{(3)}(\boldsymbol{\theta}) = \sum_{j \in \mathcal{I} \setminus i} d_j^{(3)} \{ \mu_i(\boldsymbol{\theta}), \boldsymbol{\theta} \} \{ \sigma_j(\boldsymbol{\theta}) a_{ij}(\boldsymbol{\theta}) \}^3$$

então

$$\log \{ \tilde{\pi}_{SLA}(x_i^{(s)} | \boldsymbol{\theta}, \mathbf{y}) \} = \text{constante} + \frac{1}{2} \left( x_i^{(s)} \right)^2 + \gamma_i^{(1)}(\boldsymbol{\theta}) x_i^{(s)} + \frac{1}{6} \left( x_i^{(s)} \right)^3 \gamma_i^{(3)}(\boldsymbol{\theta}) + \dots$$

no entanto, esta equação não define uma densidade uma vez que o termo de terceira ordem é ilimitado. Uma solução é ajustar uma normal assimétrica (AZZALINI; CAPITANIO, 1999) de modo que a terceira derivada avaliada na moda seja  $\gamma_i^{(3)}$ , a média seja  $\gamma_i^{(1)}$  e a

variância seja igual a 1. Dessa forma,  $\gamma_i^{(3)}$  corrigirá a aproximação Gaussiana na simetria e  $\gamma_i^{(1)}$  ajustará a locação.

Para mais detalhes da aproximação Gaussiana, Laplace e Laplace simplificada ver [Rue, Martino e Chopin \(2009\)](#).

## 2.5 INLA para Campos Aleatórios Aproximadamente Gaussianos

A principal limitação do método INLA é a de que o campo latente descrito em (2.3) deve ser Gaussiano para a metodologia estimar corretamente as densidades a posteriori. Para a condicional completa de  $\mathbf{x}$  ser bem aproximada pelas distribuições Gaussianas descritas em (2.6) e (2.9), o vetor latente deve ser, pelo menos grosseiramente, próximo a uma distribuição Gaussiana multivariada. Esta característica é basicamente garantida através da priori Gaussiana que é estabelecida em (2.3). Além disso, outro aspecto é de que as propriedades de independência condicional encontradas em campos latentes se convertem em problemas de matrizes de precisão esparsas quando o campo é Gaussiano. Isto implica em grande redução computacional quando realizando as aproximações Gaussianas, que é de extrema importância uma vez que ela deve ser computada para cada  $\theta^{(k)}$  na integração numérica da densidade condicional das variáveis latentes descritas em (2.8).

Uma das principais áreas de trabalho que vem sendo desenvolvida é a constante busca de extensões para a metodologia INLA para realizar inferência Bayesiana rápida. A extensão da metodologia INLA para modelos latentes não-Gaussianos também foi discutida em [Resende \(2011\)](#). Recentemente [Martins e Rue \(2012\)](#) propuseram uma metodologia que permite que o campo latente  $\mathbf{x}$  possua alguns componentes independentes não necessariamente Gaussianos. No entanto, tais distribuições deverão ter um comportamento “aproximadamente” Gaussiano, por este motivo é denominada esta classe como sendo campos latentes *aproximadamente* Gaussianos. A intuição mais precisa sobre a utilização deste termo será dada mais além.

O modelo que estaremos interessados terá a mesma estrutura hierárquica descrita na Seção 2.3, no entanto com uma modificação em uma das etapas. Na segunda etapa, redefine-se o campo latente como sendo

$$\text{Nível } 2^{\text{ovo}} : (\mathbf{x}_G, \mathbf{x}_{NG}) \mid \boldsymbol{\theta}_2 \sim \pi(\mathbf{x} \mid \boldsymbol{\theta}_2) = \mathcal{N}(\mathbf{x}_G; \mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta}_2)) \times \prod_i \pi(\mathbf{x}_{NGi} \mid \boldsymbol{\theta}_2),$$

onde  $\mathbf{x}_G$  e  $\mathbf{x}_{NG}$  representam os termos Gaussianos e não-Gaussianos do campo latente, respectivamente. Uma vez que uma parte do campo latente não é Gaussiano, as suposições para a aplicação da metodologia INLA estão violadas e, portanto, não poderiam ser aplicadas sem nenhuma espécie de correção desta violação.

Em [Martins e Rue \(2012\)](#), para o termo “*aproximadamente* Gaussiano” não é dada uma definição formal da estrutura de densidade para a parte do campo latente não-



Gaussiano. O importante é salientar que esta distribuição não deve fugir “muito” das características de uma distribuição Gaussiana, para que estas sejam bem aproximadas o suficiente, ao menos localmente, pela densidade Gaussiana.

Uma abordagem alternativa para resolver esta classe de modelos seria, incluir a parte não-Gaussiana como hiperparâmetros da equação (2.6). No entanto, a maior demanda computacional do método INLA está na exploração e construção da malha de pontos da densidade conjunta dos hiperparâmetros para a aproximação de (2.6) para as aproximações de  $\tilde{\pi}(x_i | \mathbf{y})$  e  $\tilde{\pi}(\theta_j | \mathbf{y})$ . Esta abordagem, apesar de resultar em aproximações mais precisas, acarretaria em uma demanda computacional muito maior o que foge totalmente do escopo da metodologia INLA que sustenta seu principal pilar na *rápida* inferência Bayesiana. A abordagem proposta fornece bons resultados sem a sobrecarga computacional.

A extensão do método leva em consideração a importância da distribuição Gaussiana das priori do modelo latente. A proposta é aproximar as prioris das componentes não-Gaussianas  $\pi(\mathbf{x}_{NG} | \theta_2)$  pela distribuição Gaussiana  $\pi_G(\mathbf{x}_{NG} | \theta_2)$  e corrigir esta aproximação usando um termo de correção

$$CT = \frac{\pi(\mathbf{x}_{NG} | \theta_2)}{\pi_G(\mathbf{x}_{NG} | \theta_2)}$$

no primeiro nível do modelo hierárquico, isto é, na verossimilhança. Conseqüentemente a verossimilhança descrita em (2.2) será multiplicada pelo termo de correção resultando em

$$\prod_{i=1}^{n_d} \pi(y_i | x_i, \theta_1) \times \pi(\mathbf{x}_{NG} | \theta_2) / \pi_G(\mathbf{x}_{NG} | \theta_2).$$

Uma maneira alternativa de escrever o primeiro nível do modelo hierárquico é definir um vetor e respostas  $\mathbf{z}$  que ora assume o valor original e ora assume o termo de correção

$$\text{Nível 1 : } \mathbf{z} | \mathbf{x}, \boldsymbol{\theta} \sim \pi(\mathbf{z} | \mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n_d+k} \pi(z_i | x_i, \boldsymbol{\theta}),$$

$$\pi(z_i | x_i, \boldsymbol{\theta}) = \begin{cases} \pi(y_i | x_i, \boldsymbol{\theta}_1) & 1 \leq i \leq n_d \\ \pi(\mathbf{x}_{NG,i} | \boldsymbol{\theta}_2) / \pi_G(\mathbf{x}_{NG,i} | \boldsymbol{\theta}_2) & n_d < i \leq n_d + k \end{cases}$$

onde  $k$  é o tamanho do campo latente não-Gaussiano. O campo latente que antes era não-Gaussiano, agora passa a ter uma distribuição Gaussiana devido a correção de  $\mathbf{x}_{NG}$

$$\text{Nível 2 : } (\mathbf{x}_G, \mathbf{x}_{NG}) | \boldsymbol{\theta}_2 \sim \pi(\mathbf{x} | \boldsymbol{\theta}_2) = \mathcal{N}(\mathbf{x}_G; \mathbf{0}, Q^{-1}(\boldsymbol{\theta}_2)) \times \pi_G(\mathbf{x}_{NG} | \boldsymbol{\theta}_2).$$

Portanto, agora está viabilizada a utilização da metodologia INLA. O último nível do modelo hierárquico continua sendo formado por uma distribuição a priori dos hiperparâmetros do modelo.

O modelo hierárquico acima agora satisfaz a representação de um Modelo Gaussiano Latente descrito na Seção 2.2 e pode portanto ser estimado utilizando a ferramenta INLA para isto. Martins e Rue (2012) mostraram que o principal impacto desta estratégia ocorre na aproximação Gaussiana da condicional completa do campo latente que agora toma a forma

$$\pi(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta}) \mathbf{x} + \sum_{i=1}^{n_d} g_i(x_i) + \sum_{n_d+1}^{n_d+k} h_i(x_i) \right\} \quad (2.10)$$

onde  $g_i(x_i) = \log \pi(y_i \mid x_i, \boldsymbol{\theta})$  como antes na equação (2.7) e

$$h_i(x_i) = \log CT_i = \log \pi(\mathbf{x}_{NG,i} \mid \boldsymbol{\theta}_2) - \log \pi_G(x_{NG,i} \mid \boldsymbol{\theta}_2).$$

Foi mostrado que a acurácia do INLA nesta situação depende do comportamento de  $h_i(x_i)$  que é diretamente afetada pela distribuição de  $\pi(x_{NG,i} \mid \boldsymbol{\theta}_2)$  dos componentes não-Gaussianos e pela aproximação Gaussiana  $\pi_G(x_{NG,i} \mid \boldsymbol{\theta}_2)$  para esta distribuição não-Gaussiana.

Martins e Rue (2012) mostraram que bons resultados são obtidos quando  $\pi_G(x_{NG,i} \mid \boldsymbol{\theta}_2)$  é escolhida ser com média zero e com uma precisão muito pequena tais que

$$\pi_G(x_{NG,i} \mid \boldsymbol{\theta}_2) \propto \text{Constante}$$

e  $\pi(x_{NG,i} \mid \boldsymbol{\theta}_2)$  possui uma distribuição não muito “distante” de uma distribuição Gaussiana. Neste paper foi introduzido o termo um tanto quanto informal de distribuições “aproximadamente Gaussianas” (do inglês, *near-Gaussian distributions*). Esta “restrição” assegura que o log do termo de correção  $h_i(x_i)$  pode ser bem aproximado por uma distribuição Gaussiana, mantendo assim uma aproximação razoável para a condicional completa do campo latente, como no caso de Modelos Gaussianos Latentes. Neste artigo, os autores não preocuparam-se em definir precisamente o que significa exatamente o termo “aproximadamente Gaussiana” para esta classe de distribuições, apesar de argumentado que estas distribuições podem ser interpretadas como uma classe de distribuições que permitem algum tipo de flexibilidade no entorno da distribuição Gaussiana, permitindo a adição de desvios, como por exemplo em termos de assimetria e curtose. Por exemplo, os autores realizaram estimações utilizando INLA para estimar modelos em que alguns elementos do campo latente possuem distribuição log-gama e t-Student.

## 3 Modelos Dinâmicos

### 3.1 Introdução

Modelos Dinâmicos são uma ampla classe de modelos com variação temporal dos parâmetros, permitindo através de uma maneira probabilística a introdução de equações que regem a evolução dos parâmetros do modelo. Eles são úteis na modelagem de séries temporais e modelos de regressão. A versão mais simples desta classe de modelos foi introduzida por [Harrison e Stevens \(1976\)](#) em que uma série temporal observada univariada  $y_t | \theta_t$  era governada por um sistema de parâmetros  $\theta_t | \mathbf{y}$  em que ambos eram normalmente distribuídos. Posteriormente generalizações foram propostas e discutidas por [Smith \(1979\)](#) e [West, Harrison e Migon \(1985\)](#).

Matematicamente, em um caso simples Gaussiano, esta classe de modelos pode ser escrita como

$$y_t = F_t' X_t + \nu_t, \quad \nu_t \sim N(0, V_t) \quad (3.1)$$

$$X_t = G_t X_{t-1} + \omega_t, \quad \omega_t \sim N(0, W_t) \quad (3.2)$$

onde  $y_t$  é uma sequência no tempo da variável dependente,  $X_t$  é uma sequência de parâmetros de estados do sistema (latentes) que o descrevem localmente. Assume-se que  $y_t$  é condicionalmente independente dado  $X_t$ .  $F_t$  é um vetor de variáveis explicativas e  $G_t$  uma matriz que descreve a evolução dos estados do sistema. A equação (3.2) é a **evolução** ou **sistema de equações**, definindo a evolução do vetor de estados. Este modelo é totalmente especificado pelo conjunto  $\{F_t; G_t; V_t; W_t\}$  e quando essas quantidades são conhecidas, o modelo é linear e os ruídos são Gaussianos,  $X_t$  pode ser estimado analiticamente através do algoritmo do filtro de Kalman (ver, por exemplo, [Harvey \(1989\)](#)).

Um caso particular dessa classe de modelos é quando o par  $\{F, G\}_t$  é constante para todo  $t$  então o modelo é referido como sendo um Modelo Dinâmico Linear de Séries Temporais. No caso mais particular ainda em que a variância das observações e das evoluções dos estados também são constantes para todo  $t$  o modelo é referido como Modelo Dinâmico Linear Constante. Desta forma, este último é caracterizado pelo conjunto  $\{F; G; V; W\}$  que essencialmente inclui todos os modelos de séries temporais lineares clássicos.

Atualmente, uma ampla opção computacional para estimação de modelos dinâmicos esta disposta na literatura para o R. O pacote `sspir` ([DETHLEFSEN C.; CHRISTENSEN, 2012](#)) realiza a estimação de modelos dinâmicos quando a distribuição de pro-

babilidade da variável resposta pertence a família exponencial. Inclui funções do filtro de Kalman e suavização. Quando ambos os ruídos do processo observacional e latente são Gaussianos e com matriz de covariâncias constantes, a estimação pode ser realizada com um algoritmo EM implementado. O pacote `d1m` (PETRIS, 2010), realiza estimação de máxima verossimilhança, filtragem e suavização de Kalman, além de algumas funções baseadas em amostradores estocásticos para modelos dinâmicos lineares Gaussianos. KFAS (HELKSKE, 2010) é um pacote que disponibiliza funções para estimação de máxima verossimilhança, amostragem estocástica, filtragem e suavização de Kalman, predição e simulação de modelos dinâmicos em que o sistema observacional tem distribuição Gaussiana, Poisson ou Binomial. Ele utiliza a abordagem univariada, também chamada de processamento sequencial, o que proporciona uma filtragem e suavização mais rápida e estável do que o usual filtro de Kalman multivariado. Para uma ampla discussão de modelos dinâmicos, ver West e Harrison (1997).

## 3.2 Modelos Dinâmicos Não-Gaussianos

Modelos dinâmicos não-Gaussianos correspondem ao sistema de equações descrito em (3.1) e (3.2) onde tanto os ruídos do sistema observável, quanto os ruídos do sistema não-observável podem ter uma distribuição de probabilidade não-Gaussiana. Foram amplamente discutidos em West (1981) e posteriormente por Kitagawa (1987).

A principal motivação para o estudo desta classe de modelos é a flexibilidade dos modelos de séries temporais a serem comportados. Séries temporais que possuem tanto mudanças abruptas quanto gradual da média não são muito bem apropriadas para a modelagem padrão em que os ruídos do processo de evolução seguem uma distribuição Gaussiana. Um modelo com baixa variância no sistema não consegue captar saltos, assim como um sistema com alta variância apresenta oscilações na tendência indesejáveis. Neste caso, uma distribuição de probabilidade para estes ruídos que comporte de maneira satisfatória os outliers é a melhor opção. Além disso, uma distribuição não-Gaussiana com caudas mais pesadas para os ruídos do sistema observado podem tratar os outliers nos dados de uma maneira mais adequada do que o convencional ruído Gaussiano. Para tratar tais situações dentro do contexto linear Gaussiano é necessário desenvolver um complicado modelo que leva em conta tanto variações bruscas e suaves nos parâmetros e existência de outliers nas observações. Entretanto, se utilizarmos ambas distribuições com caudas mais pesadas para o sistema de equações, podemos lidar com estes problemas com modelos mais simples. Para os estados latentes, nosso principal interesse está em densidades com caudas mais pesadas que possibilitam a modelagem de séries com mudanças estruturais. Modelos dinâmicos não-Gaussianos também foram amplamente discutidos em West, Harrison e Migon (1985), Harvey e Fernandes (1989), Gamerman D. (2013). Em Gamerman D. (2013) um ampla classe de modelos dinâmicos não-Gaussianos é definida onde tanto

os ruídos do sistema observado quanto latente podem seguir distribuição não-Gaussiana, porém ainda mantendo a disponibilidade analítica de funções da verossimilhança marginal e possibilitando a tratabilidade para realizar previsões de longo prazo, valores faltantes e espaçamento irregular da série.

Atualmente, quando o ruído do sistema latente não possui a distribuição Gaussiana, as estimações Bayesianas, na maioria das aplicações, devem recorrer a métodos de simulações estocásticas (DURBIN; KOOPMAN, 2000; WEST, 1981). No entanto, um dos exemplos de uma abordagem aproximativa está presente em Kitagawa (1989), Kitagawa (1990) em que a densidade não-Gaussiana é aproximada por uma mistura de Gaussianas.

### 3.3 Modelos Dinâmicos Utilizando o R-INLA

Nesta seção serão apresentados modelos dinâmicos simulados para ilustração do pacote INLA (referenciado a partir de agora como R-INLA) nesta classe de modelos. Atualmente somente modelos lineares, isto é, com estrutura linear de evolução no sistema latente, são possíveis de se realizar as estimações. Além disso, todas as estimativas dos estados latentes do modelo são estimadas com base em toda a informação disponível (estimativas suavizadas), ou seja todas as posteriores de interesse serão baseadas nas estimativas de  $\pi(x_t | \mathbf{y})$  ao invés da densidade condicional  $\pi(x_t | y_1, \dots, y_t)$ , isto é, com toda informação disponível para as estimativas.

A estrutura de um Modelo Dinâmico Generalizado de primeira ordem, por exemplo, que o R-INLA lida nas estimações é

$$\begin{aligned} y_t | \eta_t &\sim \varphi(y_t | \eta_t, V), \quad t = 1, \dots, n \\ g(\eta_t) &= F_t x_t \\ x_t &= G_t x_{t-1} + w_t, \quad w_t \sim N(0, W), \quad t = 2, \dots, n. \end{aligned}$$

em que a distribuição de  $y_t | \eta_t$  pertence a família exponencial, a estrutura de evolução  $G_t$  deve possuir uma estrutura de evolução linear e o ruído do sistema (ou sistemas) latente deve ser Gaussiano. Para utilizar o R-INLA como ferramenta de estimação apenas algumas específicas estruturas de evolução devem ser consideradas como, por exemplo, um modelo de passeio aleatório de primeira ou segunda ordem.

Seja um modelo, ilustrado na Figura 2, simulado

$$\begin{aligned} y_t &= x_t + \nu_t, & \nu_t &\sim N(0, V), & t &= 1, \dots, 50 \\ x_t &= x_{t-1} + \omega_t, & \omega_t &\sim N(0, W), & t &= 2, \dots, 50 \end{aligned}$$

com  $V = 2$  e  $W = 1$ .

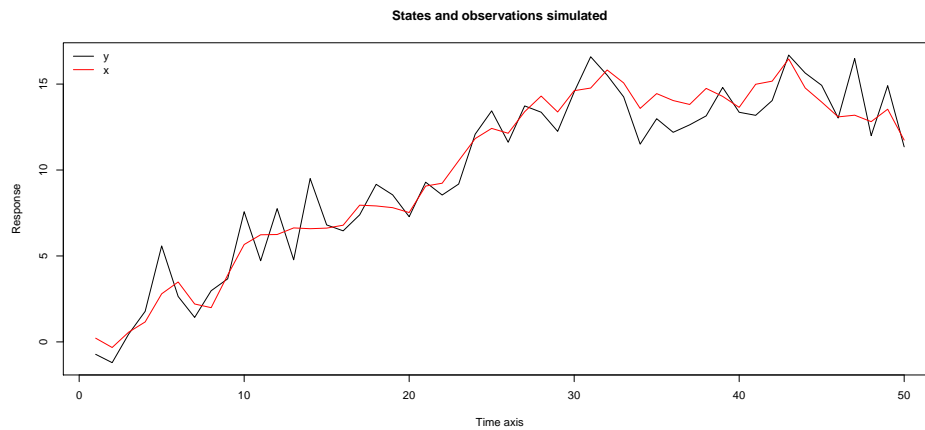


Figura 2 – Simulação de um passeio aleatório

Assumindo na estimação  $x_1 \sim N(0, W)$ ,  $V^{-1} \sim \text{Gama}(2, 4)$  e  $W^{-1} \sim \text{Gama}(1.5, 1.5)$  o código para estimar este modelo no R-INLA é

```
i = 1:n # Índices para o modelo RW1 da biblioteca INLA
formula = y ~ f(i, model="rw1", param=c(1.5, 1.5), constr=F) - 1 # Priori
para precisão do sistema latente
r = inla(formula, data = data.frame(i,y), family="gaussian",
control.family = list(hyper = list(prec = list(prior = "loggamma", param =
c(2,4)))), control.predictor=list(compute=TRUE))
```

Alguns resultado da estimação dos estados latentes  $x_t$  assim como a posteriori dos hiperparâmetros pode ser verificado utilizando o `r$summary.random` e `r$marginals.hyperpar`, respectivamente. A Figura 3 ilustra alguns exemplos gráficos que podem ser gerados com o objeto retornado pela função `inla` através da função `plot`.

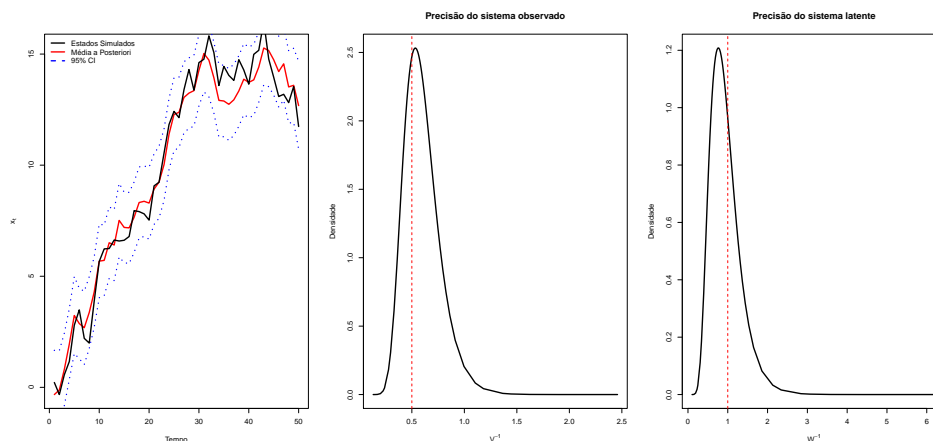


Figura 3 – Estimações do Modelo Dinâmico Gaussiano

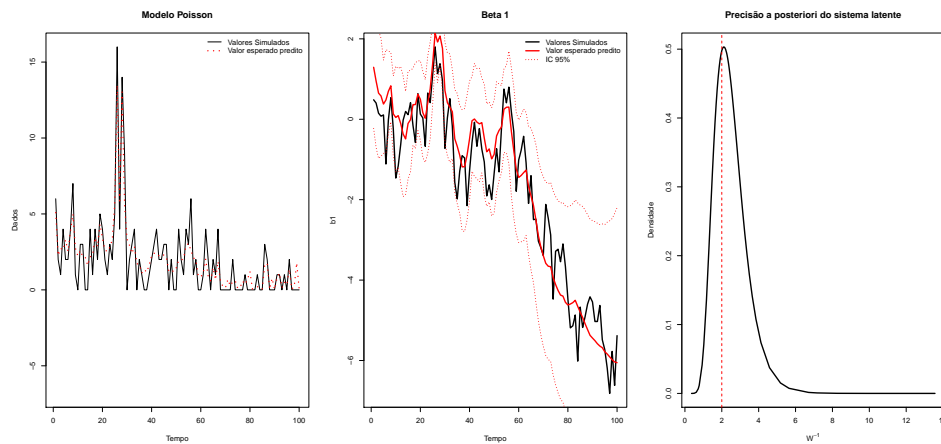


Figura 4 – Estimações do Modelo Dinâmico Poisson

Em um segundo exemplo de modelagem utilizando o R-INLA, assumamos um modelo dinâmico com resposta Poisson com a presença de uma covariável:

$$\begin{aligned}
 y_t \mid \mu_t &\sim \text{Poisson}(\mu_t) \\
 \log(\mu_t) &= \beta_0 + \beta_{1t}X_1 \quad t = 1, \dots, 100 \\
 \beta_{1t} &= \beta_{1,(t-1)} + w_{1t}, \quad w_{1t} \sim N(0, W_1) \quad t = 2, \dots, 100
 \end{aligned}$$

Observe que neste caso não possuímos hiperparâmetros associados à verossimilhança do nosso modelo, apenas a precisão ( $W_1^{-1}$ ) do ruído do parâmetro que acompanha a covariável. Assumindo  $W_1 = 0.5$  na simulação, uma priori vaga para o intercepto e uma priori informativa para a precisão do sistema latente como  $\beta_0 \sim N(0, 1000)$ ,  $W_1 \sim \text{Gamma}(3, 1.5)$  e ainda  $\beta_{11} \sim N(0, 0.5)$  temos uma modelagem utilizando o R-INLA como segue na Figura 4.

O código para a estimação deste modelo segue:

```

index = 1:n # Índices para o Beta 1
formula = y ~ f(index, x1, model="rw1", param=c(3,1.5), constr=F)
r <- inla(formula, family="poisson", data = data.frame(index,y), control.fixed
= list(mean.intercept=0, prec.intercept = 0.001), control.predictor=list(compute=TRUE))

```

## 4 Contribuição

Como visto no Capítulo 2, Modelos Gaussianos Latentes são uma ampla classe que inclui muitos dos modelos utilizados na comunidade aplicada, como por exemplo, modelos de volatilidade estocástica, modelos lineares generalizados e mapeamento de doenças. Apesar de Modelos Dinâmicos se encaixarem naturalmente na classe de Modelos Gaussianos Latentes, antes não era possível ajustar a maioria dos modelos dinâmicos com as ferramentas disponíveis no R-INLA. Ruiz-Cárdenas, Krainski e Rue (2011) apresentaram uma maneira que permite utilizar o R-INLA para realizar inferência Bayesiana em modelos de espaço de estados que não eram antes disponíveis. No entanto, a abordagem deles não inclui os modelos dinâmicos não-Gaussianos onde o erros do sistema latente possuem uma distribuição não-Gaussiana, como por exemplo distribuições com caudas mais pesadas.

Este capítulo tem por objetivo apresentar a metodologia utilizada para a estimação de Modelos Dinâmicos com estrutura de ruídos não necessariamente Gaussianos. Uma vez que o R-INLA já possui implementado diversas verossimilhanças para os dados observados, estaremos interessados em apresentar formalmente como realizar inferência Bayesiana aproximada para Modelos Dinâmicos quando não assumimos a distribuição Gaussiana para o ruído do sistema (ou sistemas) latente. Além disso, este capítulo introduz uma recente abordagem proposta para uma classe de parâmetros específica.

### 4.1 Estimação de Modelos Dinâmicos no R-INLA: Framework

Atualmente, somente uma pequena gama de modelos dinâmicos lineares podem ser estimados diretamente utilizando o R-INLA, tais como modelos auto-regressivos de primeira e segunda ordem. Ruiz-Cárdenas, Krainski e Rue (2011) apresentam como ajustar basicamente qualquer estrutura linear dinâmica Gaussiana pelo R-INLA. É válido destacar que a abordagem INLA poderia ser utilizada para estimar qualquer estrutura dinâmica que pudesse ser escrita como um modelo Gaussiano latente descrito na Seção 2.2, no entanto a abordagem apresentada aqui é motivada para superar algumas limitações do R-INLA.

Neste paper é ilustrada uma abordagem genérica para estimar modelos dinâmicos baseada em uma estrutura aumentada utilizando um simples modelo de passeio aleatório



de primeira ordem como um exemplo ilustrativo:

$$y_t = a_t + v_t, \quad v_t \sim N(0, V), \quad t = 1, \dots, n \quad (4.1)$$

$$a_t = a_{t-1} + w_t, \quad w_t \sim N(0, W), \quad t = 2, \dots, n. \quad (4.2)$$

A chave desta abordagem é igualar a zero, tantos quantos forem, os sistemas de equações latentes. Neste exemplo,

$$0 = a_t - a_{t-1} - w_t, \quad w_t \sim N(0, W), \quad t = 2, \dots, n. \quad (4.3)$$

Então é contruído um modelo alternativo em que adicionamos estas novas “observações zero falsas” de (4.3) nas verdadeiras observações  $\{y_t, t = 1, \dots, n\}$  da equação (4.1). Adicionalmente, eles assumem que estas “falsas observações” seguem uma distribuição Gaussiana com uma precisão muito alta e fixa com média dada por  $\phi_t = a_t - a_{t-1} - w_t$  para representar o fato de que estas observações seriam, na realidade, deterministicamente conhecidas. Neste exemplo, a verossimilhança terá agora dimensão  $n + (n - 1)$  onde os primeiros  $n$  elementos são correspondentes às observações reais e os demais  $n - 1$  elementos serão as observações “falsas”. Todas as demais entradas na matriz de respostas serão preenchidas com valores nulos (NA’s) e o número de colunas aumentadas será igual ao número de equações latentes no modelo dinâmico estimado. A nova estrutura deste modelo resulta em uma matriz de resposta que será fornecida para o R-INLA da seguinte forma:

$$\begin{bmatrix} y_1 & NA \\ \vdots & \vdots \\ y_n & NA \\ NA & 0 \\ \vdots & \vdots \\ NA & 0 \end{bmatrix}.$$

A estimação nesta nova estrutura se dá fornecendo diferentes verossimilhanças para as colunas das observações (duas, neste caso), uma para as  $n$  observações e outra para os  $n - 1$  “falsos zeros”. Desta forma, condicionados aos estados latentes, as observações na primeira coluna são assumidas seguir uma distribuição Gaussiana com precisão  $V^{-1}$  e, condicionados aos valores  $a_t$ ,  $a_{t-1}$  e  $w_t$ , os valores zero na equação (4.3) são deterministicamente conhecidos com variância “nula”.

Para se estimar os estados basta apenas conhecer os ruídos  $w_t, t = \{2, \dots, n\}$  que são os termos desconhecidos no sistema de equações. O grande segredo por trás desta nova formulação é de que não existe informação nenhuma sobre os estados latentes além de única e exclusivamente sua estrutura de evolução temporal. Esta evolução é acessada através de vetores de índices que conectam as observações nos seus respectivos valores

de  $a_t$ ,  $a_{t-1}$  e  $w_t$ . Os termos do  $a_{t-1}$  são modelados como cópias dos termos  $a_t$  através do atributo `copy` do R-INLA, fazendo com que a posição  $k + 1$  do índice do  $a_{t-1}$  seja igual a posição  $k$  do índice do  $a_t$ . Observando que os únicos termos estocásticos deste sistema de equações são os valores de  $w_t$ , espera-se que os estados latentes possam assumir qualquer valor do espaço paramétrico, o que induz a definição do vetor  $\mathbf{a}$  como sendo variáveis aleatórias Gaussianas com uma pequena e uma fixa e baixa precisão. Seja  $\mathbf{y}$  o vetor de respostas obtido, todas estas modificações impactam o código do R-INLA da seguinte forma:

```

m <- n-1

Y <- matrix(NA, n+m, 2)
Y[1:n, 1] <- y # Primeira coluna
Y[1:m + n, 2] <- 0 # Segunda coluna

i <- c(1:n, 2:n) # Índices para a_t
j <- c(rep(NA,n), 2:n -1) # Índices para a_{t-1}
w1 <- c(rep(NA,n), rep(-1,m)) # Pesos para a_{t-1}
l <- c(rep(NA,n), 2:n) # Índices para w_t
w2 <- c(rep(NA,n), rep(-1,m)) # Pesos para w_t

# Formulando o modelo
# Os estados a_t seguem uma distribuição Gaussiana com fixa e baixa log-precisão(-10)
# Os termos w_t seguem Distribuição Gaussiana com variância W desconhecida
(hyperpriori default no exemplo).
formula <- Y ~ f(i, model="iid", initial=-10, fixed=T) + f(j, w1, copy="i")
+ f(l, w2, model = "iid") - 1

# Duas verossimilhanças são fornecidas no argumento das famílias
# Observe que um valor alto e fixo (10) para a log-precisão das observações
falsas
modelo <- inla(formula, data = data.frame(i,j,w1,w2,l),
family = rep("gaussian", 2),
control.family = list(list(), list(initial=10, fixed=T)),
control.predictor=list(compute=TRUE))

```

Pode ser mostrado que esta modificação, quando comparada com a abordagem de modelagem padrão, possui um impacto na matriz de precisão da condicional completa  $\mathbf{y} | V, W$ . Denote por  $V^{-1} = \tau_V$ ,  $W^{-1} = \tau_W$ ,  $\mathbf{x}_{-1} = (x_2, x_3, \dots, x_n)$ ,  $\mathbf{x}_{-n} = (x_1, x_2, \dots, x_{n-1})$ ,  $\mathbf{w} = (w_2, w_3, \dots, w_n)$  e  $\mathbf{v} = (v_2, v_3, \dots, v_n)$ , podemos representar o sistema de equações

composto por (4.1) e (4.3) da seguinte forma matricial

$$\mathbf{y} = \mathbf{x} + \mathbf{v} \quad (4.4)$$

$$\mathbf{x}_{-1} = \mathbf{x}_{-n} + \mathbf{w}. \quad (4.5)$$

A verossimilhança para este modelo é dada por

$$\mathbf{y} \mid \mathbf{x}, \tau_V \propto \exp \left\{ -\frac{\tau_V}{2} (\mathbf{y} - \mathbf{x})^T (\mathbf{y} - \mathbf{x}) \right\}$$

utilizando uma priori seguindo um passeio aleatório de ordem 1 para o vetor  $\mathbf{x}$ , obtemos uma condicional completa  $\mathbf{x} \mid \tau_W \propto \exp \left\{ -\frac{\tau_W}{2} \mathbf{x}^T \mathbf{R} \mathbf{x} \right\}$ , onde esta matriz  $\mathbf{R}$  possui a seguinte estrutura

$$\mathbf{R} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -1 & 1 \end{pmatrix}$$

Então a distribuição conjunta

$$\mathbf{y} \mid \tau_V, \tau_W \propto \int \exp \left\{ -\frac{\tau_V}{2} (\mathbf{y} - \mathbf{x})^T (\mathbf{y} - \mathbf{x}) - \frac{\tau_W}{2} \mathbf{x}^T \mathbf{R} \mathbf{x} \right\} d\mathbf{x}$$

que pode ser expressada, depois de alguns cálculos, como

$$\mathbf{y} \mid \tau_V, \tau_W \propto \exp \left\{ -\frac{1}{2} \mathbf{y}^T (\tau_V \mathbf{I}_n - \tau_V^2 \mathbf{Q}_X^{-1}) \mathbf{y} \right\} d\mathbf{x}$$

com  $\mathbf{I}_n$  sendo uma matriz identidade  $n \times n$  e

$$\mathbf{Q}_X = (\tau_V \mathbf{I}_n + \tau_W \mathbf{R}).$$

Agora, para verificar qual o impacto desta abordagem utilizando esta verossimilhança “aumentada” com observações “falsas” vamos construir a mesma distribuição condicional com esta nova estrutura. Seja  $\mathbf{0}$  um vetor  $(n-1)$ -dimensional de valores nulos e  $\mathbf{d} = (x_2 - x_1, x_3 - x_2, \dots, x_n - x_{n-1})$ , a equação (4.5) pode ser reescrita

$$\mathbf{d} = \mathbf{w} \implies \mathbf{0} = \mathbf{d} - \mathbf{w}.$$

A parte da verossimilhança responsável por estas observações “falsas” será dada por

$$\mathbf{0} \mid \mathbf{d}, \mathbf{w} \propto \exp \left\{ -\frac{\tau_0}{2} (\mathbf{w} - \mathbf{d})^T (\mathbf{w} - \mathbf{d}) \right\},$$

onde  $\tau_0$  representa a alta precisão para a diferença observada “simulando” que os zeros não são estocásticos e foram observados deterministicamente. Assumindo, portanto, prioris Gaussianas para os termos estocásticos que faltam

$$\begin{aligned}\mathbf{x} \mid \tau_X &\propto \exp\left\{-\frac{\tau_X}{2}\mathbf{x}^T\mathbf{x}\right\} \\ \mathbf{w} \mid \tau_W &\propto \exp\left\{-\frac{\tau_W}{2}\mathbf{w}^T\mathbf{w}\right\}\end{aligned}$$

com  $\tau_X$  sendo uma precisão fixa e baixa para os estados  $\mathbf{x}$ . Observe que esta suposição permite que os valores dos estados possam assumir, praticamente, qualquer valor da reta dos reais.

Para obter a distribuição preditiva conjunta de  $\mathbf{y}$ , condicionada aos hiperparâmetros, se integra em relação aos termos estocásticos:

$$\begin{aligned}\mathbf{y} \mid \tau_V, \tau_W, \tau_0, \tau_X &\propto \int \int \exp\left\{-\frac{\tau_V}{2}(\mathbf{y}-\mathbf{x})^T(\mathbf{y}-\mathbf{x})-\frac{\tau_0}{2}(\mathbf{w}-\mathbf{d})^T(\mathbf{w}-\mathbf{d})\right\} \\ &\quad \times \exp\left\{-\frac{\tau_X}{2}\mathbf{x}^T\mathbf{x}-\frac{\tau_W}{2}\mathbf{w}^T\mathbf{w}\right\} d\mathbf{w}d\mathbf{x} \\ &\propto \exp\left\{-\frac{\tau_V}{2}\mathbf{y}^T\mathbf{y}\right\} \int \exp\left\{-\frac{\tau_V}{2}(\mathbf{x}^T\mathbf{x}-2\mathbf{y}^T\mathbf{x})-\frac{\tau_0}{2}\mathbf{d}^T\mathbf{d}-\frac{\tau_X}{2}\mathbf{x}^T\mathbf{x}\right\} \\ &\quad \times \int \exp\left\{-\frac{\tau_0}{2}(\mathbf{w}^T\mathbf{w}-2\mathbf{d}^T\mathbf{w})-\frac{\tau_W}{2}\mathbf{w}^T\mathbf{w}\right\} d\mathbf{w}d\mathbf{x}\end{aligned}$$

onde

$$\int \exp\left\{-\frac{\tau_0}{2}(\mathbf{w}^T\mathbf{w}-2\mathbf{d}^T\mathbf{w})-\frac{\tau_W}{2}\mathbf{w}^T\mathbf{w}\right\} d\mathbf{w} \propto \exp\left\{\frac{\tau_0^2}{2(\tau_0+\tau_W)}\mathbf{d}^T\mathbf{d}\right\}$$

e  $\mathbf{d}^T\mathbf{d} = \mathbf{x}^T\mathbf{R}\mathbf{x}$ . Além disso, temos que  $\tau_0 - \frac{\tau_0^2}{\tau_0+\tau_W} = \frac{\tau_0}{\tau_0+\tau_W}\tau_W$ , implicando que

$$\begin{aligned}\mathbf{y} \mid \tau_V, \tau_W, \tau_0, \tau_X &\propto \exp\left\{-\frac{\tau_V}{2}\mathbf{y}^T\mathbf{y}\right\} \int \exp\left\{-\frac{1}{2}\left[\mathbf{x}^T\left((\tau_V+\tau_X)\mathbf{I}_n+\tau_0\mathbf{R}-\frac{\tau_0^2}{\tau_0+\tau_W}\mathbf{R}\right)\mathbf{x}-2\tau_V\mathbf{y}^T\mathbf{x}\right]\right\} d\mathbf{x} \\ &\propto \exp\left\{-\frac{\tau_V}{2}\mathbf{y}^T\mathbf{y}\right\} \int \exp\left\{-\frac{1}{2}(\mathbf{x}^T\mathbf{Q}_{X^*}\mathbf{x}-2\mathbf{y}^T\mathbf{x})\right\} d\mathbf{x} \\ &\propto \exp\left\{-\frac{1}{2}(\tau_V\mathbf{y}^T\mathbf{y}-\tau_V^2\mathbf{y}^T\mathbf{Q}_{X^*}^{-1}\mathbf{y})\right\} \\ &\propto \exp\left\{-\frac{1}{2}\mathbf{y}^T(\tau_V\mathbf{I}_n-\tau_V^2\mathbf{Q}_{X^*}^{-1})\mathbf{y}\right\}\end{aligned}$$

onde

$$\mathbf{Q}_{X^*}^{-1} = (\tau_V+\tau_X)\mathbf{I}_n + \left(\frac{\tau_0}{\tau_0+\tau_W}\tau_W\right)\mathbf{R}. \quad (4.6)$$

Observando (4.6) nota-se que o impacto de utilizar esta estrutura aumentada se dá em dois fatores na matriz  $\mathbf{Q}_{X^*}^{-1}$  da distribuição conjunta de  $\mathbf{y}$  condicionado aos hiperparâmetros. O primeiro é a mudança de nível de  $\tau_V$  dada pela baixa precisão assumida para o campo latente  $\tau_X$ , a segunda é a redução da precisão  $\tau_W$  pelo fator  $\frac{\tau_0}{(\tau_0+\tau_W)}$ . É trivial verificar que a medida que  $\tau_X \rightarrow 0$  e  $\tau_0 \rightarrow \infty$  implica em  $\mathbf{Q}_{X^*}^{-1} \rightarrow \mathbf{Q}_X$ .

## 4.2 Utilizando o R-INLA para Modelos Dinâmicos Não-Gaussianos

Estamos interessados agora em combinar a metodologia apresentada na Seção 2.5 de utilização do INLA para Campos Aleatórios Aproximadamente Gaussianos utilizando a nova estrutura estudada. O objetivo principal é formalizar o link de ambas estruturas e como podemos utilizar esta nova metodologia para a estimação de modelos dinâmicos quando assumimos uma distribuição diferente da Gaussiana para o ruído do sistema latente usando o R-INLA. Além disto, é importante enfatizar que modelos dinâmicos representam uma classe de modelos em que existe uma estrutura de dependência condicional para os estados latentes. As contribuições a seguir podem ser divididas em dois aspectos: 1) Não estamos necessariamente assumindo modelos independentes como exigido em [Martins e Rue \(2012\)](#), 2) Estamos estendendo [Ruiz-Cárdenas, Krainski e Rue \(2011\)](#) explorando o formato lá apresentado e viabilizando o uso do R-INLA para Modelos Dinâmicos não-Gaussianos, fazendo com que a abordagem deles seja uma subclasse dos modelos apresentados aqui.

Ao invés de assumir que as observações “falsas” de (4.3) seguem uma distribuição Gaussiana com alta e fixa precisão e média dada por  $\phi_t = a_t - a_{t-1} - w_t$ , assumiremos que estas observações artificiais seguem uma distribuição Gaussiana com variância  $W$  e média  $\eta_t^* = a_t - a_{t-1}$ , o que é uma representação equivalente, uma vez que esta modificação facilitará a transição de conceitos quando estendermos esta abordagem para os modelos dinâmicos de interesse.

Para entender melhor qual o impacto dessa modificação, assim como entender como fazer a analogia entre o INLA para Campos Aleatórios Aproximadamente Gaussianos e o framework do R-INLA utilizado em [Ruiz-Cárdenas, Krainski e Rue \(2011\)](#), acreditamos que escrevendo o modelo utilizando a estrutura hierárquica similar definida na Seção 2.2 será mais fácil de entender como essa estrutura afeta a definição do modelo e como tal definição do modelo afeta o INLA. Similarmente, temos um vetor de resposta estendido (ou aumentado) com  $z_i = y_i$  se  $i \leq n$  e  $z_i = 0$  se  $n < i \leq 2n - 1$  e

- **Nível 1.**  $\mathbf{z} \mid \mathbf{x}, \boldsymbol{\theta} \sim \pi(\mathbf{z} \mid \mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^{2n-1} \pi(z_i \mid x_i, \boldsymbol{\theta})$ , onde

$$\pi(z_i \mid x_i, \boldsymbol{\theta}) = \begin{cases} \pi(y_i \mid x_i, V) & \text{para } 1 \leq i \leq n \\ \pi(z_i \mid x_i, W) & \text{para } n < i \leq 2n - 1 \end{cases} \quad (4.7)$$

com  $\pi(y_i \mid x_i, V) \sim \mathcal{N}(y_i; x_i, V)$  e  $\pi(z_i \mid x_i, W) \sim \mathcal{N}(0; x_i, W)$ . Notando que, como mencionado anteriormente, foi utilizada uma distribuição Gaussiana com variância dada por  $W$  para a verossimilhança das observações artificiais. Internamente, para o R-INLA, o campo latente é definido como

- **Nível 2.**  $\mathbf{x} = (\eta_1, \dots, \eta_n, \eta_{n+1}, \dots, \eta_{2n-1}, a_1, \dots, a_n)$

onde  $\mathbf{a}$  é dada por prioris Gaussianas independentes com baixa e fixa precisão,  $\eta_t = a_t + s_t$  é o preditor linear conectado com a observação  $y_t$ , para  $t = 1, \dots, n$  e  $\eta_t = a_t - a_{t-1} + s_t$  é o preditor linear conectado com as observações zeros, para  $t = n + 1, \dots, 2n - 1$ , e  $s_t$  é um pequeno ruído representado por uma distribuição Gaussiana com média zero e alta precisão para eliminar um problema de definição de posto da matriz de precisão de  $\mathbf{x}$  nesta representação. Por fim, prioris são designadas para os hiperparâmetros do modelo

- **Nível 3.**  $V \sim IG(a_\nu, b_\nu)$ ,  $W \sim IG(a_w, b_w)$ .

Agora, analisando a representação hierárquica acima sob o enfoque da estrutura discutida na Seção 2.5 vemos que a abordagem da estrutura aumentada de Ruiz-Cárdenas, Krainski e Rue (2011) é, na realidade, uma aproximação da forma

$$\pi(\mathbf{a}) \propto \text{constante} \quad (4.8)$$

para uma distribuição Gaussiana com matriz de precisão  $Q_a = W^{-1}\mathbf{R}$  onde  $\mathbf{R}$  é dada por

$$\begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -1 & 1 \end{pmatrix}$$

onde esta aproximação é corrigida na verossimilhança através da adição de seguinte termo de correção

$$CT = \prod_{i=n+1}^{2n-1} \pi(z_i | x_i, W)$$

com  $\pi(z_i | x_i, W)$  definida na equação (4.7). Uma vez identificado isto, nota-se que ambas a log-verossimilhança  $g_i(x_i)$  e o log do termo de correção  $h_i(x_i)$  da equação (2.10) possuem formas quadráticas, o que implica que, no contexto de Ruiz-Cárdenas, Krainski e Rue (2011), a condicional completa do campo latente  $\pi(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta})$  é Gaussiana, implicando no fato de que o R-INLA fornece resultados exatos ao menos de um pequeno erro de integração.

Agora, para estender os resultados obtidos de Ruiz-Cárdenas, Krainski e Rue (2011) e Martins e Rue (2012) para ajustar modelos onde o ruído do sistema latente possua uma distribuição não-Gaussiana vamos utilizar o seguinte exemplo ilustrativo

$$y_t = a_t + v_t, \quad v_t \sim N(0, V), \quad t = 1, \dots, n \quad (4.9)$$

$$a_t = a_{t-1} + w_t, \quad w_t \sim t(0, \tau, \nu), \quad t = 2, \dots, n. \quad (4.10)$$

que pode ser visto como a seguinte estrutura hierárquica

$$\begin{aligned} y_t &| a_t, V \sim N(a_t, V) \\ a_t &| a_{t-1}, \tau, \nu \sim t(a_{t-1}, \tau, \nu) \\ \tau &\sim \pi(\tau), \quad \nu \sim \pi(\nu), \quad V \sim \pi(V) \end{aligned}$$

salientando o fato de que o campo latente não é Gaussiano. Note que a distribuição de  $v_t$  em (4.9) poderia possuir uma distribuição não-Gaussiana também, uma vez que funções de verossimilhanças não-Gaussianas já estão implementadas no R-INLA, mas utilizando uma distribuição Gaussiana aqui fará com que o impacto da não-Gaussianidade de  $w_t$  seja mais facilmente visível e analisável. Como mencionado no Capítulo 1, a motivação de se utilizar distribuições com caudas mais pesadas, como a distribuição t-Student, no lugar da usual Gaussiana no ruído do sistema latente é com o objetivo de robustificar o modelo. Neste caso, robustez do modelo significa que o sistema dinâmico é menos sensível a diferentes tipos de outliers e permitindo uma maior variabilidade de  $w_t$  poderíamos acomodar melhor o que são chamados *outliers inovativos* na literatura de séries temporais (FOX, 1972; MASRELIEZ; MARTIN, 1977; QUARRIE; TSAI, 2003). Este tipo de outlier é caracterizado pelo efeito nas observações subsequentes, tal como uma mudança estrutural na média do processo, como pode ser exemplificado na Figura 5.

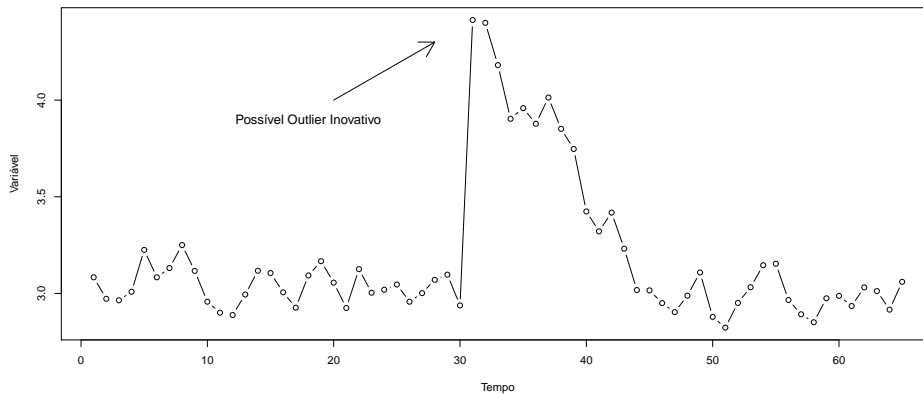


Figura 5 – Exemplo de Outlier Inovativo

Através de um argumento similar feito na Seção 4.1, nota-se que nós somente precisamos saber os termos estocásticos  $\{w, t = 2, \dots, n\}$  e a dinâmica do sistema em (4.10) para estimar  $\mathbf{a} = (a_1, \dots, a_n)$ . Conseqüentemente, se adicionarmos estas informações na função de verossimilhança através do termo de correção, estamos aptos a designar prioris Gaussianas independentes com média zero e baixa e fixa precisão para  $\mathbf{a}$ , que acarretará no seguinte modelo hierárquico

- **Nível 1.**  $z|\mathbf{x}, \boldsymbol{\theta} \sim \pi(z|\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^{2n-1} \pi(z_i|x_i, \boldsymbol{\theta})$ , onde

$$\pi(z_i|x_i, \boldsymbol{\theta}) = \begin{cases} \pi(y_i|x_i, V) & \text{for } 1 \leq i \leq n \\ \pi(z_i|x_i, \tau, \nu) & \text{for } n < i \leq 2n - 1 \end{cases} \quad (4.11)$$

com  $\pi(y_i|x_i, V) \stackrel{d}{=} \mathcal{N}(y_i; x_i, V)$  e  $\pi(z_i|x_i, \tau, \nu) \stackrel{d}{=} t(0; x_i, \tau, \nu)$ .

- **Nível 2.**  $\boldsymbol{x} = (\eta_1, \dots, \eta_n, \eta_{n+1}, \dots, \eta_{2n-1}, a_1, \dots, a_n)$

onde  $\boldsymbol{a}$  e  $\boldsymbol{\eta}$  são os mesmos como definidos anteriormente. Por fim, prioris são designadas aos hiperparâmetros do modelo

- **Nível 3.**  $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$  com  $\boldsymbol{\theta} = (V, \tau, \nu)$

Nota-se que o modelo hierárquico descrito acima é muito similar àquele descrito na Seção 2.2 e a diferença é no termo de correção

$$CT = \prod_{i=n+1}^{2n-1} \pi(z_i|x_i, \tau, \nu) \quad (4.12)$$

que não é mais Gaussiano, acarretando na condicional completa da forma (2.10) com logaritmo do termo de correção não-Gaussiano  $h_i(x_i)$ . Esta configuração se ajusta no modelo da Seção 2.5 e todos os resultados obtidos em Martins e Rue (2012) são ainda válidos no contexto de modelos dinâmicos lineares não-Gaussianos, além de superar a limitação da exigência anterior para somente modelos independentes. Isto significa que a aplicação do R-INLA dentro deste contexto irá produzir resultados acurados desde que os componentes de (4.12) são distribuídos de acordo com uma distribuição flexível no entorno da Gaussiana, como no caso da t-Student.

### 4.3 Escolha da priori para $\nu$

Neste capítulo será introduzida uma recente abordagem proposta por Martins e Rue (2013) para a escolha da priori dos graus de liberdade,  $\nu$ , da distribuição t no estudo de simulação subsequente e em uma das aplicações deste trabalho.

A escolha da priori para os graus de liberdade  $\nu$  da distribuição t se mostra um dos grandes problemas em inferência Bayesiana, uma vez que ele não possui uma interpretação direta para o seu valor, porém deve ser especificado. Alguns trabalhos já foram realizados sobre a questão. Em Fonseca, Ferreira e Migon (2008) é derivada a priori de Jeffreys para os graus de liberdade da distribuição t-Student no contexto de modelos de regressão linear para a distribuição dos erros t e assumindo que tanto, a priori marginal dos efeitos fixos e a priori conjunta do parâmetro de escala e  $\nu$  são independentes. Outros trabalhos com discussões sobre a escolha de priori para  $\nu$  podem ser encontrado em Villa C. (2013) e Fernandez e Steel (1999).



Uma maneira de robustificar e/ou generalizar um modelo contra alguma peculiaridade específica dos dados é generalizar o modelo mais “básico” adicionando um parâmetro específico, doravante referenciado como “parâmetro de flexibilidade”. Na abordagem Bayesiana, *prioris* devem ser designadas para todos os parâmetros de um modelo, incluindo o referido parâmetro. Nesta dissertação, uma vez que as abordagens robusta são avaliadas através de suposição residual através de uma generalização da distribuição Gaussiana, a t-Student, o modelo básico é a distribuição Gaussiana e o flexível é o modelo t, indexado pelos graus de liberdade,  $\nu$ , ilustrados aqui como o “parâmetro de flexibilidade”. Uma das principais motivações por trás da construção de *proris* específicas para esta classe de parâmetros é a falta de interpretabilidade que eles representam na modelagem, devido a sua natureza de generalização.

A abordagem adotada assume que o modelo básico desempenha um papel central dentro do modelo flexível e os hiperparâmetros associados a esta distribuição são os parâmetros que controlam o grau de flexibilidade desejado no entorno do modelo básico. Algumas das vantagens da construção desta *priori* são a *generalidade*, pois a ideia pode se estender para outros parâmetros além do  $\nu$  da distribuição t, *robustez com relação aos hiperparâmetros*, pois as posteriores suportam o modelo básico enquanto podem indicar o modelo mais flexível quando existe evidência amostral para tal, e *interpretabilidade*, pois a *priori* é designada à distância entre o modelo básico e flexível.

O R-INLA parametriza a distribuição t em termos de sua precisão marginal  $\tau$  e graus de liberdade  $\nu$  de maneira que as distribuições Gaussianas e t podem ser representadas pelo mesmo parâmetro de precisão  $\tau$ . Seja  $f(y; \boldsymbol{\phi})$  representar o modelo básico e  $g(y; \boldsymbol{\phi}, \psi)$  representar modelo flexível onde  $\boldsymbol{\phi} = (\mu, \tau)$  e  $\psi = \nu$ , a parametrização Gaussiana utilizando  $\mu$  para a média e  $\tau$  para a precisão é a seguinte

$$f(y; \boldsymbol{\phi}) = \left(\frac{\tau}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2}(y - \mu)^2\right\}.$$

No caso da distribuição t-Student a parametrização de  $\boldsymbol{\phi} = (\mu, \tau)$  e  $\psi = \nu > 2$  se dá da seguinte maneira

$$y = \mu + \tau^{-\frac{1}{2}}x$$

onde  $x \sim T_\nu$  é uma distribuição t-Student padrão com  $\nu$  graus de liberdade com seguinte distribuição de probabilidade

$$g(x; \psi) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) [(\nu-2)\pi]^{\frac{1}{2}}} \left(1 + \frac{x^2}{(\nu-2)}\right)^{-\frac{\nu+1}{2}}.$$

Uma das suposições da distribuição a priori do parâmetro de flexibilidade é de que o modelo básico desempenha um papel central dentro do modelo flexível e tenha valor modal no valor específico que recupera o modelo básico. A outra suposição imposta é de que deve-se supor que modelos próximos do modelo básico possuem mais densidade

a priori quando comparados com os que estão mais longe do modelo básico. Seja  $f(x)$  e  $g(x; \nu)$  sendo as distribuições Gaussiana e t, respectivamente, com média zero e precisão um, a divergência  $d$  das distribuições é definida como

$$d(f(\cdot), g(\cdot; \nu)) = d(\nu)$$

onde a medida escolhida como dissimilaridade é a divergência de Kullback-Leiber ([KULLBACK S., 1951](#)) definida como

$$KL_{\nu}(g, f) = \int g(x; \nu) \log \frac{g(x; \nu)}{f(x)} dx$$

os autores do paper justificam que a escolha desta medida com o argumento de que uma maior densidade a priori é atribuída para  $\nu$  onde é difícil de discriminar  $g(x; \nu)$  e  $f(x)$ . E atribuir menos densidade a priori para  $\nu$  onde temos grande poder de discriminação entre  $g(x; \nu)$  e  $f(x)$ . A intuição disto é de que se existe uma evidência clara nos dados de que existe uma cauda pesada, esta característica estará refletida na distribuição a posteriori indicando que existe informação, mesmo que razoável, para a estimativa de  $\nu$ , apesar da baixa densidade a priori para valores baixos.

A medida de divergência utilizada é uma transformação da distância de Kullback-Leiber com a seguinte estrutura

$$d(\nu) = (2KL_{\nu}(g, f))^{\frac{1}{2}}. \quad (4.13)$$

Um dos motivos para se usar (4.13) são interpretativos. Por exemplo,  $d(25) \approx 0.05$  significa que a distância entre  $g(x; \nu = 25)$  a  $f(x)$  é aproximadamente a mesma distância entre as distribuições  $N(0, 1)$  e  $N(0.05, 1)$ . Ou seja, esta transformação da distância de Kullback-Leiber pode ser vista como parâmetro de localização.

Por fim, a priori utilizada neste contexto é a distribuição exponencial do tipo

$$\pi(d | \lambda) = \lambda \exp(-\lambda d)$$

onde o parâmetro  $\lambda$  desempenha um papel interpretativo fundamental para o grau de flexibilidade que o usuário escolherá, pois se escrevermos

$$\lambda = \frac{-\log(df)}{d(\nu^*)}$$

significa que a priori de  $\nu$ , neste caso da distribuição t-Student, atribui  $(100 \times df)\%$  de probabilidade no intervalo  $(2, \nu^*]$ , onde  $df$  representa o “grau de flexibilidade” (do inglês, “degrees of flexibility”).

Refere-se [Martins e Rue \(2013\)](#) para estudos de simulação realizados, comparações destas prioris com outras já utilizadas na literatura, assim como mais detalhes sobre prioris para parâmetros de flexibilidade.

## 5 Simulações

Neste capítulo é conduzido uma simulação de Monte Carlo para o passeio aleatório de primeira ordem apresentado no Capítulo 4. Foi escolhido realizar um estudo de contaminação similar a [Pinheiro, Liu e Wu \(2001\)](#) e [Martins e Rue \(2012\)](#) para contaminar o ruído  $w_t$  na equação (4.2) com a seguinte mistura de distribuições Gaussianas

$$w_t \sim (1 - p) \times N(0, W) + p \times f \times N(0, W) \quad t = 1, \dots, n$$

onde  $p$  é a percentagem esperada de *outliers inovativos* no sistema latente e  $f$  é um valor fixo indicando a magnitude da contaminação. Foram rodadas 1000 iterações sob 54 cenários com todas as combinações de variações dos parâmetros no processo gerador dos dados para  $n = 100, 250, 500$ ,  $p = 0, 0.05, 0.1, 0.15, 0.20, 0.25$  e  $f = 2, 4, 8$ . O verdadeiro valor da variância de ambos os ruídos é  $V, W = 2$ .

Conforme já comentado, no R-INLA a verossimilhança da  $t$  é parametrizada em termos de sua precisão marginal  $\tau$  e graus de liberdade  $\nu$ . Isto representa uma vantagem porque a precisão sob as distribuições Gaussianas e  $t$  podem ser representadas pelo mesmo parâmetro  $\tau$ , permitindo a mesma priori a ser usada para  $\tau$  seja para referenciar o modelo Gaussiano ou o modelo  $t$ . Neste experimento de Monte Carlo nós utilizamos uma Gama<sup>1</sup> com  $\alpha$  e  $\beta$  dados por 1 e 2,375 tanto para a precisão do sistema observado quanto latente. A priori para  $\nu$  é baseada no framework de [Martins e Rue \(2013\)](#) descrita na Seção 4.3. A especificação da priori consiste na escolha do parâmetro  $df$ ,  $0 < df < 1$ , que representa a percentagem de massa atribuída aos graus de liberdade entre 2 e  $\nu^*$ , onde optou-se por  $df = 0.3$  e  $\nu^* = 10$ .

Espera-se que à medida que a percentagem esperada e a magnitude dos outliers inovativos neste modelo dinâmico aumente, a abordagem proposta resulte em robustez quando comparada com a abordagem Gaussiana, justificando portanto o uso do R-INLA para Campos Aleatórios Aproximadamente Gaussianos no contexto de modelos dinâmicos. Para avaliar estas comparações foram computados os erros quadráticos médios (MSE do inglês *mean squared error*), CPO (do inglês *conditional predictive ordinate*) ([GELFAND; DEY; CHANG, 1992](#); [DEY; CHEN; CHANG, 1997](#)), e DIC (do inglês *Deviance Information Criteria*) ([SPIEGELHALTER et al., 2002](#)). Todas as estimativas da distribuição marginal dos estados latentes foram obtidas através do método de Laplace Simplificado.

Seja  $a_{tj}$  ser o verdadeiro estado latente de interesse, do tempo  $t$  na iteração  $j$  estimado por  $E(a_{tj,G} | \mathbf{y}) = \widehat{a}_{tj,G}$  e  $E(a_{tj,T} | \mathbf{y}) = \widehat{a}_{tj,T}$  onde  $a_{tj,G} | \mathbf{y}$  e  $a_{tj,T} | \mathbf{y}$  representam

<sup>1</sup> se  $X \sim Gama(a, b)$  então  $E(X) = \frac{a}{b}$

as posteriores dos modelos Gaussiano e t, respectivamente. A eficiência do modelo t sobre o Gaussiano para cada iteração  $j$  é dada por

$$E_j = \frac{\sum_{j=1}^n (\widehat{a_{tj,G}} - a_{tj})^2}{\sum_{j=1}^n (\widehat{a_{tj,T}} - a_{tj})^2}$$

que pode ser vista como a razão dos respectivos MSEs.

A Figura 6 representa a mediana sobre as todas as iterações  $\{E_j, j = 1, \dots, 1000\}$  para cada cenário. Os resultados foram conforme esperado. Houve uma pequena melhoria de eficiência para padrões de contaminação menores, enquanto que o ganho de eficiência se torna mais evidente a medida que analisamos padrões de contaminação maiores, alcançando ganhos de eficiência maiores do que 15% para cenários mais críticos. O ganho de eficiência é alto para percentagens de contaminação esperadas moderados, em torno de 10% no caso, e este comportamento não-monotônico pode ser explicado pelo fato que uma vez que os dados se tornam muito contaminados, nem mesmo um modelo mais flexível baseada na distribuição t-Student pode fornecer resultados continuamente melhores quando comparados com o modelo Gaussiano, apesar de que o modelo mais flexível continua a demonstrar uma eficiência maior.

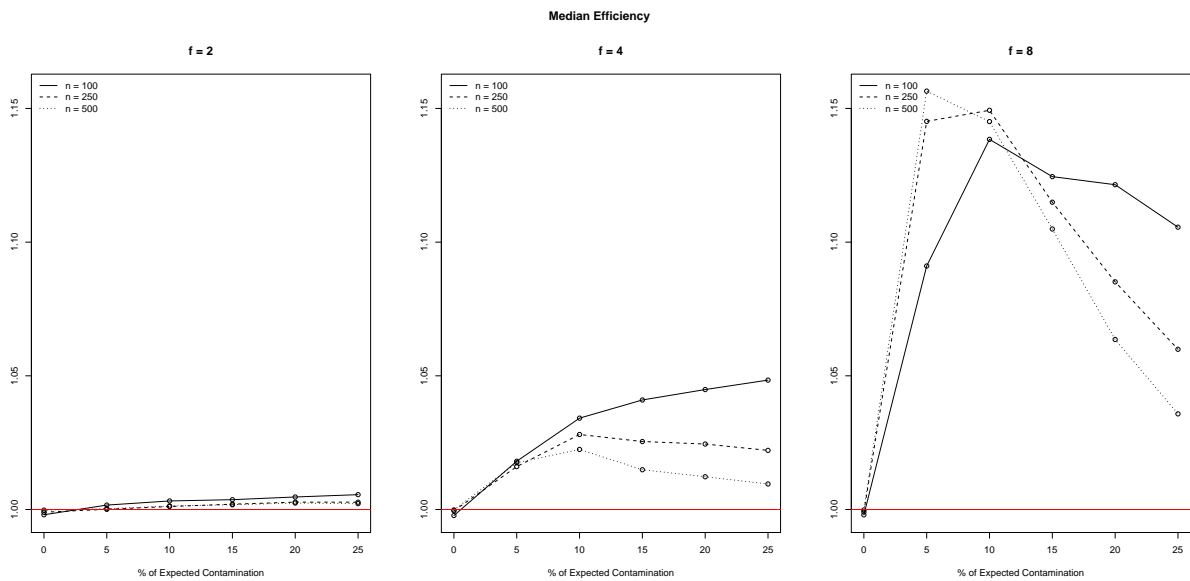


Figura 6 – Mediana das eficiências para magnitude  $f = 2$  (esquerda),  $f = 4$  (centro) and  $f = 8$  (direita),  $n = 100$  (linha sólida),  $n = 250$  (linha traçada) e  $n = 500$  (linha pontilhada). O eixo x representa a percentagem de contaminação esperada e a mediana da eficiência no eixo y.

Para comparar o ajuste do modelo foi utilizado os critérios DIC, assim como o CPO. Primeiramente, define-se o DIC relativo (RDIC) como

$$RDIC_j = \frac{DIC_{Gj} - DIC_{tj}}{DIC_{tj}}, \tag{5.1}$$

para cada uma das amostras simuladas,  $j = 1, \dots, 1000$ . Na parte de cima da Figura 7 é plotada a mediana dos valores de RDIC obtidos pelo ajuste do modelo Gaussiano e pelo ajuste do modelo t para cada cenário. Desta figura, observa-se o mesmo padrão da Figura 6.

A estatística de resumo fornecida pelo critério do CPO é chamada de logaritmo da pseudo verossimilhança (LPML) que avalia o poder preditivo do modelo. Portanto, para comparar ambos os modelos a diferença entre LPML é utilizada. Para tornar esta medida comparável com outra medida, por exemplo o DIC, nós definimos o -LPML por

$$-\text{LPML}_j = - \left( \sum_{i=1}^n \log \left\{ \pi \left( y_i \mid \mathbf{y}_{-i} \right) \right\} \right)_j$$

onde  $j$  é a  $j$ -ésima amostra de um dado cenário. Nesta definição, baixos valores de -LPML indicam um melhor poder de predição. A fim de comparar ambas as abordagens, foi computado o logaritmo do Pseudo Fator de Bayes (IPsBF) (GEISSER; EDDY, 1979) para cada iteração. Esta medida é definida como

$$\text{IPsBF}_j = -\text{LPML}_{Gj} - (-\text{LPML}_{tj}) = \text{LPML}_{tj} - \text{LPML}_{Gj}.$$

Para tornar a comparação equivalente ao RDIC apresentado na Equação (5.1) define-se o IPsBF relativo (RPsBF) como

$$\text{RPsBF}_j = \frac{\text{LPML}_{tj} - \text{LPML}_{Gj}}{-\text{LPML}_{tj}}.$$

Analisando parte de baixo da Figura 7, todas as conclusões do MSE e RDIC podem ser aplicadas no contexto do RPsBF, porém os ganhos se tornam mais evidentes. Além disto, pode-se ver que quando o cenário simulado é estável com baixa proporção esperada e baixa contaminação, a mediana do RPsBF é pequena e não muito significativa. No entanto, para grandes tamanhos de amostra e contaminação é mostrado que a abordagem Student-t é preferível para a maioria dos cenários e enaltece esta escolha quando a magnitude da contaminação aumenta alcançando valores desta medida ainda mais altos do que 10% em alguns casos. Um fato curioso é de que no cenário mais crítico onde  $p = 0.25$ ,  $f = 8$  e  $n = 500$  os valores do RPsBF apontaram incisivamente para a abordagem Gaussiana indicando que, uma vez que o processo gerador possui muita contaminação e gera tantos outliers inovativos, mesmo a abordagem Student-t não é apta a controlar este comportamento produzindo medidas de predição menos acuradas.

A partir do estudo de simulação, pode-se concluir que o modelo mais flexível é preferível sobre o modelo tradicional na maioria dos casos analisados e a distância entre as abordagens são maiores quando um número moderado de outliers inovativos são envolvidos.

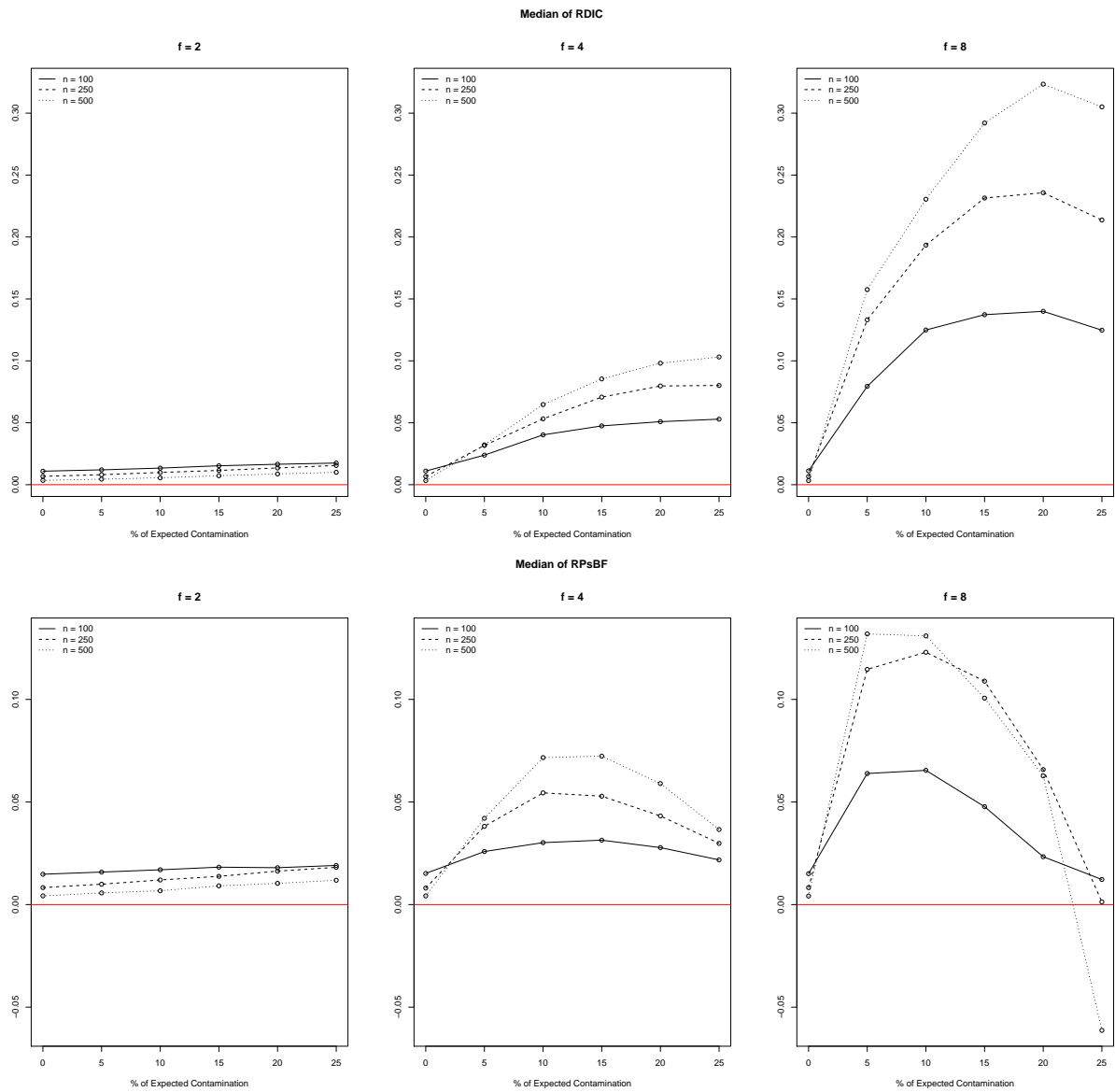


Figura 7 – Topo: Mediana de RDIC, no eixo y, para todos cenários; Abaixo: Mediana de RPsBF, no eixo y, para todos cenários. O percentual de contaminação esperada se encontra no eixo x e todos os cenários são:  $f = 2$  (esquerda),  $f = 4$  (centro) and  $f = 8$  (direita),  $n = 100$  (linha sólida),  $n = 250$  (linha traçada) e  $n = 500$  (linha pontilhada).

## 6 Aplicação

Este capítulo terá como objetivo a aplicação prática da nova abordagem para estimação de modelos dinâmicos utilizando inferência Bayesiana aproximada com o R-INLA. O primeiro estudo trata-se de taxas anuais de homicídios de municípios brasileiros e o estudo subsequente contempla dados mensais de dengue do estado de Minas Gerais. Todas as distribuições descritas na Etapa 2 da Seção 2.4 foram obtidas através da aproximação de Laplace Simplificada, que corresponde a aproximação padrão do R-INLA.

### 6.1 Estudo de Homicídios

Os dados utilizados serão dados de homicídios disponíveis publicamente no banco de dados do Departamento de Informática do Sistema Único de Saúde Brasileiro<sup>1</sup> onde estes dados representam as taxas de homicídio padronizadas pela população de cada município brasileiro.

Segundo [Waiselfisz \(2012\)](#), pela legislação vigente no Brasil (Lei n 15, de 31/12/73, com as alterações introduzidas pela Lei n° 6.216, de 30/06/75), nenhum sepultamento pode ser feito sem a certidão de registro de óbito correspondente. Esse registro deve ser feito à vista de declaração de óbito atestado por médico ou, na falta de médico na localidade, por duas pessoas qualificadas que tenham presenciado ou constatado a morte. A declaração de óbito, instrumento padronizado nacionalmente, normalmente, fornece dados relativos à idade, sexo, estado civil, profissão, naturalidade e local de residência da vítima. Determina a legislação, e igualmente, que o registro do óbito seja sempre feito "no lugar do falecimento", isto é, no local da ocorrência do evento. Os dados de homicídios analisados representam as taxas por 100.000 habitantes de séries históricas anuais de 1980 a 2010 dos municípios Brasileiros.

Até 1995, as causas das mortes eram classificadas pelo Subsistema de Informação sobre Mortalidade do Ministério da Saúde seguindo os capítulos da nona revisão da Classificação Internacional de Doenças (CID-9) da Organização Mundial da Saúde. A partir daquela data, o Ministério da Saúde adotou a décima revisão vigente até os dias de hoje (CID-10). As categorias utilizadas para a recuperação dos óbitos por homicídios foram os códigos E960 até E969 da CID-9 até o ano de 1995 e X85 até Y09 da CID-10 posteriormente ([CASTRO; ASSUNCAO; DURANTE, 2003](#)). Estas categorias, que correspondem as categorias habitualmente classificadas como homicídios, recebem o título genérico de

---

<sup>1</sup> [www.datasus.gov.br/](http://www.datasus.gov.br/)

*Agressões.* Tem como característica a presença de uma agressão intencional de terceiros, que utiliza qualquer meio para provocar danos, lesões ou a morte da vítima.

Os estudos de homicídio possuem um amplo campo de investigação sociológica (veja, por exemplo, [Jacobs e Richardson \(2008\)](#)). Homicídios representam uma categoria específica de crime que embora tenha uma proporção muito menor com relação aos crimes patrimoniais, como roubo e assalto, geram uma forte demanda por parte da população por políticas públicas de controle e prevenção. Um exemplo desta demanda pode ser verificado nos discursos políticos que muitas vezes pautam suas discussões em homicídios. Neste sentido, estudos que lidam com a dinâmica temporal dos homicídios procuram de uma maneira geral associar este comportamento a fatores de ordem econômica, social e política. Por exemplo, altas taxas de homicídio no Brasil são atribuídas a níveis elevados de desemprego, pobreza e desigualdade ([MIR, 2004](#)). Estrutura demográfica ([GRAHAM; BOWLING, 1995](#); [FLOOD-PAGE et al., 2000](#)) e crescimento populacional desordenado são fatores de outra ordem que podem funcionar como hipóteses explicativas ([BLAU; BLAU, 1982](#)). Por fim, a alteração no comportamento histórico criminal é associada a elementos relativos ao *law-enforcement* como aumento do efetivo policial, gastos com políticas públicas de segurança e crescimento da taxa de aprisionamento. Analisando os dados da cidade de Nova York, [Zimring \(2007\)](#) concluiu que "existe uma forte evidência que alterar a quantidade de policiais, bem como tática de policiamento, tem um importante impacto sobre o crime"(pg. 151). Especificamente no Brasil, [Goertzel e Kahn \(2009\)](#), estudando o comportamento o forte declínio das taxas de homicídios observadas no estado de São Paulo a partir da década de 2000, concluem que repressão policial e políticas de controle do armamento reduziram substancialmente homicídios e outros crimes violentos no estado. Estatisticamente, era esperado em que certas regiões, as taxas de criminalidades sofreriam quebras estruturais de tendência devido a essas intervenções o que justificaria a utilização de um modelo dinâmico com uma abordagem mais flexível, isto é, com caudas mais pesadas, na distribuição de probabilidade do sistema não-observado para a acomodação de possíveis outliers inovativos como discutido no Capítulo 5.

O modelo utilizado foi um modelo similar ao utilizado no Capítulo 5. A abordagem adotada foi considerar grupos de municípios de acordo com algum critério e modelar as taxas de homicídio. Seja  $Y_{it}$  a taxa de homicídio da cidade  $i$  para o tempo  $t$ , as análises foram realizadas de acordo com os seguintes modelos:

$$\begin{aligned} Y_{it} &= a_t + v_{it}, & v_{it} &\sim N(0, V_i), t = 1, \dots, 31 \\ a_t &= a_{t-1} + w_t, & w_t &\sim N(0, W), t = 2, \dots, 31 \end{aligned}$$

e modelo competidor

$$\begin{aligned} Y_{it} &= a_t + v_{it}, & v_{it} &\sim N(0, V_i), t = 1, \dots, 31 \\ a_t &= a_{t-1} + w_t, & w_t &\sim Student - t(0, \nu, \tau), t = 2, \dots, 31. \end{aligned}$$



O indexador  $t$  é o indexador que varia ao longo dos anos totalizando em uma amostra de tamanho 31 para cada cidade e o indexador  $i$  varia com as cidades. Portanto, foi assumido prioris diferentes para a precisão de cada cidade na equação observada e uma priori de precisão para o sistema latente. Para a priori do sistema observado, optou-se por uma priori que cobriria com alta densidade de probabilidade o conjunto de todas as variâncias das cidades Brasileiras. Uma priori  $V_i \sim Gama(5, 500)$ ,  $i \in \{Conjunto das Cidades\}$  foi adotada, pois assume uma massa de 90% de probabilidade entre o primeiro e terceiro quartil dessa distribuição de variâncias. É importante salientar que o estado latente representado pelo vetor  $\mathbf{a}$  de dimensão 31, pode ser interpretada como a tendência não-observada do conjunto de cidades de cada análise.

Salientamos que devido os dados analisados possuem uma distribuição de Poisson  $z_{it} | \lambda_{it} \sim Poisson(E_{it}\lambda_{it})$  onde  $z_{it}$ ,  $\lambda_{it}$  e  $E_{it}$  representariam, respectivamente, o número de casos da cidade, taxa de homicídio e população  $i$  no tempo  $t$ , uma transformação descrita na Seção 4.3 de Rojas (2007) poderia ser realizada, fazendo com que  $\sqrt{Y_{it}}$  teria uma distribuição aproximadamente Gaussiana. Além de estar interessados no comportamento do método com um modelo similar aos Capítulos 4 e 5, tal transformação não é feita basicamente por dois motivos. Primeiramente, necessitaríamos trabalhar com as populações anuais interpoladas entre os anos censitários o que afeta a análise e, secundamente, a variância da aproximação Gaussiana de  $\sqrt{Y_{it}}$  dependeria de  $E_{it}$  o que inviabilizaria o uso do R-INLA, neste contexto.

Para as priori da precisão do sistema latente Gaussiano do primeiro modelo uma  $Gama(1, 0.1)$  foi assumida para todos os casos e para o modelo Student-t a mesma priori para o parâmetro de precisão  $Gama(1, 0.1)$  foi assumida e a priori apresentada no Capítulo 4.3 com massa de 30% de probabilidade entre os valores 2 e 10.

A primeira abordagem adotada foi estimar cada modelo para um conjunto de capitais nas taxas de homicídio e outro modelo que representaria os vizinhos de primeira ordem (cidades que fazem fronteira) destas capitais. Então para todo conjunto de capitais, temos as estimativas das tendências dos vizinhos de primeira ordem do respectivo grupo. Três divisões foram propostas por motivos sociológicos em termos de tamanho da população e importância a conjuntura econômica brasileira. A primeira divisão, onde os resultados do modelo t podem ser analisados na Figura 8, foi realizada nas 5 regiões, Sul, Sudeste, Nordeste, Centro-Oeste e Norte, do Brasil. Portanto para a região Sul, por exemplo, temos um modelo para as capitais Porto Alegre, Florianópolis e Curitiba e apenas um modelo para todos os vizinhos de primeira ordem de todas estas capitais e assim por diante para cada região. Em todas as regiões foram retiradas cidades pequenas que apresentaram valores faltantes nas taxas, assim como valores nulos de homicídio (com exceção da região Norte que foram retiradas apenas as com valores faltantes, devido ao excesso de cidades que apresentavam ao menos um valor nulo).

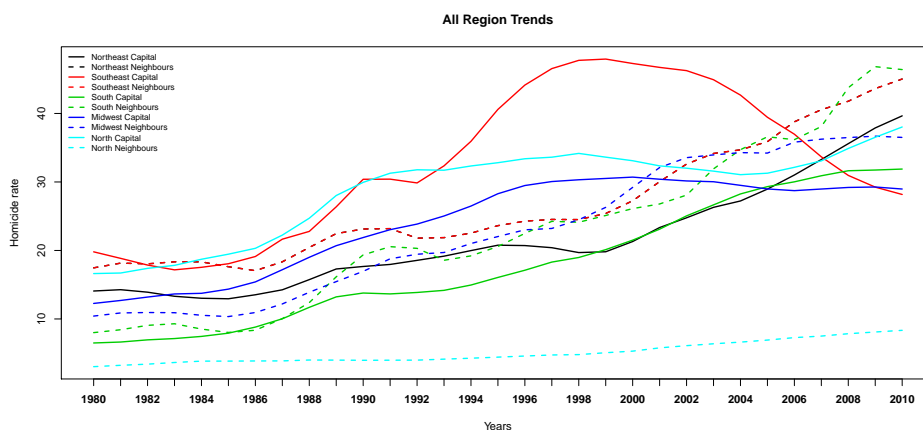


Figura 8 – Tendência das capitais e vizinhos das regiões Brasileiras

A segunda abordagem adotada foi a criação de duas divisões afim de estimar ambos os modelos. A primeira divisão é composta pelas seguintes capitais.

- Grupo 1 (G1) - São Paulo e Rio de Janeiro
- Grupo 2 (G2) - Belo Horizonte, Recife, Vitória e Porto Alegre
- Grupo 3 (G3) - Todas 21 capitais restantes

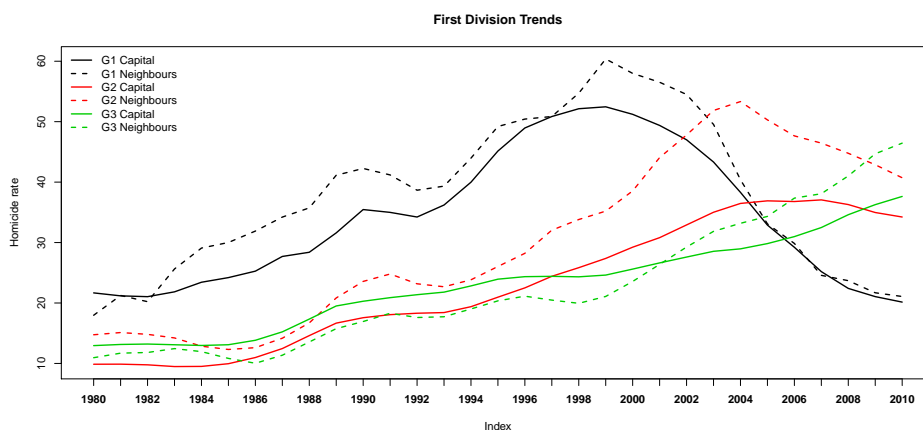


Figura 9 – Tendências da primeira divisão para cada grupo

Para entender as tendências temporais, a literatura sociológica analisa o contexto da maneira que a violência interpessoal acontece. Analisando uma série histórica de mais de 30 anos de crimes e comportamentos delinquentes, [Shaw, Clifford e McKay \(1942\)](#) verificaram que não somente crimes, mas diversos problemas sociais foram relacionados com um ambiente de desorganização social. A *Teoria da Desorganização Social* estabelece que o processo de urbanização não-planejado ajuda a compreender o comportamento

delinquencial. No Brasil, o desenvolvimento das cidades foi desencadeado conjuntamente com o surgimento de enclaves urbanos (como favelas e áreas de ocupação desordenada) onde o enfraquecimento dos mecanismos tradicionais de controle social favorece a ocorrência de um ambiente de oportunidade criminal diferenciada (SUTHERLAND; CRESSEY; LUCKENBILL, 1992). Para mais informações sobre a *Teoria da Desorganização Social* veja Shaw, Clifford e McKay (1942) e Kubrin e Weitzer (2003).

A Figura 9 representa a tendência a posteriori,  $\alpha$ , de cada grupo e vizinhos de capitais para o modelo Student-t. A tendência do primeiro grupo nos diz que seu processo de urbanização começou mais cedo, quando comparado com os outros grupos porque a taxa de homicídio começou a aumentar antes dos demais. Outro aspecto a salientar foi de que até o ano 2000, todas as tendências foram aparentemente lineares para todos os grupos. No entanto, uma reversão da tendência foi observada como resultado de investimentos e políticas de controle criminal estabelecidos no G1 (GOERTZEL; KAHN, 2009). Os outros dois grupos estão se direcionando para o mesmo comportamento, porém sem demonstrar de maneira tão evidente as políticas de segurança como observada no G1. No G2 alguns estados já adotaram algumas medidas de segurança, como por exemplo Minas Gerais, a qual a capital é Belo Horizonte, com a criação da *Integração da Gestão em Segurança Pública* (IGESP<sup>2</sup>) em Maio de 2005, e Pernambuco, a qual capital é Recife, com a criação do *Pacto pela Vida*<sup>3</sup> em Maio de 2007.

A segunda divisão foi motivada pela criação de um limiar populacional de 600.000 habitantes no ano de 1980 entre os grupos G2 e G3 anteriores, resultando na seguinte divisão:

- Grupo 1 (G1) - São Paulo e Rio de Janeiro
- Grupo 2 (G2) - Salvador, Brasília, Fortaleza, Belo Horizonte, Manaus, Curitiba, Recife, Porto Alegre, Belém e Goiânia
- Grupo 3 (G3) - Todas 15 capitais restantes

A segunda divisão estabelecida, que possui os resultados apresentados na Figura 10, possui praticamente as mesmas interpretações e conclusões que a primeira, salvo o fato de que algumas diferenças são observadas após o ano 2000. Podemos ver que adicionando cidades menores no grupo intermediário, a tendência da taxa de homicídio se reduz e decai significativamente no que tange os vizinhos de primeira ordem.

A fim de avaliar a qualidade do ajuste desta abordagem robusta, nos computamos -LPMLs e DICs para todos os modelos e se optou por apresentar os resultados para a primeira divisão dos grupos, como pode ser visto na Tabela 1. Desta tabela, todos os

<sup>2</sup> [www.sids.mg.gov.br/igesp](http://www.sids.mg.gov.br/igesp)

<sup>3</sup> [www.pactopelavida.pe.gov.br/pacto-pela-vida](http://www.pactopelavida.pe.gov.br/pacto-pela-vida)

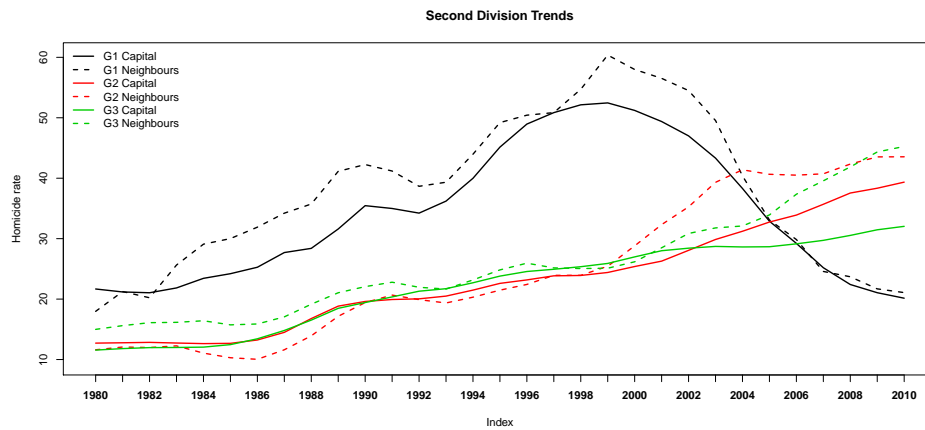


Figura 10 – Tendências da segunda divisão para cada grupo

		G1		G2		G3	
		Gaussiana	Student-t	Gaussiana	Student-t	Gaussiana	Student-t
-LPML	Capitais	229.09	227.81	496.72	496.45	2481.70	2473.55
	Vizinhos	3536.78	3517.38	2379.98	2375.89	3026.54	3002.48
DIC	Capitais	460.22	459.64	997.44	996.60	4960.98	4946.32
	Vizinhos	7063.15	7030.68	4753.67	4746.94	6040.70	6004.27

Tabela 1 – Medidas de qualidade da primeira divisão. G1: São Paulo e Rio de Janeiro; G2: Belo Horizonte, Recife, Vitória e Porto Alegre; G3: Demais 21 capitais

critérios, -LPMLs e DICs, apontaram para a abordagem robusta assumindo a distribuição  $t$  para o ruído do sistema. Para verificar a evidência de que a abordagem robusta supera significativamente a tradicional nos dados reais, optou-se analisar a Tabela 2 que foi proposta e utilizada em Prates et al. (2010). Desta tabela, pode-se ver que o PsBF, que é a diferença de -LPML como apresentada no Capítulo 5, possui uma evidência na modelagem das capitais no G1 ( $2PsBF = 2.56$ ) e uma forte evidência para o modelo dos vizinhos de G2 ( $2PsBF = 8.18$ ). No entanto, uma maior diferença é verificada quando compara-se ambas abordagens para os vizinhos de G1 e G3 assim como nas capitais deste último grupo, com todas tendo forte evidência em termos de diferença de -LPML. Como para cada grupo tem-se diversas séries temporais referentes aos vizinhos de primeira ordem (para G1, G2 e G3 existem 30, 24 e 152 vizinhos respectivamente) acredita-se que a medida que o número de cidades aumenta existe a demanda de uma abordagem mais robusta conforme verificado.

Nesta aplicação a dados reais, o ganho em termos de poder preditivo foi muito claro. Também é válido apontar que existe uma evidência de desvio da distribuição Gaussiana, quando analisamos a distribuição a posteriori do  $\nu$  no modelo Student-t. Da Tabela 3, pode-se ver a mediana a posteriori e os intervalos de credibilidade de 95%. A medida mediana indica que a distribuição Student-t é concentrada em valores médios de graus de

2PsBF	Evidência contra Gaussiana
$(-1,1]$	mencionar
$(1,5]$	positiva
$(5,9]$	forte
$(9,\infty)$	muito forte

Tabela 2 – Critério do Pseudo Bayes Factor

liberdade mas o intervalo de credibilidade se mostra muito assimétrico alcançando altos valores.

Grupo	Tipo	Mediana	95% Intervalo de Credibilidade
G1	Capitais	35.53	(5.84; 441.30)
	Vizinhos	28.46	(6.04; 339.38)
G2	Capitais	39.93	(6.73; 533.77)
	Vizinhos	31.69	(5.56; 445.12)
G3	Capitais	32.23	(5.14; 434.54)
	Vizinhos	55.91	(21.69; 556.67)

Tabela 3 – Medidas a posteriori do  $\nu$ . G1: São Paulo e Rio de Janeiro; G2: Belo Horizonte, Recife, Vitória e Porto Alegre; G3: Demais 21 capitais

## 6.2 Estudo de Casos de Dengue

O vírus da dengue constitui uma doença que se mostra presente em todo o mundo, principalmente em países de região tropical. Em todo o mundo cerca de 2,5 bilhões de pessoas residem em regiões de clima tropical constituindo uma população com alto risco de contaminação da doença. O estado de Minas Gerais reside na região sudeste brasileira e em períodos anteriores a 2000, agrupando por regiões, a Sudeste foi a que registrou o maior número de casos, sendo também a de maior população e disponibilidades de recursos para diagnóstico e notificação, seguidas de Nordeste, Centro-Oeste, Sul e Norte (BORGES, 2001). Nos períodos posteriores, o Brasil vem sofrendo ondas de epidemias sendo a primeira constatada em 2002. As altas temperaturas, grande volume de chuvas e presença do tipo 1 do vírus explicam parte das epidemias<sup>4</sup>.

O banco de dados analisado trata-se de dados podendo ser acessados publicamente no banco de dados do Departamento de Informática do Sistema Único de Saúde Brasileiro<sup>5</sup> e representam a série histórica do número bruto de casos notificados de dengue do estado de Minas Gerais do primeiro mês de aparecimento dos sintomas. A base constitui as informações por município de residência dos pacientes abrangendo informações desde Janeiro de 2008 até Dezembro de 2012.

<sup>4</sup> Número de casos de dengue no País sobe 109% - O Estado de S.Paulo, 26 de Fevereiro de 2010

<sup>5</sup> [www.datasus.gov.br/](http://www.datasus.gov.br/)

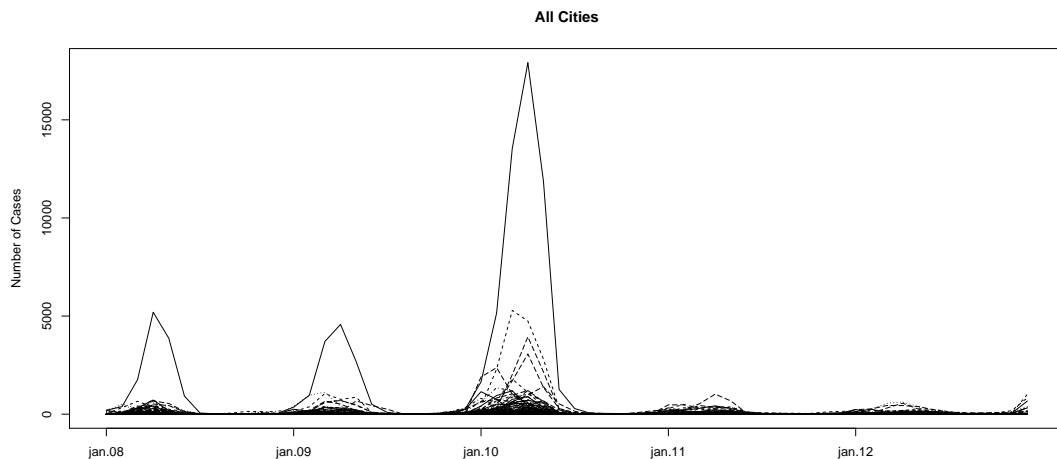


Figura 11 – Número de casos de todas as cidades de Minas Gerais de Janeiro de 2008 à Dezembro de 2012

Analisando a Figura 11 muitas características deste conjunto de dados são constatadas. O forte padrão sazonal dos meses de notificação da dengue são presentes em períodos de início de ano que configuram o verão e, conseqüentemente, uma maior frequência de períodos chuvosos. Além disso, podemos notar que algumas séries apresentam valores muito mais discrepantes das demais, por exemplo a linha em destaque que apresenta meses com valores brutos acima de 15000, trata-se da capital Belo Horizonte que apresenta a maior população do estado com mais de 2,3 milhões de habitantes e, neste mesmo período, uma série apresenta valores próximos a 5000 casos de notificação e ilustra a série referente à cidade de Contagem que apresenta terceira maior população de Minas Gerais com 608.715 habitantes<sup>6</sup>, população muito próxima a da segunda cidade mais populosa, Uberlândia. A última característica a ser destacada trata-se do comportamento anômalo presente nos dados no período dos primeiros meses de 2010.

Um dos objetivos foi verificar o ajuste e poder preditivo de modelos estimados conjuntamente para grupos de cidades e, tendo em vista o comportamento anômalo presente no início do ano de 2010, justificar a abordagem robusta assumindo uma distribuição não-Gaussiana para melhorar estimação em termos de qualidade e predição. Porém antes de estimar modelos conjuntos, avaliar-se-á o desempenho da metodologia em 10 cidades selecionadas que representam grande parte do número de casos de dengue no estado de Minas Gerais. A abordagem inicial neste conjunto de dados foi a construção de modelos separados por municípios, a fim de avaliar o melhor modelo que se ajusta e prevê melhor os dados. Optou-se por omitir os dados referentes a 2012 e ajustar três modelos com os

<sup>6</sup> Dado referente ao censo de 2010

quatro anos restantes. O primeiro modelo (M1) é,

$$\begin{aligned} y_t | \mu_t &\sim \text{Poisson}(\mu_t) \\ \log(\mu_t) &= \eta_t = a_t \\ a_t &= \phi a_{t-1} + \varepsilon_{1t}, \quad \phi \in (0, 1), \end{aligned}$$

o segundo modelo (M2)

$$\begin{aligned} y_t | \mu_t &\sim \text{Poisson}(\mu_t) \\ \log(\mu_t) &= \eta_t = a_t \\ a_t &= a_{t-1} + \varepsilon_{1t}, \end{aligned}$$

por fim, o último modelo (M3) sendo

$$\begin{aligned} y_t | \mu_t &\sim \text{Poisson}(\mu_t) \\ \log(\mu_t) &= \eta_t = a_t \left( \text{sen} \left( \frac{2\pi t}{12} \right) + 2 \right) + T_t \\ a_t &= a_{t-1} + \varepsilon_{1t} \\ T_t &= T_{t-1} + \varepsilon_{2t} \end{aligned} \tag{6.1}$$

Os dois primeiros modelos foram estimados a fim de verificar se uma abordagem mais simples seria suficiente de ajustar bem os dados em um universo que compreende os 10 municípios mais populosos do estado. O último modelo tentar capturar a tendência sazonal incluindo um termo trigonométrico não-negativo que possui picos e vales nos meses de maneira muito semelhantes aos apresentados na análise descritiva de todas as cidades. Além disso, para contemplar uma possível tendência presente na série sob estudo, o termo  $T_t$  é introduzido aumentando a sua complexidade de ajuste. Para os termos de ruídos  $\varepsilon_{1t}$  as abordagens Gaussiana e t-Student foram assumidas, com proris vagas para os hiperparâmetros associados. Para o terceiro modelo apenas dois cenários de suposição de distribuição dos erros foram assumidos sendo ora ambos Gaussianos, ora ambos t-Student. As comparações foram feitas em termos de ajuste através do DIC e previsão do ano de 2012 através do erro quadrático médio (MSE) dado pela média das diferenças ao quadrado dos valores previstos, através da mediana a posteriori, e verdadeiros. Por exemplo, para o município de Belo Horizonte, podemos analisar na Figura 12 os valores ajustados do Modelo 3, assumindo distribuição Gaussiana para ambos os ruídos, para Belo Horizonte, o preditor linear  $\eta_t$  dos dados e os anos de 2011 e 2012 no gráfico mais abaixo. Neste caso o erro de previsão é dado pela média das diferenças dos quadrados entre a predição e o valor real, a partir de Janeiro de 2012.

Além destes modelos separados por cidade, um modelo conjunto para estas 10

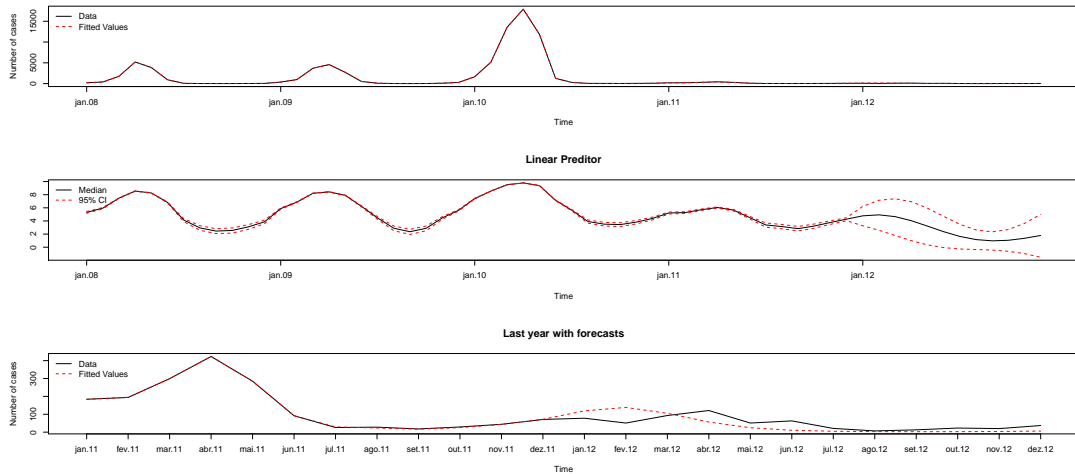


Figura 12 – Topo: Ajuste de M3 para Belo Horizonte (Gaussiano e Gaussiano), no eixo y, número de casos; Centro: Preditor Linear com 95% de intervalo de credibilidade; Abaixo: anos de 2011 e 2012 com os dados reais e ajustados.

idades também foi estimado com a seguinte estrutura

$$\begin{aligned}
 y_{it} \mid \mu_{it} &\sim Poisson(\mu_{it}) \\
 \log(\mu_{it}) &= \eta_t = a_t \left( \text{sen} \left( \frac{2\pi t}{12} \right) + 2 \right) + T_t \\
 a_t &= a_{t-1} + \varepsilon_{1t} \\
 T_t &= T_{t-1} + \varepsilon_{2t}
 \end{aligned}$$

onde  $i$  representa a cidade  $i$ . Neste caso, não existe diferença entre os preditores lineares para cada cidade o que afeta significativamente as estimativas de predição individual. A Figura 13 ilustra o ajuste para o modelo que supõe distribuição Gaussiana para ambos os sistemas de equações, sem previsão destes dados para estas 10 cidades. Note que, neste caso, as estimativas representam uma espécie de tendência média das cidades representando a tendência do número de casos de dengue de MG, no entanto, tal suposição não resultaria em boas previsões no contexto global da série.

Os resultados preliminares destas abordagens podem ser analisadas na Tabela 4. Verificando os valores de predição das abordagens, nota-se que o modelo que contempla o ajuste com a presença do termo trigonométrico para a característica sazonal é superior em todas estas cidades selecionadas. No entanto, a diferença entre elas se dá na distribuição dos erros associados ao vetor  $\mathbf{a}_t$  e  $\mathbf{T}_t$ , ora apontando para o modelo Gaussiano como em Belo Horizonte e Contagem, ora apontando para o modelo robusto como Uberlândia e Betim.

Uma vez tendo analisado previamente o comportamento e desempenho destes modelos neste conjunto de dados, realizamos modelagens para conjuntos de cidades optando pelas características constatadas com os três modelos anteriores. Primeiramente, na aná-



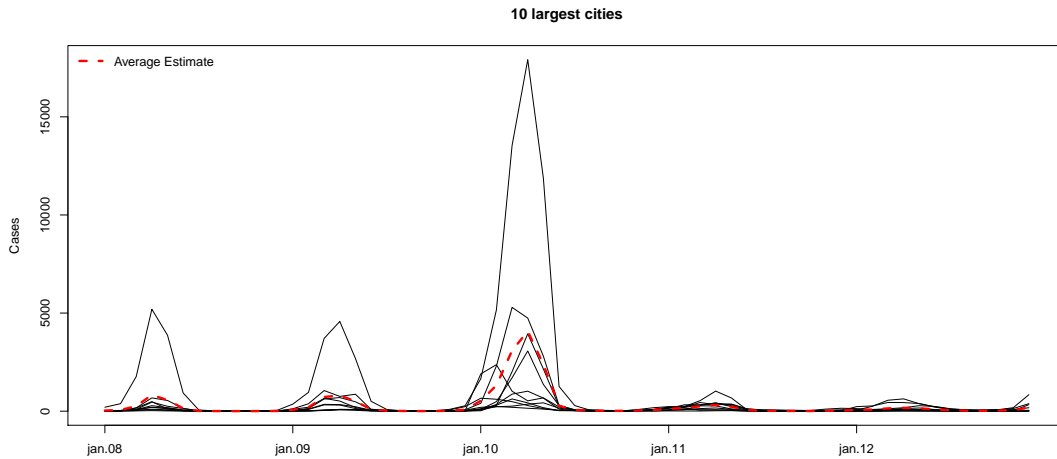


Figura 13 – Estimativa do modelo Gaussiano conjunto para as 10 cidades mais populosas de Minas Gerais. A linha tracejada vermelha representa o exponencial do preditor linear.

lise descritiva fica evidente a relação direta da população do município com o número de casos de dengue, sugerindo a contemplação disto na modelagem com a inclusão de *offset* representando a população de cada município na verossimilhança. Outra característica constatada foi de que na modelagem que inclui o termo trigonométrico, o vetor latente  $\mathbf{T}_t$  se mostrava resultando valores muito próximos de zero e com pouca contribuição no preditor linear, causando assim sua exclusão nas análises subsequentes.

O objetivo foi verificar a melhor abordagem para um conjunto de cidades, contemplando e ajustando os pontos levantados anteriormente, a fim de melhorar a estimação e previsão de novos casos de dengue para elas. Para isto, estimou-se dois modelos, o primeiro deles, C1, sendo

$$y_{it} \mid E_i \mu_{it} \sim \text{Poisson}(E_i \mu_{it})$$

$$\log(E_i \mu_{it}) = \log(E_i) + \beta_0 + \theta_i + a_t \left( \text{sen} \left( \frac{2\pi t}{12} \right) + 2 \right)$$

$$a_t = a_{t-1} + \varepsilon_{1t}$$

e o segundo, C2, modelo sendo

$$y_{it} \mid E_i \mu_{it} \sim \text{Poisson}(E_i \mu_{it})$$

$$\log(E_i \mu_{it}) = \log(E_i) + \beta_0 + a_t \left( \text{sen} \left( \frac{2\pi t}{12} \right) + 2 \right)$$

$$a_t = a_{t-1} + \varepsilon_{1t}$$

onde  $E_i$  representa a população IBGE de 2011 da cidade  $i$  e  $\varepsilon_{1t}$  assume ora distribuição Gaussiana e ora distribuição t-Student.

Tabela 4 – Medidas de ajuste das 10 cidades mais populosas

Cidade	Modelo	Dist. Erro	DIC	MSE
Belo Horizonte	M1	Gaussiana	439,13	2514,21
		Student	439,17	2565,28
	M2	Gaussiana	438,93	2453,12
		Student	439,15	2691,07
	M3	Gaussiana	439,43	2183,98
		Student	439,73	2378,69
Uberlândia	M1	Gaussiana	325,35	1680,28
		Student	325,43	1679,27
	M2	Gaussiana	325,39	1680,29
		Student	325,39	1678,86
	M3	Gaussiana	326,74	1505,17
		Student	326,21	1425,78
Contagem	M1	Gaussiana	348,49	2096,17
		Student	348,58	2094,84
	M2	Gaussiana	348,18	2091,38
		Student	348,00	2099,94
	M3	Gaussiana	349,59	2020,52
		Student	348,98	2041,52
Juiz de Fora	M1	Gaussiana	308,48	18,08
		Student	308,60	17,55
	M2	Gaussiana	307,51	19,66
		Student	310,57	17,23
	M3	Gaussiana	308,77	12,62
		Student	5530,69 <sup>7</sup>	12,66
Betim	M1	Gaussiana	353,23	560,61
		Student	353,53	561,06
	M2	Gaussiana	352,30	562,29
		Student	352,62	561,71
	M3	Gaussiana	354,24	546,33
		Student	354,17	532,10
Montes Claros	M1	Gaussiana	354,52	12668,20
		Student	354,39	12670,66
	M2	Gaussiana	354,73	12665,40
		Student	354,05	12661,30
	M3	Gaussiana	353,51	12644,33
		Student	355,66	12651,14
Ribeirão das Neves	M1	Gaussiana	292,08	22,19
		Student	292,03	22,21
	M2	Gaussiana	292,11	22,19
		Student	21426,95 <sup>8</sup>	22,31
	M3	Gaussiana	292,17	22,12
		Student	292,55	22,08
Uberaba	M1	Gaussiana	313,95	86273,47
		Student	314,19	86015,86
	M2	Gaussiana	314,11	86044,56
		Student	364,28	86575,55
	M3	Gaussiana	316,08	78728,24
		Student	315,61	83985,65
Governador Valadares	M1	Gaussiana	366,70	58945,02
		Student	366,74	58914,61
	M2	Gaussiana	366,88	58891,40
		Student	366,97	58863,85
	M3	Gaussiana	366,74	57958,57
		Student	366,55	56442,49
Ipatinga	M1	Gaussiana	360,13	73168,32
		Student	360,02	73195,14
	M2	Gaussiana	360,01	73170,83
		Student	360,11	74186,68
	M3	Gaussiana	360,76	70860,41
		Student	358,83	70728,08

Um modelo alternativo ao C2 com a ausência do intercepto  $\beta_0$  também foi avaliado, no entanto produzindo resultados insatisfatórios. No primeiro modelo C1, o termo

<sup>7</sup> Problema numérico proveniente de alguma observação

<sup>8</sup> Problema numérico proveniente de alguma observação

$\theta_i$  que poderia ser interpretado como uma estimativa do nível médio de cada série é assumido de duas maneiras. A primeira delas é suposição de que ele é um efeito aleatório Gaussiano independente e a segunda maneira assume uma estrutura de que existe a dependência espacial dos vizinhos de primeira ordem da cidade (BESAG, 1974). A Tabela 5 e a Figura 14 apresenta resultados preliminares estimados com base nas mesmas 10 cidades anteriores para os modelos C1 (efeito aleatório  $\theta_i$  independente) e modelo C2. Nota-se, claramente, que a qualidade de ajuste reduziu quando comparado com a abordagem de modelos separados por cidade, isto se deve ao fato de que agora o ajuste dos parâmetros deve ser conjunto, tendo apenas um fator de correção composto pelo *offset* e/ou pelo efeito aleatório de cada cidade. No entanto, é válido destacar que as estimativas das predições foram semelhantes e, em alguns casos, superior à abordagem anterior.

Tabela 5 – Medidas de ajuste dos modelos conjuntos estimados com as 10 cidades mais populosas

Medida	Cidade	Gaussiana		t-Student	
		C1	C2	C1	C2
MSE	Belo Horizonte	2228,32	1216,88	2242,04	1218,34
	Uberlândia	1707,76	912,31	1707,30	911,80
	Contagem	1474,64	1132,54	1471,25	1128,03
	Juiz de Fora	82,00	112,73	83,01	114,22
	Betim	284,59	347,99	288,45	349,01
	Montes Claros	12534,75	12518,95	12539,97	12523,44
	Ribeirão das Neves	21,54	46,46	21,65	46,95
	Uberaba	92268,07	90650,74	92265,09	90644,32
	Governador Valadares	62514,49	61908,11	62502,69	61892,48
	Ipatinga	77785,98	78450,77	77791,40	78453,41
DIC	Belo Horizonte	7793,695	11131,14	7793,786	11131,18
	Uberlândia <sup>9</sup>	1551,937	16247,82	1551,882	16247,81
	Contagem	2385,468	4197,599	2385,428	4197,574
	Juiz de Fora	8186,377	8371,127	8186,461	8371,217
	Betim <sup>10</sup>	5178,921	11156,05	5178,984	11156,09
	Montes Claros	9781,283	9757,297	9781,267	9757,282
	Ribeirão das Neves	1139,773	2612,539	1139,812	2612,588
	Uberaba	7102,831	8566,895	7102,788	8566,861
	Governador Valadares	7964,591	8137,77	7964,605	8137,787
	Ipatinga	9440,753	9844,417	9440,746	9844,4
MSE Médio		25090,21	24729,75	25091,28	24728,20
DIC Médio <sup>11</sup>		6724,35	7827,35	6724,36	7827,36

O mais importante a observar é a característica de dualidade das estimativas. O modelo mais simples de um modo geral, possui valores de DIC inferiores ao modelo mais complexo que se ajusta melhor aos dados. No entanto, as estimativas de predição do modelo mais complexo são, também de um modo geral, inferiores às do modelo mais

<sup>9</sup> A cidade apresentou problema numérico em alguma observação do modelo t

<sup>10</sup> A cidade apresentou problema numérico em alguma observação do modelo t

<sup>11</sup> Valor calculado sem Uberlândia e Betim

simples, ressaltando ainda mais a dificuldade de obtenção de um modelo ideal em que é minimizado tanto qualidade de ajuste, quanto qualidade de predição. Enquanto um modelo, que possui mais parâmetros e se ajusta bem os dados, tem as suas predições viciadas, o modelo mais simples se ajusta de maneira inferior aos dados, porém com estimativas mais acuradas.

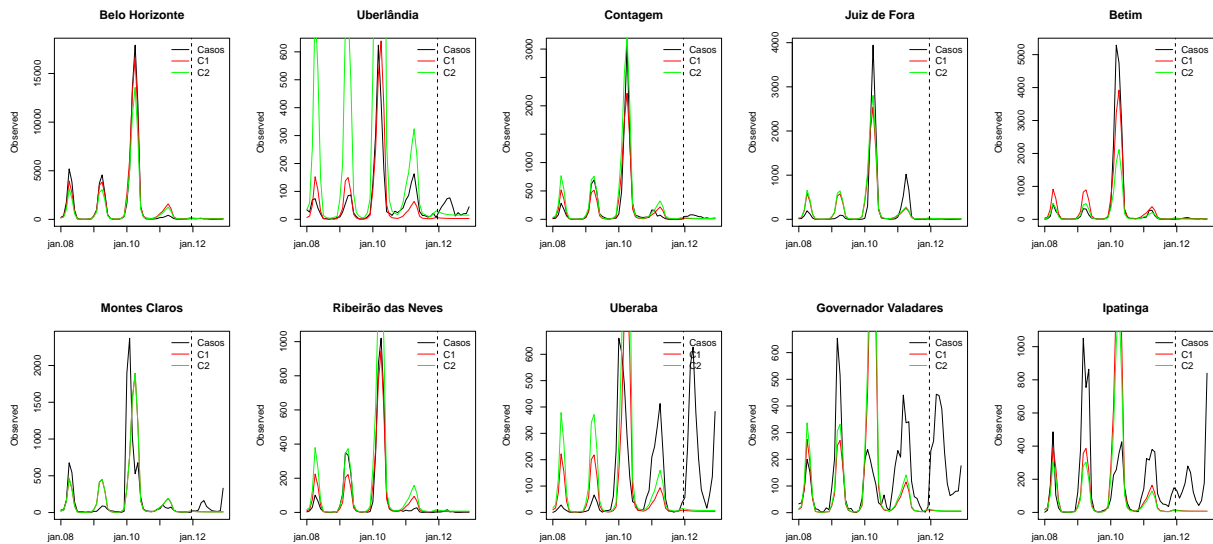


Figura 14 – Estimativas Gaussianas de C1, sem efeito espacial, e C2 para as 10 cidades mais populosas de Minas Gerais. As linhas vermelhas e verdes representam as estimativas para C1, sem efeito espacial, e C2, respectivamente. A linha vertical separa os dados de 2012.

Por fim, os modelos C1, assumindo  $\theta_i$  ora independente e ora com dependência espacial, e C2 foram estimados com base em todos os 853 municípios de MG. Um dos objetivos principais era avaliar uma possível melhoria das estimativas e ajuste assumindo esta estrutura de dependência espacial. A Tabela 6 apresenta o valor mediano de DIC e o valor médio de MSE de todas as cidades, nota-se claramente que os valores são muito próximos entre si, com destaque apenas para os valores de DIC mediano encontrados para as estimativas de cada cidade separadamente do modelo C2.

Tabela 6 – Medidas de ajuste e predição dos modelos estimados com base em todos municípios

	C1				C2	
	Indep.		Besag		-	-
	Gaussiana	t-Student	Gaussiana	t-Student	Gaussiana	t-Student
DIC	73,048	73,052	73,061	73,064	314,275	314,275
MSE	612,807	612,802	612,818	612,830	613,301	613,401

Os ganhos quando analisados por uma medida resumo de todas as cidades está mascarado, devido número excessivo de cidades. Portanto, para analisarem um contexto mais individual, verificou-se qual o modelo, dentre os seis, possuía o menor valor de DIC

e menor valor de MSE para cada cidade. O resultado pode ser verificado na Tabela 7 que ilustra o número bruto de cidades que o respectivo modelo foi superior.

Tabela 7 – Número de cidades presentes em cada modelo. As linhas representam o critério de qualidade de ajuste (DIC); as colunas representam o critério de predição (MSE).

		MSE						Total	
		C1 Indep.		C1 Besag		C2			
		Gaussiana	t-Student	Gaussiana	t-Student	Gaussiana	t-Student		
DIC	C1 Indep.	Gaussiana	54	9	11	18	28	29	149
		t-Student	15	38	3	8	11	7	82
	C1 Besag	Gaussiana	60	34	68	53	49	59	323
		t-Student	52	19	76	52	34	39	272
	C2	Gaussiana	2	3	2	2	4	4	17
		t-Student	2	2	0	3	1	2	10
<b>Total</b>		185	105	160	136	127	140	853	

Analisando esta tabela podemos concluir, por exemplo, que em 54 cidades, dentre todas as 853 do estado de MG, o modelo C1 supondo  $\theta_i$  com estrutura independente e  $\varepsilon_{1t}$  Gaussiano produziu o menor DIC e MSE. Por outro lado, em 59 cidades, o menor DIC resultou do modelo C1 assumindo  $\theta_i$  com estrutura espacial e  $\varepsilon_{1t}$  Gaussiano e, nas mesmas 59 cidades, o menor MSE foi proveniente do modelo C2 com  $\varepsilon_{1t}$  com distribuição t-Student. Observamos novamente que o modelo C2 não foi capaz de vencer os modelos C1 em termos de DIC, no entanto, nesta tabela fica evidente o ganho da inclusão do efeito aleatório com estrutura de dependência espacial introduzida quando comparado com este efeito independente. O número de cidades presentes no modelo C1 Besag nas medidas de DIC é bem superior aos outros dois modelos. No entanto, novamente constatamos uma grande homogeneidade de desempenho nas predições realizadas para todos os modelos. Isto pode estar sendo oriundo da dinâmica temporal adotada que depende somente de um valor do passado onde a clara estrutura sazonal é captada somente pelo termo senoidal que se demonstra ainda insuficiente para contemplar os picos de dengue presentes nos inícios de ano. Por fim, um fato importante para enaltecer novamente é a dificuldade de obter um modelo que seja superior tanto em ajuste, quanto uma capacidade de predição superior quando comparado com modelos competidores. Pela Tabela 7 observamos que apenas 25,56% dos municípios são concordantes em termos de ajuste e predição, com destaque para os 68 municípios que o modelo "C1+Besag+Gaussiano" foi superior em termos de DIC e MSE, ou seja, dificilmente existe um consenso em qual modelo optar.

Para concluir os estudos de aplicação com esta base de dados, a fim de termos comparativos de tempo computacional e complexidade, um modelo alternativo ao modelo (6.1), sem o termo  $\mathbf{T}_t$  (M4), também foi estimado separadamente para cada uma das cidades, assumindo distribuição Gaussiana e t-Student para  $\varepsilon_{1t}$ . Tais informações podem ser analisadas na Tabela 8.

Analisando estes tempos de estimação e complexidade de cada modelo é sugerida a escolha do modelo mais parcimonioso, C2, em que não é incluído o efeito aleatório no

Tabela 8 – Tempo (em minutos) para cada modelo

Modelo	Efeito Aleatório	Dist. Erro	Tempo (min)
C1	Indep.	Gaussiana	14,35
		t-Student	23,02
	Besag	Gaussiana	19,21
		t-Student	30,67
C2	-	Gaussiana	6,97
	-	t-Student	11,45
M4	-	Gaussiana	19,72
	-	t-Student	29,58

preditor linear. Além disso, o tempo computacional demandado para a estimação deste modelo mais simples é significativamente inferior ao tempo computacional dos modelos mais complexos, reiterando a sua escolha em termos de parcimônia e praticidade.

No entanto, cabe ressaltar que analisando conjuntamente os resultados apresentados nas Tabelas 7 e 8 podemos segregar as opções de escolha dependendo do interesse de pesquisa. Em termos preditivos, o modelo C2 pode ser escolhido pelo tempo computacional, parsimônia e devido à homogeneidade do número de cidades ganhadoras dentre todos os modelos estimados, pois nenhum modelo se destacou. Por outro lado, se o objetivo é ajuste em termos de DIC, o C2 é muito inferior, sendo escolhido apenas uma minoria das vezes. Neste caso, algum C1 deve ser optado onde as estimativas assumindo  $\theta_i$  com estrutura espacial figuraram uma boa escolha devido à parsimônia e melhor ajuste sem um custo de tempo considerável.

Uma vez que os diversos modelos adotados para este estudo longitudinal da dengue em MG mostraram a dificuldade de lidar com ajuste e predição nesta específica classe de dados. Métodos de seleção de modelos que fogem do escopo natural de medidas de qualidade à luz dos dados, como, por exemplo, tempo computacional, podem ser considerados como instrumento de decisão nas escolhas. Como os modelos estimados resultaram em heterogeneidade de decisão, a escolha por algum modelo para futuras análises estaria de alguma forma justificada.

## 7 Discussões

Esta dissertação descreve como realizar inferência Bayesiana utilizando a abordagem INLA, através do R-INLA, para estimar modelos dinâmicos generalizados não-Gaussianos, em outras palavras quando o ruído do sistema de evolução, ou sistema observado, possui uma distribuição não-Gaussiana. A estrutura como ela foi construída foi motivada pela apresentação da classe de modelos em que o método INLA trabalha, uma descrição exemplificada desta abordagem e de que maneira ela pode ser utilizada para a estimação destes modelos dinâmicos não-Gaussianos para modelagem de séries temporais.

Tais modelos podem ser vistos como parte de modelos hierárquicos latentes onde um campo latente não-Gaussiano é assumido e, portanto, invalidando o uso direto da metodologia INLA que requer que o campo latente seja Gaussiano.

Utilizando um exemplo de um passeio aleatório de primeira ordem apresentamos como utilizar uma estrutura aumentada para superar a limitação Gaussiana do R-INLA para o campo latente. A chave para entender o motivo que nossa abordagem funciona reside no fato que o campo latente não-Gaussiano é aproximado por uma distribuição Gaussiana e corrigido na função de verossimilhança tentando minimizar a perda desta aproximação para modelos dependentes. Discutimos e explicamos os motivos para fazer esta aproximação e, especialmente, onde nos cálculos do R-INLA isto irá impactar. Além disto, realizamos aplicações em dados reais para dois conjuntos de dados ilustrando a nossa abordagem.

Através de simulações, nós mostramos a necessidade de modelagens mais robustas quando séries temporais sofrem repentinas mudanças estruturais. A partir dos nossos resultados, observamos que modelos Gaussianos são sensíveis a mudanças estruturais enquanto que a nossa abordagem robusta assumindo uma distribuição t-Student é mais robusta. Especificamente, nosso estudo de simulação apresentou uma demanda incisiva para evitar a usual suposição Gaussiana na maioria dos cenários contaminados.

Existem traços de que algumas políticas públicas para controle do crime podem gerar um efeito positivo na tendência temporal dos crimes permitindo/criando mudanças estruturais. É evidente que outros fatores podem ajudar a confirmar estas hipóteses, no entanto é muito provável que investimentos em segurança, tais como às políticas aplicadas nos grupos G1 e G2 na Seção 6.1, têm contribuições na mudança da dinâmica observada. A nossa aplicação nas taxas de homicídios, conforme esperado, apontaram que políticas públicas podem desempenhar um importante papel para explicar dinâmicas ho-

micidas através de uma abordagem robusta devido às características deste tipo de dado. No contexto epidemiológico, o estudo da dengue se mostra importante no contexto de saúde pública e apresentou características a serem relevadas como a presença de surtos epidemiológicos na forma de quebra estrutural de padrão sazonal em anos específicos. As modelagens realizadas puderam mostrar as dificuldades de seleção de modelos no contexto de ajuste e predição de séries temporais. Apesar das motivações de termos escolhido estas aplicações, estamos cientes de que a extensão apresentada poderia também ser justificada em outros campos como em modelos de volatilidade estocástica em, por exemplo, [Jacquier, Polson e Rossi \(2004\)](#).

As perspectivas para futuros trabalhos englobam algumas vertentes. A estrutura sazonal dos dados da dengue captadas através do termo senoidal aparentemente não foi suficiente para contemplar bem os surtos no início dos períodos de previsão. Os modelos estimados poderiam ser aprimorados, assim como modelos alternativos poderiam ser testados, como por exemplo a inclusão de termos não somente de dependência condicional de primeira ordem mas de 12<sup>a</sup> ordem. Além disto, uma vez que a classe de modelos apresentada na dissertação engloba uma imensa gama de problemas, diferentes verossimilhanças para os dados poderiam ser assumidas além da Gaussiana e Poisson, como, por exemplo, Binomial, Skew-Normal ou Gama. Outra extensão seria investigar modelos não-Gaussianos no INLA para casos onde os termos de erro do sistema latente possuam estrutura de dependência mais complexa, como, por exemplo, estrutura espacial, o que abrangeria a aplicação da abordagem INLA para Modelos não-Gaussianos Latentes mais complexos. No entanto, a principal vantagem da estrutura apresentada nesta dissertação é de que ela permite o usuário ajustar diversos modelos dinâmicos não-Gaussianos estruturados com uma boa acurácia e rapidez utilizando uma ferramenta amigável já disponível.

Acreditamos que a comunidade aplicada pode fazer bom uso desta metodologia quando demandada. Para dados de séries temporais reais, não é raro observarmos quebras estruturais e uma abordagem robusta, como a apresentada, pode representar uma melhor adequação para ajustar este tipo de dado. Por fim, formalizamos como utilizar o R-INLA para modelos dinâmicos generalizados não-Gaussianos de uma maneira simples.



## Referências

- AZZALINI, A.; CAPITANIO, A. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society. Series B.*, 1999. v. 61, n. 3, p. 579–602, 1999. Citado na página 22.
- BARRON, A.; SCHERVISH, M.; WASSERMAN, L. The consistency of posterior distributions in nonparametric problems. *Annals of Statistics*, 1999. JSTOR, v. 27, n. 2, p. 536–561, 1999. Citado na página 19.
- BESAG, J. Spatial Interaction and Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B.*, 1974. v. 36, n. 2, p. 579–602, 1974. Citado na página 58.
- BLAU, J. R.; BLAU, P. M. The cost of inequality: metropolitan structure and violent crime. *American Sociological Review*, 1982. v. 47, p. 114–129, 1982. Citado na página 47.
- BORGES, S. M. *Importancia epidemiologica do Aedes Albopictus nas Américas*. [S.l.]: GNU Free Documentation License, 2001. Citado na página 52.
- CASTRO, M. S. M.; ASSUNCAO, R. M.; DURANTE, M. O. Data comparison on homicide deaths between two information systems, Brazil. *Revista de Saúde Pública*, 2003. v. 37, n. 2, p. 168–176, 2003. Citado na página 46.
- DETHLEFSEN C., L.-C. S.; CHRISTENSEN, A. L. *sspir: State Space Models in R*. 2012. Citado na página 26.
- DEY, D. K.; CHEN, M. H.; CHANG, H. Bayesian approach for nonlinear random effects models. *Biometrics*, 1997. v. 53, p. 1239–1252, 1997. Citado na página 42.
- DURBIN, J.; KOOPMAN, S. Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society: Series B*, 2000. v. 62, p. 3–56, 2000. Citado 2 vezes nas páginas 11 e 28.
- FERNANDEZ, C.; STEEL, M. Multivariate student-t regression models: Pitfalls and inference. *Biometrika*, 1999. Biometrika Trust, v. 86, n. 1, p. 153, 1999. Citado na página 39.
- FLOOD-PAGE, C. et al. *Youth Crime: Findings from the 1998/99 Youth Lifestyles Survey*. [S.l.]: Home Office Research Study 209. London: HMSO, 2000. Citado na página 47.
- FONSECA, T.; FERREIRA, M.; MIGON, H. Objective bayesian analysis for the student-t regression model. *Biometrika*, 2008. Biometrika Trust, v. 95, n. 2, p. 325, 2008. Citado na página 39.

- FOX, A. J. Outliers in Time Series. *Journal of the Royal Statistical Society (Series B)*, 1972. v. 3, p. 350–363, 1972. Citado 2 vezes nas páginas 11 e 38.
- GAMERMAN, D.; LOPES, H. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. [S.l.]: Chapman & Hall/CRC, 2006. Citado na página 13.
- GAMERMAN D., S.-T. F. G. A non-Gaussian family of state-space models with exact marginal likelihood. *Journal of Time Series Analysis*, 2013. v. 34, p. 625–645, 2013. Citado na página 27.
- GEISSER, S.; EDDY, W. F. A predictive approach to model selection (Corr: V75 p765). *Journal of the American Statistical Association*, 1979. v. 74, p. 153–160, 1979. Citado na página 44.
- GELFAND, A. E.; DEY, D. K.; CHANG, H. Model determination using predictive distributions, with implementation via sampling-based methods (Disc: P160-167). In: BERNARDO, J. M. et al. (Ed.). *Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting*. [S.l.]: Clarendon Press [Oxford University Press], 1992. p. 147–159. Citado na página 42.
- GELMAN, A.; RUBIN, D. Inference from Iterative Simulation Using Multiple Sequences (with discussion). *Statistical Science*, 1992. v. 7, p. 457–511, 1992. Citado na página 13.
- GOERTZEL, T.; KAHN, T. The Great São Paulo Homicide Drop. *Homicide Studies*, 2009. v. 13, n. 4, p. 398–410, 2009. Citado 2 vezes nas páginas 47 e 50.
- GRAHAM, J.; BOWLING, B. *Young People and Crime*. [S.l.]: Home Office Research Study 145. London: HMSO, 1995. Citado na página 47.
- HARRISON, P.; STEVENS, C. Bayesian Forecasting (with discussion). *Journal of the Royal Statistical Society. Series B.*, 1976. v. 38, p. 205–247, 1976. Citado na página 26.
- HARVEY, A. *Forecasting, Structural Time Series Models and the Kalman Filter*. [S.l.]: Cambridge University Press, 1989. Citado na página 26.
- HARVEY, A.; FERNANDES. Time Series Models for Count or Qualitative Observations. *Journal of Business and Economic Statistics*, 1989. v. 7, n. 4, p. 407–417, 1989. Citado na página 27.
- HELKSKE, J. *KFAS: Kalman filter and smoothers for exponential family state space models*. 2010. Citado na página 27.
- JACOBS, D.; RICHARDSON, A. M. Economic Inequality and Homicide in the Developed Nations From 1975 to 1995. *Homicide Studies*, 2008. v. 12, n. 1, p. 28–45, 2008. Citado na página 47.
- JACQUIER, E.; POLSON, N.; ROSSI, P. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 2004. Elsevier, v. 122, n. 1, p. 185–212, 2004. Citado na página 63.
- KITAGAWA, G. Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American statistical association*, 1987. JSTOR, v. 82, n. 400, p. 1032–1041, 1987. Citado 2 vezes nas páginas 11 e 27.

- KITAGAWA, G. Non-gaussian seasonal adjustment. *Comput. Math. Applic.*, 1989. Elsevier, n. 18, p. 503–514, 1989. Citado na página 28.
- KITAGAWA, G. *The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother*. [S.l.]: Institute of Statistical Mathematics, Tokyo, 1990. Citado na página 28.
- KUBRIN, C. E.; WEITZER, R. New Directions in Social Disorganization Theory. *Journal of Research in Crime and Delinquency*, 2003. v. 40, n. 4, p. 374–402, 2003. Citado na página 50.
- KULLBACK S., L.-R. A. On information and sufficiency. *The Annals of Mathematical Statistics*, 1951. v. 22, p. 79–86, 1951. Citado na página 41.
- LAVINE, M. Another look at conditionally Gaussian Markov random fields. *Bayesian Statistics*, 1999. v. 6, p. 577–585, 1999. Citado na página 18.
- MARTINS, T. *Aproximações Determinísticas para Distribuições a Posteriori Marginais*. [S.l.]: Dissertação de Mestrado. Universidade Federal do Rio de Janeiro, 2010. Citado na página 19.
- MARTINS, T.; RUE, H. Extending INLA to a class of near-Gaussian latent models. *Department of Mathematical Sciences, NTNU, Norway*, 2012. 2012. Citado 7 vezes nas páginas 11, 23, 25, 36, 37, 39 e 42.
- MARTINS, T.; RUE, H. Prior for flexibility parameters: the Student's t case. *Department of Mathematical Sciences, NTNU, Norway*, 2013. 2013. Citado 3 vezes nas páginas 39, 41 e 42.
- MASRELIEZ, C. J.; MARTIN, R. D. Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter. *IEEE Transactions on Automatic Control*, 1977. v. 22, p. 361–371, 1977. Citado na página 38.
- MIGON, H. et al. Dynamic models. *Handbook of Statistics*, 2005. Elsevier, n. 25, p. 553–588, 2005. Citado 2 vezes nas páginas 11 e 18.
- MIR, L. *A guerra civil: estado e trauma*. [S.l.]: São Paulo: Geração Editorial, 2004. Citado na página 47.
- PETRIS, G. *dIm: Bayesian and Likelihood Analysis of Dynamic Linear Models*. 2010. Citado na página 27.
- PINHEIRO, J.; LIU, C.; WU, Y. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, 2001. ASA, v. 10, n. 2, p. 249–276, 2001. Citado na página 42.
- PRATES, M. et al. Intervention Analysis of Hurricane Effects on Snail Abundance in a Tropical Forest Using Long-Term Spatiotemporal Data. *Journal of Agricultural, Biological, and Environmental Statistics*, 2010. v. 16, n. 1, p. 142–156, 2010. Citado na página 51.
- QUARRIE, A. D. M.; TSAI, C. Outlier Detections in Autoregressive Models. *Journal of Computational and Graphical Statistics*, 2003. JSTOR, v. 12, n. 2, p. 450–471, 2003. Citado na página 38.

- RAFTERY, A.; LEWIS, S. How Many Iterations in the Gibbs Sampler? *Bayesian Statistics*, 1992. v. 4, p. 763–773, 1992. Citado na página 13.
- RESENDE, C. *Inferência Bayesiana Aproximada em Modelos de Espaço de Estados, Dissertação de Mestrado*. [S.l.]: Dissertação de Mestrado. Universidade Federal do Rio de Janeiro, 2011. Citado na página 23.
- ROJAS, J. C. V. *Modelos Espaço-Temporais para Dados de Área na Família Exponencial*. [S.l.]: Tese de Doutorado. Universidade Federal do Rio de Janeiro, 2007. Citado na página 48.
- RUE, H.; HELD, L. *Gaussian Markov random fields: theory and applications*. [S.l.]: Chapman & Hall, 2005. Citado na página 18.
- RUE, H.; MARTINO, S. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of statistical planning and inference*, 2007. Elsevier, v. 137, n. 10, p. 3177–3192, 2007. Citado na página 20.
- RUE, H.; MARTINO, S.; CHOPIN, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B(Statistical Methodology)*, 2009. Citeseer, v. 71, n. 2, p. 319–392, 2009. Citado 4 vezes nas páginas 11, 13, 20 e 23.
- RUIZ-CÁRDENAS, R.; KRAINSKI, E.; RUE, H. Direct fitting of dynamic models using integrated nested laplace approximations–inla. *Computational Statistics & Data Analysis*, 2011. Elsevier, 2011. Citado 3 vezes nas páginas 31, 36 e 37.
- SHAW; CLIFFORD, R.; MCKAY, H. *Juvenile Delinquency in Urban Areas*. [S.l.]: Chicago: University of Chicago Press, 1942. Citado 2 vezes nas páginas 49 e 50.
- SMITH, J. A Generalisation of the Bayesian steady forecast model. *Journal of the Royal Statistical Society. Series B.*, 1979. v. 41, p. 378–387, 1979. Citado na página 26.
- SPIEGELHALTER, D. J. et al. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 2002. v. 64, p. 583–639, 2002. Citado na página 42.
- SUTHERLAND, E.; CRESSEY, D.; LUCKENBILL, D. *Principles of Criminology*. [S.l.]: AltaMira, 1992. ISBN 0930390695. Citado na página 50.
- VILLA C., W. S. Objective Prior for t-density Degrees of Freedom. *Bayesian Analysis*, 2013. p. 1–24, 2013. Citado na página 39.
- WASELFISZ, J. *Mapa da Violência 2012: Os novos padrões da Violência Homicida no Brasil*. [S.l.]: Instituto Sangari, 2012. Citado na página 46.
- WEST, M. Robust Sequential Approximate Bayesian Estimation. *Journal of the Royal Statistical Society. Series B.*, 1981. v. 43, p. 157–166, 1981. Citado 2 vezes nas páginas 27 e 28.
- WEST, M.; HARRISON, J. *Bayesian forecasting and dynamic models*. [S.l.]: Springer Verlag, 1997. Citado 2 vezes nas páginas 11 e 27.

---

WEST, M.; HARRISON, J.; MIGON, H. Dynamic Generalized Linear Models and Bayesian Forecasting. *Journal of the American Statistical Association*, 1985. JASA, v. 80, n. 389, p. 73–83, 1985. Citado 2 vezes nas páginas 26 e 27.

ZIMRING, F. E. *The Great American Crime Decline*. [S.l.]: New York: Oxford, 2007. Citado na página 47.