

Universidade Federal do Rio de Janeiro

Departamento de Métodos Estatísticos

*Aproximações Determinísticas
para Distribuições a Posteriori
Marginais*

Thiago Guerrera Martins

Orientador: Prof. Dani Gamerman

Rio de Janeiro

Abril de 2010

Universidade Federal do Rio de Janeiro

Departamento de Métodos Estatísticos

*Aproximações Determinísticas para Distribuições a
Posteriori Marginais*

Thiago Guerrero Martins

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Estatística, Instituto de Matemática da Universidade Federal do Rio de Janeiro (UFRJ), como parte dos requisitos necessários à obtenção do título de Mestre em Estatística.

Orientador: Prof. Dani Gamerman

Rio de Janeiro

Abril de 2010

Universidade Federal do Rio de Janeiro

Departamento de Métodos Estatísticos

*Aproximações Determinísticas para Distribuições a
Posteriori Marginais*

Thiago Guerrero Martins

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Estatística, Instituto de Matemática da Universidade Federal do Rio de Janeiro (UFRJ), como parte dos requisitos necessários à obtenção do título de Mestre em Estatística.

Aprovada por:

Prof. Dani Gamerman, presidente
IM - UFRJ

Prof. Carlos Antonio Abanto-Valle
IM - UFRJ

Prof. Jorge Alberto Achcar
UNESP

Rio de Janeiro, Abril de 2010.

Ficha catalográfica

Martins, Thiago Guerrero.

Aproximações determinísticas para distribuições a posteriori marginais / Thiago Guerrero Martins.

Rio de Janeiro: UFRJ/IM, 2010.

xiii, 98f. ; 30 cm.

Dissertação (mestrado) - UFRJ/IM.

Programa de Pós-graduação em Estatística, 2010.

Orientador: Dani Gamerman.

1. Estatística matemática - Tese. 2. Teoria da decisão estatística bayesiana - Tese. I. Gamerman, Dani.

II. Universidade Federal do Rio de Janeiro.

Instituto de Matemática. III. Título.

“The man who doesn’t read good books has no advantage over the man who can’t read them.”

Mark Twain

Esse trabalho é dedicado a todas as pessoas importantes em minha vida.

Agradecimentos

Eu gostaria de começar agradecendo ao professor Paulão, que foi meu professor de física mecânica durante o ano de vestibular (ao longo de 2003) na Escola São Domingos. Esse agradecimento pode parecer um pouco estranho, já que esta tese data do ano de 2010 e é completamente não relacionada à física mecânica do jeito que me foi ensinado em 2003. No entanto, meu professor de física foi muito além das suas obrigações e, além da física, me ensinou a pensar. Isso mesmo, me ensinou a pensar, de um jeito que, apesar de ter passado grande parte dos meus 17 anos até aquele momento em uma sala de aula, nunca haviam me ensinado antes. Eu já tinha aprendido a ler e a escrever, até mesmo a memorizar fatos e fórmulas, mas pensar de forma clara e organizada era algo novo para mim. De certo modo, eu acho que devo a ele grande parte do sucesso que venho obtendo em qualquer atividade “intelectual” que participe, os fracassos são devidos única e exclusivamente as minhas limitações. E é por isso que consta esse agradecimento ao professor Paulão na minha tese de mestrado em Estatística.

Agradeço ao professor Dani Gamerman, meu orientador de mestrado, pelo tempo que disponibilizou para nossas reuniões, onde através de críticas e sugestões, ajudou a moldar o texto da minha dissertação. Foi meu orientador que me apresentou a área de Inferência Bayesiana Aproximada, uma área extremamente interessante, por sua beleza, dificuldade e importância dentro da estatística Bayesiana. Além disso, ele foi um dos grandes responsáveis pelos dois excelentes meses que passei na Universidade de Ciência e Tecnologia da Noruega (NTNU), por sugerir e viabilizar essa visita.

Tive o prazer de trabalhar com o professor Håvard Rue no período que passei na Noruega, período esse de grande importância para o amadurecimento da minha dissertação, já que Håvard Rue é o nome por trás de grande parte da metodologia citada nesta tese. Sem seus conselhos, o capítulo 5 não teria ficado pronto a tempo de ser incluído neste texto. Gostaria de mencionar os colegas Sara Martino, Alessandro Ottavi, Rupali Akerkar, Xiang

Ping e Rikke Ingebrigtsen, que fizeram parte do meu dia a dia na Noruega e que me ajudaram de alguma forma na minha estadia.

Um muito obrigado a todos os professores do programa de Pós-graduação em Estatística, pela quantidade de conhecimento que me transmitiram ao longo desses dois anos de mestrado. Em especial, gostaria de agradecer ao professor Hélio Migon, que sempre deixou as portas do programa abertas para mim, mesmo quando eu apresentava dúvidas quanto ao caminho que queria seguir, e por ter se esforçado para viabilizar minha dedicação exclusiva a pesquisa ao longo do mestrado.

A meus colegas de laboratório pelos momentos de descontração que passamos juntos e em especial para os membros da minha turminha, João Batista, Kelly Cristina, Larissa Alves, Leonardo Nassif e Rodrigo Targino. Agradeço a Rodrigo Targino por aceitar meus inúmeros convites para um cafezinho, principalmente após eu descobrir, vários meses depois, que ele não gosta de café. Muitas conversas interessantes, profissionais ou não, ocorreram durante esses momentos.

Gostaria de agradecer aos membros fundadores dos Delavorutchas, Felipe Scampini, Vitor Sapucaia, Thales Tedoldi, Ivan Cypriano, Leandro Cardoso, e aos mais novos membros Pedro Guerrero Martins, Lucas Guerrero Martins, Henrique Sapucaia e Daniel Scampini pela satisfação e tranqüilidade de saber que tenho amigos verdadeiros para me ajudar e incentivar nos momentos que eu mais precisar.

A todos os membros da minha família por me incentivarem e por se orgulharem das minhas conquistas. Não vou citar cada nome individualmente por problemas de espaço e também porque acabaria sendo injusto ao não colocar alguns nomes que sem dúvida também fazem parte das minhas vitórias.

Naturalmente tenho que reservar um agradecimento especial ao meu pai, Renato da Silveira Martins, e a minha mãe, Adélia Christina Guerrero Martins, pelo amor incondicional que eles tem por mim. Sem dúvida, sou a pessoa e o profissional que sou graças a educação, dentro e fora de casa, que eles me proporcionaram, mesmo quando exigia sacrifícios da parte de-

les. Agradeço a teimosia do meu pai para que eu aprendesse inglês e lesse livros fora da sala de aula, pois hoje, muitos anos depois, essas atividades se tornaram muito valiosas para os caminhos que resolvi seguir em minha vida. Agradeço a minha mãe pelo seu interesse no que acontece em minha vida, sempre arrumando tempo para me escutar e dar conselhos, e pelo tempo que gastou estudando comigo quando eu ainda era uma criança.

Por último, mas de forma alguma menos importante, gostaria de agradecer minha atual noiva e futura esposa Gabriela Brettas por seu amor e compreensão ao longo desses dois anos. A sua compreensão foi muito importante para que eu pudesse perseguir a mais alta qualidade em meu trabalho, mesmo que para isso eu tivesse que estar a muitos quilômetros de distância. E o seu amor foi essencial ao dar um sentido maior a minha vida e ao meu trabalho.

Thiago Guerrera Martins

Abril de 2010

Resumo

Aproximações Determinísticas para Distribuições a Posteriori Marginais

Thiago Guerrero Martins

Resumo: Um dos grandes desafios em estatística Bayesiana é obter, quando não há solução analítica disponível, aproximações para distribuições a posteriori marginais de forma precisa e eficiente. Nessa dissertação é feita uma revisão na literatura de métodos determinísticos para este fim em um contexto geral, e mostra-se que ainda há modelos de importância atual que são melhor estimados ao utilizar estes métodos em vez dos baseados em simulação. O método Integrated Nested Laplace Approximations (INLA), aplicado na importante classe de modelos que envolvem Campos Aleatórios Markovianos Gaussianos (CAMG), é descrito e, através de exemplificação, uma discussão qualitativa sobre o método é apresentada. Foi proposta a utilização do INLA para realização de inferência em modelos dinâmicos Bayesianos para processos pontuais espaço-temporais ao invés da abordagem usual que utiliza Markov Chain Monte Carlo (MCMC). Por fim, é apresentada uma importante extensão do INLA, onde a dependência entre o conjunto de dados e o campo latente, da forma como apresentada na descrição do INLA, é generalizada.

Palavras-chave. Estatística Bayesiana, Aproximações Determinísticas, INLA, Processos Pontuais.

Abstract

Deterministic Approximations to Marginal Posterior Distributions

Thiago Guerrera Martins

Abstract: A major challenge in Bayesian statistics is to obtain approximations to marginal posterior distributions in an accurate and efficient manner in cases where there is no analytic solution available. A review of the literature of deterministic methods for this purpose is presented in this dissertation in a general context, and it is shown that there are models of current interest that are better estimated by using these methods instead of those based on simulation. The method Integrated Nested Laplace approximations (INLA), designed for the important class of models involving Gaussian Markov Random Field (GMRF), is described and, through examples, a qualitative discussion about the method is presented. The INLA method is proposed to perform Bayesian inference in dynamic models for space-time point processes rather than the usual approach that uses Markov Chain Monte Carlo (MCMC). Finally, we present an important extension of INLA, where the dependency between the data set and the latent field, the way it is presented in the description of INLA, is generalized.

Keywords. Bayesian Statistics, Deterministic Approximations, INLA, Point Processes.

Sumário

1	Introdução	1
1.1	Introdução	1
1.2	Notação	4
1.3	Nota Computacional	5
1.4	Matrizes esparsas	6
1.5	Grafos não direcionados	7
1.6	Distribuição Normal Multivariada e suas propriedades	7
1.7	Distribuições	8
1.7.1	Distribuição Uniforme	8
1.7.2	Distribuição Gama	9
1.7.3	Distribuição Gama-Inversa	9
1.7.4	Distribuição Beta	9
1.8	Processo Gaussiano	10
1.9	Teorema de Taylor	10
1.10	Algoritmo Newton-Raphson	11
1.11	Aproximação Gaussiana	13
1.11.1	Caso Univariado	13
1.11.2	Caso Multivariado	13
1.11.3	Caso Específico de interesse	14
1.11.4	Escolha do ponto x_0	14
2	Aproximações determinísticas	17
2.1	Revisão da literatura	18

2.1.1	Abordagem de Reilly	18
2.1.2	Quadratura de Gauss-Hermite	19
2.1.3	Aproximação de Laplace	21
2.2	Parametrização adequada de Ψ	21
2.3	Exploração da grade	24
2.4	Aplicação: Modelo de Black-Scholes Fracionário	26
2.4.1	Simulação e Resultados	28
3	INLA e aplicações	32
3.1	Introdução	33
3.2	Método INLA	34
3.2.1	Objetivos	34
3.2.2	Aproximação para $\pi(\boldsymbol{\theta} \mathbf{y})$	35
3.2.3	Aproximação para $\pi(x_i \boldsymbol{\theta}, \mathbf{y})$	36
3.3	Aplicações	39
3.4	Discussão sobre o INLA	49
4	Processos Pontuais Espaço-Temporais: Inferência Bayesiana Aproximada	53
4.1	Processos Pontuais Espaço-Temporais	54
4.1.1	Formulação geral do modelo	54
4.1.2	Formulação do modelo discreto	55
4.2	Modelos Dinâmicos Bayesianos para Processos Pontuais Espaço-Temporais	56
4.3	Inferência Bayesiana Aproximada: Aspectos Computacionais	58
5	INLA - Extensões	67
5.1	Introdução	68
5.1.1	Modelo de Volatilidade Estocástica Assimétrica	68
5.2	Extensões	70
5.2.1	Aproximação Gaussiana	71
5.2.2	Aproximação de Laplace Simplificado	74

5.3	Modelo de Volatilidade Estocástica Assimétrica - Resultados	76
6	Discussão Final	80

Capítulo 1

Introdução

1.1 Introdução

Em inferência Bayesiana, usualmente nos deparamos com o seguinte contexto: seja $l(\mathbf{y}; \Psi)$ a função de verossimilhança de Ψ proveniente de observações \mathbf{y} e $\pi(\Psi)$ a densidade a priori; utilizando o Teorema de Bayes temos que a distribuição a posteriori de Ψ é dada por

$$\pi(\Psi|\mathbf{y}) = \frac{l(\mathbf{y}; \Psi)\pi(\Psi)}{\int l(\mathbf{y}; \Psi)\pi(\Psi)d\Psi}. \quad (1.1)$$

Se estamos interessado na densidade marginal de $\Psi_{\mathbf{I}}$, onde $\mathbf{I} \subseteq \{1, \dots, k\}$ são os índices dos componentes de interesse, sendo k a dimensão do vetor paramétrico Ψ , nós temos simplesmente que integrar sob $\Psi_{\mathbf{I}'}$, onde \mathbf{I}' é o complemento de \mathbf{I} com relação a $\{1, \dots, k\}$, para obter

$$\pi(\Psi_{\mathbf{I}}|\mathbf{y}) = \int \pi(\Psi|\mathbf{y})d\Psi_{\mathbf{I}'}. \quad (1.2)$$

A partir das equações (1.1) e (1.2) pode-se obter diversas quantidades de interesse, como esperanças, variâncias, entre outras. Do mesmo modo como feito em Naylor & Smith (1982), vamos definir o operador $S_{\mathbf{I}}$ tal que

$$S_{\mathbf{I}}(q(\Psi)) = \int q(\Psi)l(\mathbf{y}; \Psi)\pi(\Psi)d\Psi_{\mathbf{I}'}, \quad (1.3)$$

escrevendo S no lugar de $S_{\mathbf{I}}$ se $\mathbf{I}' = \{1, \dots, k\}$ e deixando a dimensão do espaço de integração definida implicitamente e o intervalo de integração sendo todo o espaço paramétrico. Com isso, todas as integrais necessárias para calcular e resumir as equações (1.1) e (1.2) são casos especiais de (1.3) para escolhas particulares de \mathbf{I} e $q(\Psi)$. Como exemplo temos que o denominador de (1.1) é dado por $S(1)$, a equação (1.2) é dada por $S_{\mathbf{I}}(1)/S(1)$ e a média a posteriori $E_{\Psi|\mathbf{y}}(\Psi) = S(\Psi)/S(1)$.

Com isso, podemos concluir que um dos maiores desafios na aplicação da inferência Bayesiana é obter a solução de integrais do tipo (1.3) de forma precisa e eficiente. Naturalmente, nos casos onde $l(\mathbf{y}; \Psi)$ e $\pi(\Psi)$ pertencem à família exponencial e à sua família conjugada correspondente, respectivamente, as integrais contidas em (1.1) e (1.2) podem ser calculadas analiticamente. Porém, fora desse contexto, como geralmente ocorre na prática, tal forma analítica não é, em geral, possível de se obter, o que faz com que métodos de aproximações sejam necessários para obter a solução dessas integrais. Com isso, não é de se estranhar que uma grande quantidade de artigos foi elaborada propondo diferentes métodos para a solução de integrais dentro do contexto de inferência Bayesiana.

Até o final da década de 80, a atenção estava voltada para aproximações numéricas e analíticas, bem representadas por (Reilly, 1976; Naylor & Smith, 1982; Tierney & Kadane, 1986; Smith *et al.*, 1987; Tierney *et al.*, 1989) para citar alguns autores. Tais aproximações foram um grande avanço ao permitir que usuários de inferência Bayesiana usassem prioris e modelos mais próximos da realidade do problema em questão. Naturalmente, os métodos tem suas limitações, a principal sendo que o tempo computacional necessário para os cálculos cresce drasticamente à medida que a dimensão paramétrica aumenta.

Os avanços (e problemas) com aproximações determinísticas foram deixados de lado a partir da década de 90, quando os influentes artigos de Gelfand & Smith (1990) e Gordon *et al.* (1993) mostraram a utilidade da aborgagem baseada em amostragem para aproximar integrais multi-dimensionais no contexto Bayesiano, sendo que o segundo artigo visava problemas sequenciais.

Nas duas décadas que se seguiram, a comunidade científica interessada em obter soluções para (1.3) viu nessa nova corrente uma área promissora de pesquisa, o que levou a uma grande quantidade de resultados relacionados a essa idéia serem publicados para diferentes contextos específicos (ver por exemplo, Robert & Casella, 2004; Gamerman & Lopes, 2006).

Do mesmo modo como antes, a metodologia de aproximações baseadas em amostragem, interessante como é, tem também várias desvantagens, como por exemplo o fato de ser extremamente difícil diagnosticar convergência, o custo computacional ser proibitivamente caro em determinadas situações (modelos espaciais e espaço-temporais para citar dois exemplos) e erros na média de Monte Carlo serem aditivos e de ordem $O(N^{-1/2})$, onde N é o número de iterações, o que significa que é necessário 100 vezes a mais de tempo computacional para se obter um dígito correto a mais na estimativa. Consequentemente, temos que é fácil obter estimativas brutas, mas quase impossível obtê-las de forma precisa.

O capítulo 2 irá revisitar a literatura de aproximações numéricas e determinísticas para obtenção de distribuições a posteriori marginais em um contexto geral, abordando o assunto de forma estruturada a fim de que tais métodos possam ser aplicados de forma mais eficiente, e por fim, irá demonstrar que modelos de interesse atual podem ser melhor estimados utilizando técnicas determinísticas ao invés das baseadas em simulação, tanto com relação a tempo computacional e precisão quanto com relação à facilidade de aplicação do método, característica de extrema importância, principalmente quando utilizada por não especialistas em estatística.

Rue *et al.* (2009) trouxeram de volta a atenção para aproximações determinísticas com o método *Integrated Nested Laplace Approximations (INLA)*, que permite fazer inferência Bayesiana aproximada para a classe de modelos Gaussianos latentes de maneira rápida e precisa. Tal metodologia teve um grande impacto na comunidade da estatística Bayesiana pois a classe em questão engloba grande parte dos modelos Bayesianos estruturados, como por exemplo, modelos de regressão, modelos dinâmicos, modelos espaciais

e espaço-temporais. O capítulo 3 irá descrever a metodologia e, através de exemplos, explicitar detalhes que não foram abordados de forma detalhada em Rue *et al.* (2009). O capítulo 4 irá propor a utilização do INLA para realização de inferência em modelos dinâmicos Bayesianos para processos pontuais espaço-temporais (Reis, 2008; Reis *et al.*, 2010) ao invés da abordagem usual que utiliza *Markov Chain Monte Carlo* (MCMC).

O método INLA proposto em Rue *et al.* (2009) assume que cada elemento do vetor de dados \mathbf{y} está conectado a somente um elemento do campo latente \mathbf{x} , o que impossibilita sua aplicação a diversos modelos encontrados na prática. O capítulo 5 é dedicado a uma extensão do INLA, onde tal restrição entre a dependência entre o conjunto de dados \mathbf{y} e o campo latente \mathbf{x} é eliminada.

Por fim, o capítulo 6 irá apresentar uma discussão final sobre o material apresentado nesta dissertação e discutir deficiências e problemas que ainda persistem na teoria de aproximação determinística para distribuições a posteriori marginais, o que mostra ser um ótimo campo para trabalhos futuros.

1.2 Notação

Vetores e matrizes serão denotados em caracter negrito, sendo que vetores são representados com letra minúscula, \mathbf{a} , enquanto matrizes serão definidas com letras maiúsculas, \mathbf{A} . \mathbf{a}_{-i} representa o vetor \mathbf{a} com o elemento i excluído, ou seja, se $\mathbf{a} = (a_1, \dots, a_n)^T$ então $\mathbf{a}_{-i} = (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n)^T$. $\mathbf{A}_{[-i, -i]}$ representa a matriz \mathbf{A} com a i -ésima linha e a i -ésima coluna removidas. \mathbf{A}^T representa o transposto da matriz \mathbf{A} . Os símbolos $\mathbf{0}$ e $\mathbf{1}$ representam um vetor de zeros e um vetor de uns, respectivamente, onde a dimensão estará implícita no contexto onde se insere a notação. A dimensão das integrais estará implicitamente definida e o intervalo de integração será sempre todo o espaço paramétrico, a menos que seja explicitamente definido de outro modo. A n -ésima derivada de uma função $f(x)$ em relação a x será denotada por $f^{(n)}(x)$. Assim, $f^{(n)}(x_0)$ será a n -ésima derivada avaliada no ponto x_0 .

$f(x) \propto g(x)$ significa que $f(x)$ é igual a $g(x)$ a menos de uma constante multiplicativa, de modo que $f(x) = \text{constante} \times g(x)$ e $f(x) \overset{c}{\propto} g(x)$ significa que $f(x)$ é igual a $g(x)$ a menos de uma constante aditiva, de modo que $f(x) = \text{constante} + g(x)$. O símbolo \approx significa aproximadamente e $x \perp y$ significa que a variável aleatória x é independente da variável aleatória y . **dom** f representa o domínio da função f . Sejam f e g duas funções definidas em um intervalo comum, possivelmente infinito. Seja z_0 um ponto nesse intervalo (pode ser $-\infty$ ou ∞). É necessário que $g(z) \neq 0$ para todo $z \neq z_0$ em uma vizinhança de z_0 . Então, temos que

$$f(z) = O(g(z))$$

se existe uma constante M tal que $|f(z)| \leq M|g(z)|$ quando $z \rightarrow z_0$. Por exemplo, $\frac{n+1}{3n^2} = O(n^{-1})$, e fica entendido que está sendo considerado $n \rightarrow \infty$.

1.3 Nota Computacional

Todos os exemplos nesta dissertação foram programados pelo autor utilizando os softwares (gratuitos) R (Bates *et al.*, 2004) ou WinBUGS (Lunn *et al.*, 2000), e os cálculos foram realizados em um laptop com um único processador 2.1-GHz.

O objetivo principal do autor nesta dissertação ao utilizar programação própria foi adquirir pleno conhecimento dos métodos utilizados e não o de elaborar algoritmos eficientes, apesar de grande esforço ter sido colocado na tentativa de programar de forma inteligente. Conseqüentemente, só haverá comparação de tempo computacional entre métodos distintos quando as operações necessárias para ambos os métodos forem semelhantes. Suponha, por exemplo, que a maior parte do custo computacional dos métodos esteja na avaliação da função de verossimilhança, como ocorre na seção 2.4, então ambos os métodos seriam beneficiados caso houvesse um ganho de eficiência nesta etapa do algoritmo. Caso os métodos a serem comparados sejam de natureza distinta e/ou programados em plataformas diferentes, como ocorre

nos capítulos 3 e 5, será evitado a comparação de tempo computacional de forma precisa, sendo indicado quando há uma diferença em ordem de magnitude entre os métodos, como por exemplo, no caso em que um algoritmo realiza em minutos o que o outro levaria horas para realizar.

Para a abordagem do MCMC, foi escolhido utilizar o WinBUGS para estimar o modelo de volatilidade estocástica no capítulo 3 e o modelo de volatilidade estocástica assimétrica no capítulo 5 porque tal programa exige esforço parecido, por parte do usuário, ao necessário para implementar o INLA da forma como disponibilizado em <http://www.r-inla.org/>. Para não utilizar o WinBUGS (e escapar de suas ineficiências) é necessário um tempo considerável de programação a fim de elaborar algoritmos específicos para tais modelos, o que não faz parte dos objetivos desta dissertação.

1.4 Matrizes esparsas

Uma matriz é dita ser esparsa quando tem uma quantidade relativamente pequena de elementos não-nulos. Tais matrizes aparecem com frequência em diversas áreas do conhecimento, como por exemplo física, engenharia e matemática. Nesta dissertação serão abordados diversos problemas em estatística onde esse tipo de matriz estará presente, mais especificamente em problemas envolvendo campos aleatórios Markovianos Gaussianos, abreviado por CAMG (ver capítulo 3). Deste modo, parece razoável utilizar a vasta literatura existente sobre métodos numéricos para matrizes esparsas na implementação dos algoritmos utilizados dentro desse contexto, já que ganhos são obtidos tanto em armazenamento de dados quanto na manipulação das matrizes. Como citado em Rue & Held (2005), é recomendado que a tarefa de construir e implementar algoritmos numéricos para matrizes esparsas seja deixado a cargo de especialistas da ciência da computação e análise numérica. No entanto, estatísticos devem usar tais resultados e bibliotecas para elaborar programas estatísticos eficientes. Tais algoritmos estão eficientemente implementados em diversas bibliotecas, como por exemplo em Bates & Maechler

(2007) e Rue & Follestad (2002). Para mais detalhes sobre matrizes esparsas, sugere-se consultar George & Liu (1981) e Duff *et al.* (1989).

1.5 Grafos não direcionados

Iremos usar grafos não-direcionados para representar a estrutura de independência condicional em um CAMG (ver capítulo 3). Um grafo não direcionado G é uma dupla $G = (V, E)$, onde V é o conjunto de nós (ou elementos) em um grafo, e E é o conjunto de ligações $\{i, j\}$, onde $i, j \in V$ e $i \neq j$. Se $\{i, j\} \in E$, 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 $i \neq j$.

1.6 Distribuição Normal Multivariada e suas propriedades

Definição 1.1. (Distribuição Normal Multivariada) Um vetor aleatório $\mathbf{x} = (x_1, \dots, x_n)$ tem distribuição Normal multivariada com média $\boldsymbol{\mu}$ e matriz de variância-covariância $\boldsymbol{\Sigma}$, denotada por $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, se sua densidade pode ser descrita por

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

Vamos dividir \mathbf{x} em duas partes, $\mathbf{x} = (\mathbf{x}_A^T, \mathbf{x}_B^T)$ e dividir $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ de modo que

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

A seguir, algumas propriedades da distribuição Normal multivariada serão apresentadas:

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

2. $\Sigma_{AB} = 0$ se e somente se \mathbf{x}_A e \mathbf{x}_B são independentes;
3. A distribuição condicional de $(\mathbf{x}_A|\mathbf{x}_B)$ é $N(\boldsymbol{\mu}_{A|B}, \Sigma_{A|B})$, onde

$$\begin{aligned}\boldsymbol{\mu}_{A|B} &= \boldsymbol{\mu}_A + \Sigma_{AB}\Sigma_{BB}^{-1}(\mathbf{x}_B - \boldsymbol{\mu}_B) \\ \Sigma_{A|B} &= \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA};\end{aligned}$$

4. Se $\mathbf{x} \sim N(\boldsymbol{\mu}, \Sigma)$ e $\mathbf{x}' \sim N(\boldsymbol{\mu}', \Sigma')$ são independentes, então $\mathbf{x} + \mathbf{x}' \sim N(\boldsymbol{\mu} + \boldsymbol{\mu}', \Sigma + \Sigma')$;
5. Seja $\Sigma = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ uma decomposição de Σ onde as colunas de \mathbf{U} são autovetores unitários e $\boldsymbol{\Lambda}$ uma matrix diagonal de autovalores. Então temos que

$$\mathbf{x} \sim N(\boldsymbol{\mu}, \Sigma) \iff \mathbf{x} = \boldsymbol{\mu} + \mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{z},$$

onde $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I})$;

6. Seja $\mathbf{x} \sim N(\boldsymbol{\mu}, \Sigma)$ e $\mathbf{Q} = \Sigma^{-1}$ a matrix de precisão de \mathbf{x} , então para $i \neq j$,

$$x_i \perp x_j | \mathbf{x}_{-ij} \iff Q_{ij} = 0.$$

1.7 Distribuições

Nesta seção serão definidas algumas distribuições de probabilidade que serão utilizadas ao longo desta dissertação.

1.7.1 Distribuição Uniforme

Seja X uma variável aleatória com distribuição Uniforme com parâmetros a e b , denotada por $U[a, b]$. Sua função de densidade de probabilidade é dada

por

$$\pi(x) = \frac{1}{b-a}, \quad (1.4)$$

onde $a \leq x \leq b$, $-\infty < a < b < \infty$.

1.7.2 Distribuição Gama

Seja X uma variável aleatória com distribuição Gama com parâmetros α e β , denotada por $G(\alpha, \beta)$. Sua função de densidade de probabilidade é dada por

$$\pi(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad (1.5)$$

onde $x \geq 0$, $\alpha, \beta > 0$ e Γ é a função Gama.

1.7.3 Distribuição Gama-Inversa

Seja X uma variável aleatória com distribuição Inversa-Gama com parâmetros α e β , denotada por $GI(\alpha, \beta)$. Sua função de densidade de probabilidade é dada por

$$\pi(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\beta/x}, \quad (1.6)$$

onde $x, \alpha, \beta > 0$ e Γ é a função Gama. Temos que

$$X \sim G(\alpha, \beta) \iff \frac{1}{X} \sim GI(\alpha, \beta). \quad (1.7)$$

1.7.4 Distribuição Beta

Seja X uma variável aleatória com distribuição Beta com parâmetros α e β , denotada por $B(\alpha, \beta)$. Sua função de densidade de probabilidade é dada por

$$\pi(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad (1.8)$$

onde $0 < x < 1$, $\alpha, \beta > 0$ e Γ é a função Gama.

1.8 Processo Gaussiano

O processo Gaussiano é definido como o processo estocástico $x(\cdot)$ na região $D \in \mathfrak{R}^d$, com d inteiro e D fixo e contínuo, tal que, para $n \geq 1$ e localizações espaciais s_1, \dots, s_n , o vetor $(x(s_1), \dots, x(s_n))$ tem distribuição Normal multivariada com vetor de médias \mathbf{m} e matriz de variância-covariância Σ .

As suposições usuais são:

- **estacionariedade**, que implica que $\mathbf{m} = \mu \mathbf{1}$ e $\Sigma = \sigma^2 \mathbf{R}$, onde $\mu \in \mathfrak{R}$, $\sigma^2 > 0$ e \mathbf{R} é uma matriz de correlações tal que $r_{ij} = \rho_\kappa(s_i - s_j) = \rho(s_i - s_j; \kappa)$ para uma função de autocorrelação adequada ρ ;
- **isotropia**, que implica que a função de correlação ρ_κ depende apenas da distância $\|s_i - s_j\|$ entre as localizações s_i e s_j .

A notação

$$x(\cdot) | \mu, \sigma^2, \kappa \sim PG[\mu; \sigma^2; \rho(\cdot; \kappa)] \quad (1.9)$$

será utilizada neste texto para denotar um processo Gaussiano estacionário e isotrópico com média μ , variância σ^2 e função de correlação espacial ρ .

1.9 Teorema de Taylor

O teorema de Taylor fornece uma aproximação polinomial para uma função f e tal aproximação será amplamente utilizada ao longo desta dissertação.

Teorema 1.2. (*Expansão de Taylor*) *Seja $n > 0$ um inteiro e f uma função n vezes continuamente diferenciável no intervalo fechado $[a, x]$ e $n + 1$ vezes diferenciável no intervalo aberto (a, x) , então*

$$f(x) = f(a) + \frac{f^{(1)}(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x).$$

onde

$$R_n(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x-x_0)^{n+1}$$

para algum ponto ξ no intervalo entre x e x_0 .

Desse modo, o teorema de Taylor fornece, além da aproximação polinomial, uma expressão para o erro da aproximação. Como exemplo, temos que, ao utilizar a aproximação de Taylor até o termo de segunda ordem para uma função de forma quadrática, o erro da aproximação será zero, pois nesse caso, $f^{(3)}(\xi) = 0$ para todo ξ entre x e x_0

1.10 Algoritmo Newton-Raphson

Essa seção irá descrever o algoritmo de Newton-Raphson, que é um dos vários métodos usados para resolver um problema de otimização sem restrição, ou seja, o objetivo dessa seção é maximizar $f(\mathbf{x})$ onde $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$ é côncava e duas vezes continuamente diferenciável. No que vem a seguir ∇f e $\nabla^2 f$ são o gradiente e a matriz Hessiana de f , respectivamente.

O algoritmo de Newton-Raphson, em sua forma mais simples, pode ser descrito como:

Algorithm 1. (*Newton-Raphson*)

Dado um valor inicial $\mathbf{x} \in \mathbf{dom} f$.

repita

1. Calcule o passo de Newton. $\Delta \mathbf{x} = -\nabla^2 f(\mathbf{x})^{-1} \nabla f(\mathbf{x})$;
2. Atualize $\mathbf{x}_1 = \mathbf{x}_0 + \Delta \mathbf{x}$
3. Cheque a regra de parada.

A regra de parada utilizada nos exemplos desta dissertação é dada por

$$\mathbf{x}_1 - \mathbf{x}_0 < \epsilon \mathbf{1}, \tag{1.10}$$

onde $\epsilon = 0,001$.

No entanto, dependendo da complexidade do problema, há casos onde o algoritmo acima pode demorar a convergir ou até mesmo divergir dependendo do valor inicial escolhido. Quando tais problemas acontecem, usualmente é porque as primeiras iterações do algoritmo não estão aumentando f de forma 'suficiente'. Para lidar com esse possível problema é necessário tornar

o algoritmo acima mais robusto, ao acrescentar o método *busca em linha* (do inglês *Line-Search*), como a seguir

Algorithm 2. (*Newton-Raphson com busca em linha*)

Dado um valor inicial $\mathbf{x} \in \mathbf{dom} f$.

repita

1. Calcule o passo de Newton. $\Delta\mathbf{x} = -\nabla^2 f(\mathbf{x})^{-1}\nabla f(\mathbf{x})$;
2. Busca em linha. Escolha o tamanho do passo $t > 0$;
3. Atualize $\mathbf{x}_1 = \mathbf{x}_0 + t\Delta\mathbf{x}$
4. Cheque a regra de parada.

O segundo item é chamado de *busca em linha* já que a seleção do tamanho do passo t determina aonde ao longo da linha $\{x + t\Delta x_{nt} | t \in \mathbb{R}_+\}$ estará a nova iteração.

Há mais de um algoritmo do tipo busca em linha, mas o seguinte, chamado de *busca em linha retroativo* (do inglês *backtracking line search*), é o que será usado nessa dissertação:

Algorithm 3. (*Busca em linha retroativa*)

Dado o passo $\Delta\mathbf{x}$ para f em $\mathbf{x} \in \mathbf{dom} f$, $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$.

1. $t := 1$
2. **enquanto** $f(\mathbf{x} + t\Delta\mathbf{x}) > f(\mathbf{x}) + \alpha t\nabla f(\mathbf{x})^T$, $t := \beta t$.

Esse método é chamado de *busca em linha retroativo* porque começa com o passo de tamanho unitário ($t = 1$) e começa a reduzi-lo com um fator β até a condição de parada $f(\mathbf{x} + t\Delta\mathbf{x}) > f(\mathbf{x}) + \alpha t\nabla f(\mathbf{x})^T$ ser satisfeita. O parâmetro α é tipicamente escolhido entre 0,01 e 0,3, o que significa que aceita-se um aumento de f entre 1% e 30% da previsão baseada na extrapolação linear da função no ponto atual. O parâmetro β é usualmente escolhido entre 0,1, que corresponde a uma busca mais grosseira, e 0,8, que corresponde a uma busca mais detalhada. Para mais detalhes sobre métodos de otimização sugere-se Boyd & Vandenberghe (2004).

1.11 Aproximação Gaussiana

A seguir será descrito como aproximar uma função de densidade $\pi(x)$ por uma função de densidade Gaussiana. Esta aproximação é a base da metodologia apresentada nos capítulos 3 e 5.

1.11.1 Caso Univariado

Uma aproximação Gaussiana para uma densidade da forma

$$\pi(x) \propto \exp\{f(x)\}, \quad (1.11)$$

onde $f(x)$ satisfaz as condições da seção 1.9, pode ser obtida do seguinte modo: construa uma expansão de Taylor de $f(x)$ ao redor de um valor escolhido x_0 até o termo de segunda ordem

$$f(x) \approx f(x_0) + f^{(1)}(x_0)(x - x_0) + \frac{f^{(2)}(x_0)}{2}(x - x_0)^2 \quad (1.12)$$

$$\stackrel{c}{\propto} b(x_0)x - \frac{1}{2}c(x_0)x^2 \quad (1.13)$$

onde $b(x_0) = f^{(1)}(x_0) - f^{(2)}(x_0)x_0$ e $c(x_0) = -f^{(2)}(x_0)$. Agora basta substituir (1.13) em (1.11) de modo a obter

$$\tilde{\pi}_G(x) \propto \exp \left\{ -\frac{1}{2}c(x_0)x^2 + b(x_0)x \right\} \quad (1.14)$$

onde $\tilde{\pi}_G(x)$ é uma aproximação Gaussiana de $\pi(x)$ e, portanto, tem precisão $c(x_0)$ e média $b(x_0)/c(x_0)$.

1.11.2 Caso Multivariado

A extensão para o caso multivariado é trivial, bastando substituir a derivada primeira e a derivada segunda de f pelo gradiente (∇f) e pela matriz Hessiana ($\mathbf{H}f$) de f respectivamente, de modo que

$$\tilde{\pi}_G(\mathbf{x}) \propto \exp \left\{ -\frac{1}{2}\mathbf{x}^T \mathbf{C}(\mathbf{x}_0)\mathbf{x} + \mathbf{b}(\mathbf{x}_0)\mathbf{x} \right\} \quad (1.15)$$

é uma aproximação Gaussiana com matrix de precisão $\mathbf{C}(\mathbf{x}_0)$ e média $\mathbf{C}^{-1}(\mathbf{x}_0)\mathbf{b}(\mathbf{x}_0)$, onde $\mathbf{b}(\mathbf{x}_0) = \nabla f(\mathbf{x}_0) - \mathbf{H}f(\mathbf{x}_0)\mathbf{x}_0$ e $\mathbf{C}(\mathbf{x}_0) = -\mathbf{H}f(\mathbf{x}_0)$.

1.11.3 Caso Específico de interesse

Muitas vezes é de interesse obter uma aproximação Gaussiana para densidades da forma

$$\pi(\mathbf{x}) \propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \sum_{i \in \mathbf{I}} g_i(x_i) \right\} \quad (1.16)$$

Nesse caso a expansão de Taylor até os termos de segunda ordem no ponto x_0^i efetuada em $g_i(x_i)$ fornece

$$g_i(x_i) \approx a_i(x_0^i) + b_i(x_0^i)x_i - \frac{1}{2}c_i(x_0^i)x_i^2 \quad (1.17)$$

$$\stackrel{c}{\propto} b_i(x_0^i)x_i - \frac{1}{2}c_i(x_0^i)x_i^2 \quad (1.18)$$

onde do mesmo modo como no caso univariado

$$b_i(x_0^i) = g_i^{(1)}(x_0^i) - g_i^{(2)}(x_0^i)x_0^i \quad e \quad c_i(x_0^i) = -g_i^{(2)}(x_0^i). \quad (1.19)$$

Logo,

$$\sum_{i \in \mathbf{I}} g_i(x_i) \approx \sum_{i \in \mathbf{I}} b_i(x_0^i)x_i - \frac{1}{2}c_i(x_0^i)x_i^2 \quad (1.20)$$

$$= -\frac{1}{2} \mathbf{x}^T \mathbf{C}(\mathbf{x}_0) \mathbf{x} + \mathbf{b}(\mathbf{x}_0) \mathbf{x} \quad (1.21)$$

onde $\mathbf{C}(\mathbf{x}_0)$ é uma matriz diagonal cujos elementos $c_i(x_0^i)$ estão definidos em (1.19) e $c_i(x_0^i) = 0$ caso $i \notin \mathbf{I}$. Similarmente, $\mathbf{b}(\mathbf{x}_0)$ é um vetor com elementos $b_i(x_0^i)$ definidos em (1.19) e $b_i(x_0^i) = 0$ caso $i \notin \mathbf{I}$. Substituindo (1.21) em (1.16) temos que

$$\tilde{\pi}_G(\mathbf{x}) \propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T (\mathbf{Q} + \mathbf{C}(\mathbf{x}_0)) \mathbf{x} + \mathbf{b}(\mathbf{x}_0) \mathbf{x} \right\} \quad (1.22)$$

é uma aproximação Gaussiana para $\pi(\mathbf{x})$ com matrix de precisão $\mathbf{Q}^*(\mathbf{x}_0) = (\mathbf{Q} + \mathbf{C}(\mathbf{x}_0))$ e média $\mathbf{Q}^{*-1}(\mathbf{x}_0)\mathbf{b}(\mathbf{x}_0)$.

1.11.4 Escolha do ponto \mathbf{x}_0

Agora basta definir qual o melhor ponto x_0 para efetuar a expansão de Taylor nas seções anteriores. Essa escolha pode ser melhor entendida através

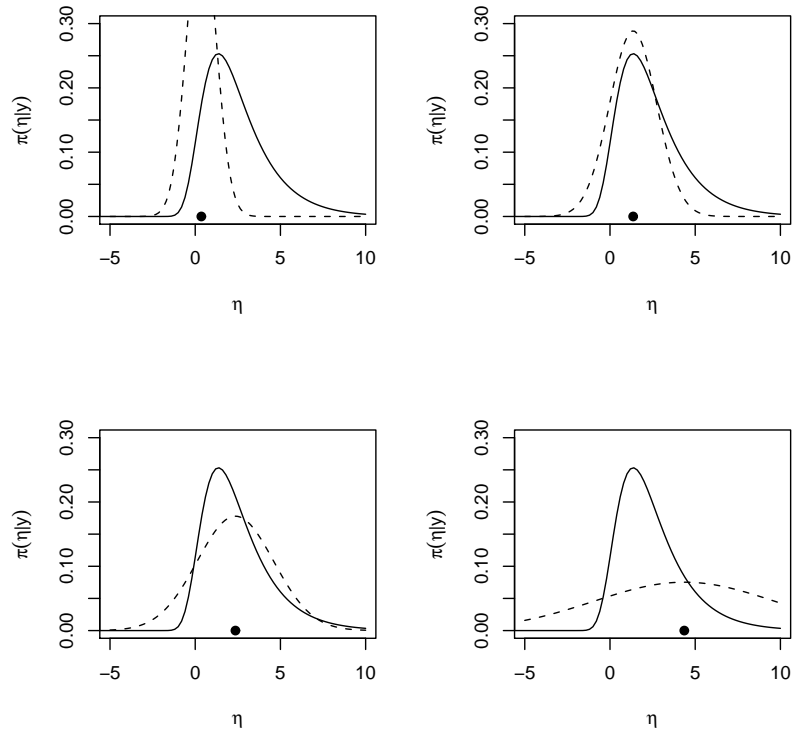


Figura 1.1: Aproximação Gaussiana para a verossimilhança de η para diferentes valores de x_0 : (linha solida) Verossimilhança de η (linha pontilhada) Aproximação Gaussiana.

do seguinte exemplo: Suponha que temos uma única observação y de uma distribuição Normal com média 0 e variância $\exp(\eta)$ e η possui uma priori Normal com média 0 e variância 100.

$$y \sim N(0, \exp(\eta)); \quad \eta \sim N(0, 100);$$

$$y = 2$$

Agora vamos efetuar uma aproximação Gaussiana para a posteriori (não-Gaussiana)

$$\pi(\eta|y) \propto \exp \left\{ -\frac{1}{2} \left(\frac{\eta^2}{100} + \frac{y^2}{e^\eta} + \eta \right) \right\} \quad (1.23)$$

para diferentes valores x_0 em (1.14). Olhando a figura 1.1, pode-se concluir que quanto mais perto x_0 está da moda da distribuição $\pi(\eta|y)$ melhor é a aproximação Gaussiana. O mesmo é válido para o caso multivariado. Assim, vimos que em muitos casos de interesse onde a distribuição a ser aproximada é unimodal, faz sentido escolher a moda da distribuição para efetuar uma aproximação Gaussiana.

Capítulo 2

Aproximações determinísticas

A seção 2.1 contém uma pequena revisão da literatura sobre aproximação determinística para distribuições a posteriori marginais. Porém, os métodos mencionados são os que o autor achou necessário para a compreensão do conteúdo que se segue. Com isso, tal revisão está longe de ser exaustiva. Para uma revisão um pouco mais completa, porém ainda longe de cobrir todos os métodos que foram desenvolvidos em aproximação determinística, sugere-se Smith *et al.* (1985). A seção 2.2 irá abordar o tema reparametrização com o objetivo de tornar o problema "mais bem comportado", de forma a tornar as aproximações determinísticas mais eficientes. A seção 2.3 irá descrever como construir uma grade (do inglês *grid*) de forma ótima, de modo que a escolha dos pontos será feita levando em consideração as especificidades do problema em questão. Por fim, a seção 2.4 irá chamar a atenção, através de um exemplo, para o fato que nem sempre métodos de simulação estocástica são a melhor opção para realizar inferência Bayesiana aproximada (apesar de serem os mais disseminados em computação Bayesiana atualmente).

2.1 Revisão da literatura

2.1.1 Abordagem de Reilly

Reilly (1976) abordou o problema de obter distribuições a posteriori marginais usando discretização dos parâmetros envolvidos, avaliando a posteriori em diversos pontos de uma grade e substituindo as integrais por somatórios para achar a constante normalizadora de (1.1) e as marginais de (1.2). Na época, uma das grandes vantagens desse método era a possibilidade de tratar problemas independente da forma e do tipo do modelo, descartando a necessidade de linearizar modelos não-lineares por exemplo, e também o fato de eliminar restrições com relação às prioris usadas, pois deixou de ser necessário escolher distribuições a priori com o intuito de facilitar as contas necessárias para a implementação do método, como ocorre em certos casos na implementação de esquemas de MCMC por exemplo. Esse foi um primeiro passo para que usuários de inferência Bayesiana pudessem ajustar modelos mais próximos da realidade ao invés de se restringirem aos modelos onde a solução podia ser obtida analiticamente. No entanto, a grande desvantagem do método é o alto custo computacional, tanto com relação ao tempo para executar os cálculos quanto com relação à demanda de armazenamento, ambos crescendo exponencialmente de acordo com a dimensão k do vetor paramétrico Ψ . Além disso, essa técnica não é eficaz nos casos onde há alta correlação no vetor paramétrico, casos esses em que a escolha da grade tem que ser feita de forma mais cuidadosa do que a apresentada em Reilly (1976) (onde a grade era obtida por tentativa e erro até se obter a região que contém toda a massa de probabilidade) para que não se gaste tempo computacional avaliando funções em pontos de baixa densidade. Isso sendo de extrema importância em problemas onde o custo de se avaliar uma função não é negligenciável e/ou quando o tamanho do vetor paramétrico é alto.

2.1.2 Quadratura de Gauss-Hermite

Inicialmente, suponha Ψ univariado, Naylor & Smith (1982) usaram o fato que, para uma grande classe de problemas, $l(\mathbf{y}; \Psi)$ e $\pi(\Psi)$ satisfazem condições de regularidade que garantem que a densidade a posteriori é aproximadamente normal. Com isso, através de uma parametrização adequada de Ψ (ver seção 2.2) é razoável supor que, para tamanhos de amostra moderados ou grandes, a forma de $l(\mathbf{y}; \Psi)\pi(\Psi)$ pode ser adequadamente aproximada pelo produto de uma densidade normal e um polinômio em Ψ . Ou seja, aproximada por

$$g(\Psi) = h(\Psi)(2\pi\sigma^2)^{-1/2} \exp \left\{ -\frac{1}{2} \left(\frac{\Psi - \mu}{\sigma} \right)^2 \right\} \quad (2.1)$$

onde $h(\Psi)$ é um polinômio. Desse modo, teríamos que

$$\int g(\Psi) d\Psi = \int \frac{1}{\sqrt{\pi}} h(\mu + \sqrt{2}\sigma\Psi) e^{-\Psi^2} d\Psi \quad (2.2)$$

É fato conhecido que integrais da forma (2.2) podem ser aproximadas usando quadratura de Gauss-Hermite, o que nos leva a

$$\int g(\Psi) d\Psi \approx \sum_{i=1}^n m_i g(z_i), \quad (2.3)$$

onde

$$m_i = w_i \exp(t_i^2) \sqrt{2}\sigma, \quad z_i = \mu + \sqrt{2}\sigma t_i \quad (2.4)$$

Valores de t_i e w_i estão disponíveis para $n = 1, \dots, 20$ (Salzer *et al.*, 1952) e o erro da aproximação será pequeno caso $h(z)$ seja aproximadamente um polinômio e será zero se $h(z)$ for um polinômio. Resta apenas encontrar uma densidade Normal que, quando multiplicada por um polinômio em Ψ , forneça uma aproximação adequada para $\pi(\Psi|\mathbf{y})$. Uma possível escolha é a densidade Normal com média μ e variância σ^2 iguais a média e variância a posteriori de Ψ . Com isso, a equação (2.3) pode ser aplicada para obter o valor de $S(1)$ assim como para obter a média e a variância a posteriori de Ψ , já que esse procedimento só irá multiplicar $l(\mathbf{y}; \Psi)\pi(\Psi)$ por um polinômio de

no máximo ordem 2. Desse modo, um método iterativo é usado onde a média e a variância encontrada em uma iteração são usadas na iteração seguinte até que a convergência seja obtida. Ou seja, escolhe-se valores iniciais para μ e σ^2 e aplica-se (2.3) escolhendo g de forma a obter a média e a variância a posteriori de ψ . Agora, substituí-se μ e σ^2 pela média e variância a posteriori obtidas anteriormente, respectivamente e aplica-se novamente (2.3). Esse processo é repetido até que a média e a variância a posteriori de Ψ de uma iteração estejam próximas das obtidas na iteração anterior.

Para o caso multivariado foi proposto usar uma regra cartesiana multiplicativa baseada em (2.3), que pode ser escrita da forma

$$\int \dots \int g(t_1, \dots, t_k) dt_1 \dots dt_k \approx \sum_{i_k} m_{i_k}^{(k)} \dots \sum_{i_1} m_{i_1}^{(1)} g(z_{i_1}^{(1)}, \dots, z_{i_k}^{(k)}), \quad (2.5)$$

onde $m_{i_j}^{(j)}$ e $z_{i_j}^{(j)}$ são achados usando (2.4) com μ e σ^2 iguais a média e a variância marginal a posteriori de ψ_j . É muito importante ressaltar que o correto seria usar a média e variância condicional de ψ_j no lugar de μ e σ^2 em (2.4), porém o custo computacional seria muito alto uma vez que seriam necessárias iterações separadas para cada ponto. A justificativa para usar os momentos marginais envolve suposições adicionais de independência e homocedasticidade a posteriori. O problema é que em muitos casos há forte correlação a posteriori entre os elementos de Ψ . Uma proposta para contornar esse problema está em buscar uma reparametrização de Ψ (ver seção 2.2) na tentativa de achar um conjunto paramétrico tendo uma simetria esférica aproximada a posteriori.

Parte da eficiência do método está no fato de que a mesma grade e pesos (2.4) podem ser usados para o cálculo de integrais do tipo (1.3) para diversas formas de $q(\Psi)$. Isso implica que, além das distribuições a posteriori marginais, uma série de informações podem ser obtidas simultaneamente, como médias, variâncias e distribuições preditivas. No entanto, o custo computacional ainda cresce exponencialmente com a dimensão k do vetor paramétrico Ψ .

2.1.3 Aproximação de Laplace

Tierney & Kadane (1986) utilizaram o método de Laplace para obter as densidades marginais a posteriori de interesse para os casos onde a distribuição a posteriori é unimodal. Para entender seu trabalho, particione Ψ de modo que $\Psi = (\psi_1, \Psi_2)$, sendo ψ_1 o parâmetro de interesse para se obter a distribuição marginal. Defina $g(\Psi) = \log \pi(\Psi) + \log l(\Psi; \mathbf{y})$. Seja $\hat{\Psi} = (\hat{\psi}_1, \hat{\psi}_2)$ a moda a posteriori e seja Σ menos o inverso da Hessiana de $g(\Psi)$ avaliada em $\hat{\Psi}$. Para um dado ψ_1 , seja $\hat{\Psi}_2^* = \hat{\Psi}_2^*(\psi_1)$ o argumento que maximiza a função $h(\cdot) = g(\psi_1, \cdot)$, que é equivalente a $g(\Psi)$ com ψ_1 fixo e seja $\Sigma^* = \Sigma^*(\psi_1)$ menos o inverso da Hessiana de $h(\cdot)$. Aplicando o método de Laplace (ver De Bruijn, 1981) nas integrais do numerador e do denominador da expressão da densidade marginal de ψ_1 ,

$$\pi(\psi_1 | \mathbf{y}) = \frac{\int \pi(\psi_1, \Psi_2) l((\psi_1, \Psi_2); \mathbf{y}) d\Psi_2}{\int \pi(\Psi) l(\Psi; \mathbf{y}) d\Psi} \quad (2.6)$$

temos a seguinte aproximação

$$\tilde{\pi}(\psi_1 | \mathbf{y}) = \left(\frac{|\Sigma^*(\psi_1)|}{2\pi|\Sigma|} \right)^{1/2} \frac{\pi(\psi_1, \hat{\Psi}_2^*) l((\psi_1, \hat{\Psi}_2^*); \mathbf{y})}{\pi(\hat{\Psi}) l(\hat{\Psi}; \mathbf{y})} \quad (2.7)$$

A aproximação (2.7) é muito precisa, apresentando um erro para a forma funcional de ordem $O(n^{-3/2})$. O grande custo computacional do método está na obtenção de $\hat{\Psi}_2^*(\psi_1)$ para cada valor de ψ_1 , o que implica realizar uma otimização em um espaço de dimensão $(k-1)$ para cada valor de ψ_1 avaliado, o que é extremamente caro caso a dimensão k de Ψ seja alta.

2.2 Parametrização adequada de Ψ

Usualmente, quando queremos aplicar métodos numéricos para a obtenção de distribuições a posteriori marginais, como os citados na seção 2.1, é mais prático, ou em alguns casos necessário (ver seção 2.1.2), reparametrizar o vetor original Ψ de modo a obter um novo vetor, digamos \mathbf{z} , cuja distribuição $\pi(\mathbf{z} | \mathbf{y})$ tenha uma simetria esférica aproximada a posteriori, ou seja, de modo

que a nova distribuição seja o mais próximo possível de uma densidade Normal padrão multivariada, por exemplo. Tal transformação irá possibilitar a aplicação de métodos que exijam tal configuração, e além disso, irá facilitar em muito a construção de grades otimizadas (ver seção 2.3), onde as regiões de maior massa de probabilidade recebem um número de pontos superior aos das regiões com menor massa de probabilidade. Apesar de vários artigos [e.g. (Hills & Smith, 1992, 1993; Tibshirani & Wasserman, 1994)] terem sido devotados ao propósito de achar a melhor reparametrização, a simples abordagem descrita em (Smith *et al.*, 1987) oferece ótimos resultados na prática. Tal procedimento envolve duas etapas:

1. reparametrize os parâmetros individuais de forma que eles fiquem definidos na reta. Parâmetros com suporte em intervalos do tipo $(0, \infty)$ e (a, b) podem ser transformados utilizando $\log(\Psi)$ e $[\log(\Psi - a) - \log(b - \Psi)]$ respectivamente. Vamos denominar esse novo conjunto de parâmetros de Ψ^* , de modo que $\Psi^* \in \mathfrak{R}^k$, onde k é a dimensão do vetor paramétrico original, Ψ .
2. após a conclusão do item anterior, transforme os parâmetros novamente para um conjunto centrado, padronizado e 'mais' ortogonal. Um modo de obter tal conjunto é achar a moda da densidade a posteriori de Ψ^* , digamos Ψ_{moda}^* e calcular a matrix Hessiana avaliada em Ψ_{moda}^* , denominada aqui simplesmente de \mathbf{H} . O novo conjunto de parâmetros \mathbf{z} é tal que $\Psi^*(\mathbf{z}) = \Psi_{moda}^* + \mathbf{V}\Lambda^{1/2}\mathbf{z}$, onde $\Sigma = -\mathbf{H}^{-1}$ e $\Sigma = \mathbf{V}\Lambda\mathbf{V}^T$ é uma decomposição de Σ .

O primeiro passo é importante por pelo menos dois motivos, primeiro porque ele faz com que a distribuição dos parâmetros fique mais perto de ser Gaussiana, como pode ser visto na Figura 2.1 referente ao parâmetro ϕ da equação (3.17) do modelo de volatilidade estocástica apresentado no capítulo 3, que originalmente está definido no espaço $(-1, 1)$ e que, após reparametrização, $\phi^* = \log(1 + \phi) - \log(1 - \phi)$, está definido na reta Real.

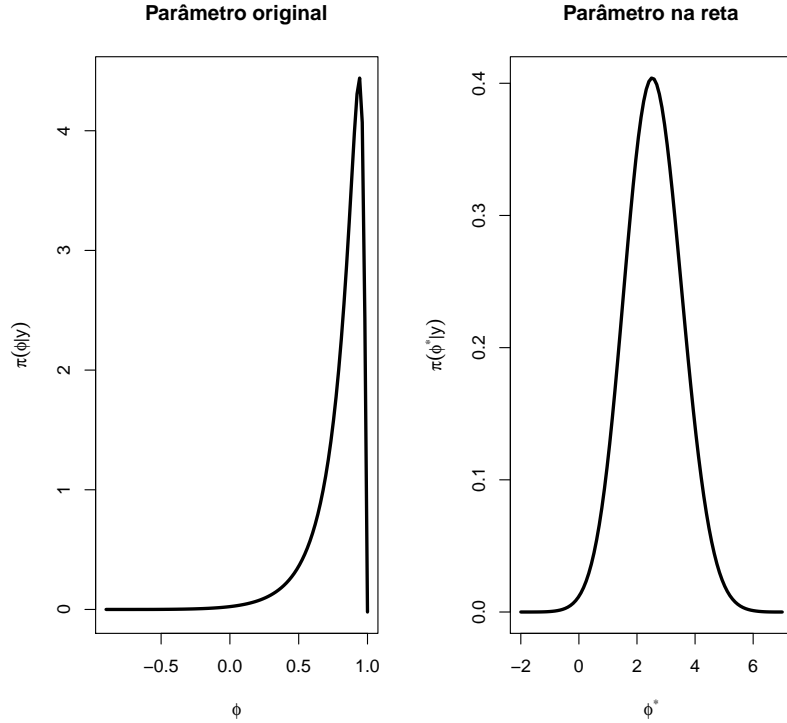


Figura 2.1: (Esquerda) Densidade do parâmetro original definido em $(-1, 1)$ (Direita) Densidade do parâmetro definido na reta Real.

Além disso, o fato que agora os parâmetros de trabalho estarão definidos em \mathcal{R}^k irá facilitar muito operações computacionais, como a otimização da densidade conjunta a posteriori por exemplo, pois algoritmos de maximização sem restrição como os definidos na seção 1.10 poderão ser aplicados.

Caso a distribuição conjunta de Ψ^* fosse uma Normal multivariada, então pela propriedade 5 da seção 1.6, \mathbf{z} teria uma distribuição Normal padrão. Apesar de Ψ^* não ter distribuição Normal multivariada, espera-se que após a execução do passo 1 acima tal distribuição seja próxima da Normal. Com isso, o passo 2 fará com que \mathbf{z} tenha uma distribuição próxima da Gaussiana padrão, como pode ser visto na figura 2.2, onde o primeiro gráfico mostra a distribuição conjunta de $\Psi = (\tau, \phi)$ do modelo de volatilidade estocástica da seção 3.3, o segundo gráfico mostra a distribuição conjunta de

$\Psi^* = (\log(\tau), \log(1 + \phi) - \log(1 - \phi))$ como descrito no passo 1, e o terceiro gráfico mostra a distribuição conjunta de \mathbf{z} , após aplicação do passo 2. Pode-se ver claramente que a distribuição conjunta obtida após a aplicação da reparametrização descrita nessa seção está perto de ter uma simetria esférica a posteriori, característica que não estava presente na densidade dos parâmetros originais.

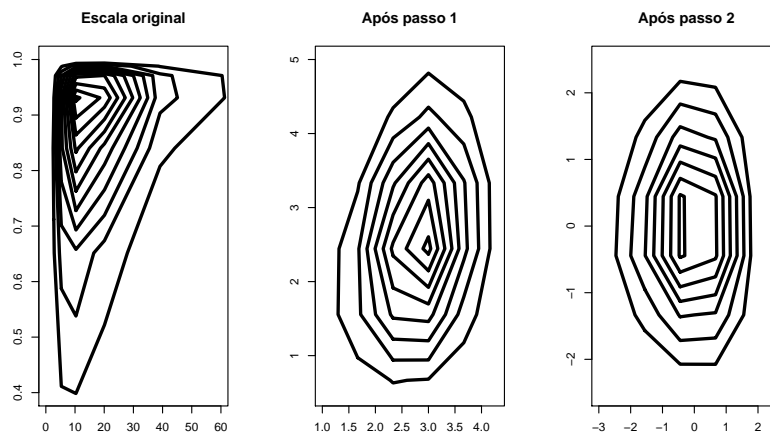


Figura 2.2: (Esquerda) Densidade dos parâmetros originais (ϕ, τ) . (Centro) Densidade após a realização do passo 1. (Direita) Densidade após a realização do passo 2.

2.3 Exploração da grade

Um ponto crucial quando se está trabalhando com aproximações determinísticas é descobrir onde se concentra a maior parte da massa de probabilidade da distribuição conjunta dos parâmetros e então usar essa informação de modo a construir uma grade otimizada, no sentido que um maior número de pontos será atribuído às regiões com maior massa de probabilidade. Possivelmente, existe mais de um modo de construir tal grade, porém a abordagem descrita aqui é a utilizada em Rue *et al.* (2009).

A idéia é explorar o log da distribuição, digamos $\log\{\pi(\Psi|\mathbf{y})\}$, utilizando a parametrização \mathbf{z} obtida como descrito na seção 2.2. Comece pela moda ($\mathbf{z} = \mathbf{0}$) e vá na direção positiva de z_1 com espaçamento de tamanho δ_z , digamos $\delta_z = 1$, até quando a condição

$$\log[\pi\{\Psi(\mathbf{0})|\mathbf{y}\}] - \log[\pi\{\Psi(\mathbf{z})|\mathbf{y}\}] < \delta_\pi \quad (2.8)$$

for válida, onde por exemplo $\delta_\pi = 7,5$. Depois, use o mesmo procedimento na direção negativa de z_1 . As outras coordenadas são tratadas do mesmo modo. O resultado até esse momento é o que denominamos de pontos pretos. Agora, complete a construção da grade ao fazer todas as combinações de pontos possíveis utilizando os pontos pretos, verificando em cada combinação se a condição (2.8) é satisfeita. Essas combinações serão denominadas de pontos cinzas, como pode ser visto na figura 2.3, a qual mostra a construção da grade utilizada no exemplo de volatilidade estocástica da seção 3.3. Comparando o gráfico da direita da figura 2.3 com o gráfico da esquerda da figura 2.2, fica claro que a grade construída utilizando a abordagem descrita acima atribui mais pontos nas regiões de maior densidade, que é exatamente o nosso objetivo.

Caso a distribuição a posteriori conjunta de Ψ fosse uma distribuição Normal padrão multivariada, os pontos da grade selecionados usando $\delta_z = 1$ e $\delta_\pi = 7,5$ seriam suficientes para cobrir a região que concentra a maior parte da massa de probabilidade da distribuição, como pode ser visto na figura 2.4 para o caso univariado. Sabemos que a distribuição a posteriori conjunta de Ψ não é Normal padrão, mas, como pode ser visto na figura 2.2, espera-se obter uma distribuição próxima da Normal padrão após a reparametrização descrita na seção 2.2.

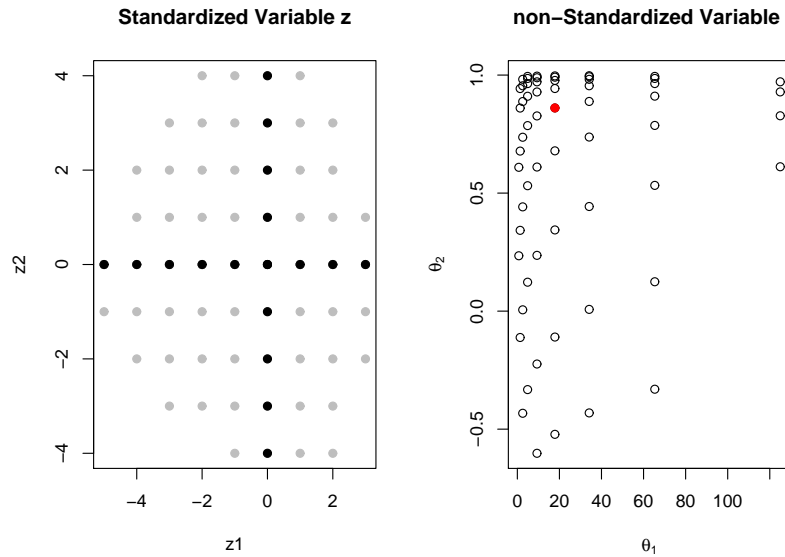


Figura 2.3: Exploração da grade do modelo de volatilidade estocástica do capítulo 3: (Esquerda) Exploração da grade na parametrização z (Direita) Grade otimizada na parametrização original.

2.4 Aplicação: Modelo de Black-Scholes Fracionário

Nesta seção, usaremos a metodologia de Tierney & Kadane (1986) apresentada na seção 2.1.3 para obter as distribuições a posteriori marginais dos parâmetros de um modelo de Black-Scholes fracionário (Cheridito, 2003) para um determinado ativo financeiro. Apesar da baixa dimensão paramétrica, esse modelo foi escolhido para demonstrar a utilidade de métodos de aproximações determinísticos porque apresenta grandes desafios para a aplicação de um esquema de MCMC eficiente e para métodos determinísticos implementados sem se preocupar com preliminares importantes, como as apresentadas nas seções 2.2 e 2.3. Tais dificuldades decorrem do fato que, como será visto adiante, o custo de avaliar a função de verossimilhança do modelo cresce drasticamente à medida que a quantidade de dados aumenta.

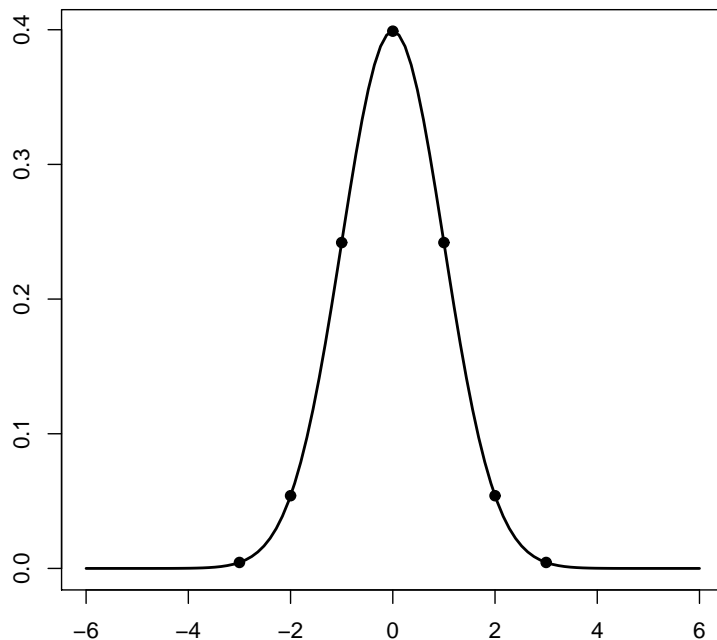


Figura 2.4: Pontos selecionados na exploração da grade ao utilizar $\delta_\pi = 7,5$ e $\delta_z = 1$ caso a distribuição seja Normal padrão.

Antes de definir o modelo, vamos definir o que é um movimento Browniano fracionário:

Definição 2.1. (Movimento Browniano fracionário) O movimento Browniano fracionário, $B^H = (B_t^H)_{t \geq 0}$, é um processo Gaussiano com função de médias igual a zero e função de covariâncias dada por

$$\Gamma_H(s, t) = \frac{1}{2}(|s|^{2H} + |t|^{2H} - |t - s|^{2H})$$

onde $H \in (0, 1)$.

Quando $H = 1/2$ o movimento Browniano fracionário reduz-se ao movimento Browniano. O modelo de Black-Scholes fracionário para um determi-

nado ativo financeiro é dado por

$$y_t = \exp\{\mu + \sigma B_t^H\}, \quad t \in [0, T], \quad H \in (0, 1) \quad (2.9)$$

onde y_t é o valor do ativo. Apesar do modelo acima ser um modelo a tempo contínuo, o que de fato se observa são $n_d + 1$ observações discretas, que no nosso caso serão os log-retornos definido como $r_t = [\log(y_{t+1}) - \log(y_t)]$. Desse modo, utilizando a equação (2.9) temos que

$$r_t = \sigma(B_{t+1}^H - B_t^H), \quad t = 1, \dots, n_d.$$

Portanto, o modelo é dado por

$$\mathbf{r} \sim N\left(\mathbf{0}, \frac{\sigma^2}{2} \boldsymbol{\Sigma}\right) \quad (2.10)$$

onde $\boldsymbol{\Sigma}(i, j) = [|i - j + 1|^{2H} - 2|i - j|^{2H} + |i - j - 1|^{2H}]$ e $\mathbf{r} = (r_1, \dots, r_{n_d})$. Os parâmetros a serem estimados são $\boldsymbol{\Psi} = (H, \sigma^2)$, sendo que usaremos uma $B(1/2, 1/2)$ como priori para H e uma $GI(1, 0.01)$ como priori para σ^2 .

Como dito anteriormente, o custo de avaliar a função de verossimilhança cresce drasticamente à medida que a quantidade de dados n_d aumenta, como pode ser observado na tabela 2.1. Isso se deve ao fato da matriz de variância-covariância $\boldsymbol{\Sigma}$ ter dimensão $n_d \times n_d$.

n_d	100	500	1000
Tempo	1, 291	70, 857	694, 438

Tabela 2.1: Tempo necessário para avaliar 100 vezes a função de verossimilhança do modelo de Black-Scholes fracionário, em segundos.

2.4.1 Simulação e Resultados

Três conjuntos de dados foram simulados ($n_d = 100, 500$ e 1000), utilizando $H = 0,6$ e $\sigma^2 = 0,4$. As marginais obtidas utilizando aproximação de Laplace (ver seção 2.1.3) estão representadas nas figuras 2.5 e 2.6. Os tempos

n_d	100	500	1000
Marginal H	3, 816	98, 850	969, 965
Marginal σ^2	4, 318	108, 817	948, 979

Tabela 2.2: Tempo necessário para a obtenção das marginais dos parâmetros do modelo de Black-Scholes fracionário, em segundos.

necessários para a obtenção de cada marginal em cada conjunto de dados podem ser vistos na tabela 2.2.

Para a aplicação de (2.7) é necessário escolher uma grade para a obtenção de cada uma das marginais, e essas foram escolhidas de acordo com as seções 2.2 e 2.3, utilizando $\delta_z = 1$ e $\delta_\pi = 7,5$. Isso significou que, nesse exemplo, foi necessário avaliar somente 8 pontos para cada uma das marginais. Tal economia de pontos é essencial em exemplos como esse, onde a função de verossimilhança se torna muito custosa de se avaliar a medida que o conjunto de dados cresce, como pode ser visto no aumento de tempo necessário para obter as marginais ocorrido a medida que n_d aumenta (ver tabela 2.2). Além disso, tal aproximação é muito precisa, apresentando erro relativo da ordem de $O(n^{-3/2})$. Obviamente, um esquema MCMC pode ser elaborado para obter as marginais de interesse desse problema, porém tal estratégia será ineficiente, devido tanto ao alto custo de se avaliar a função de verossimilhança (ver tabela 2.1) quanto ao tempo necessário para diagnosticar convergência da cadeia, tarefa longe de ser trivial. Além disso, uma vez programado adequadamente, o algoritmo utilizado nesse exemplo pode ser facilmente usado para resolver esse e outros problemas similares sem a intervenção do usuário.

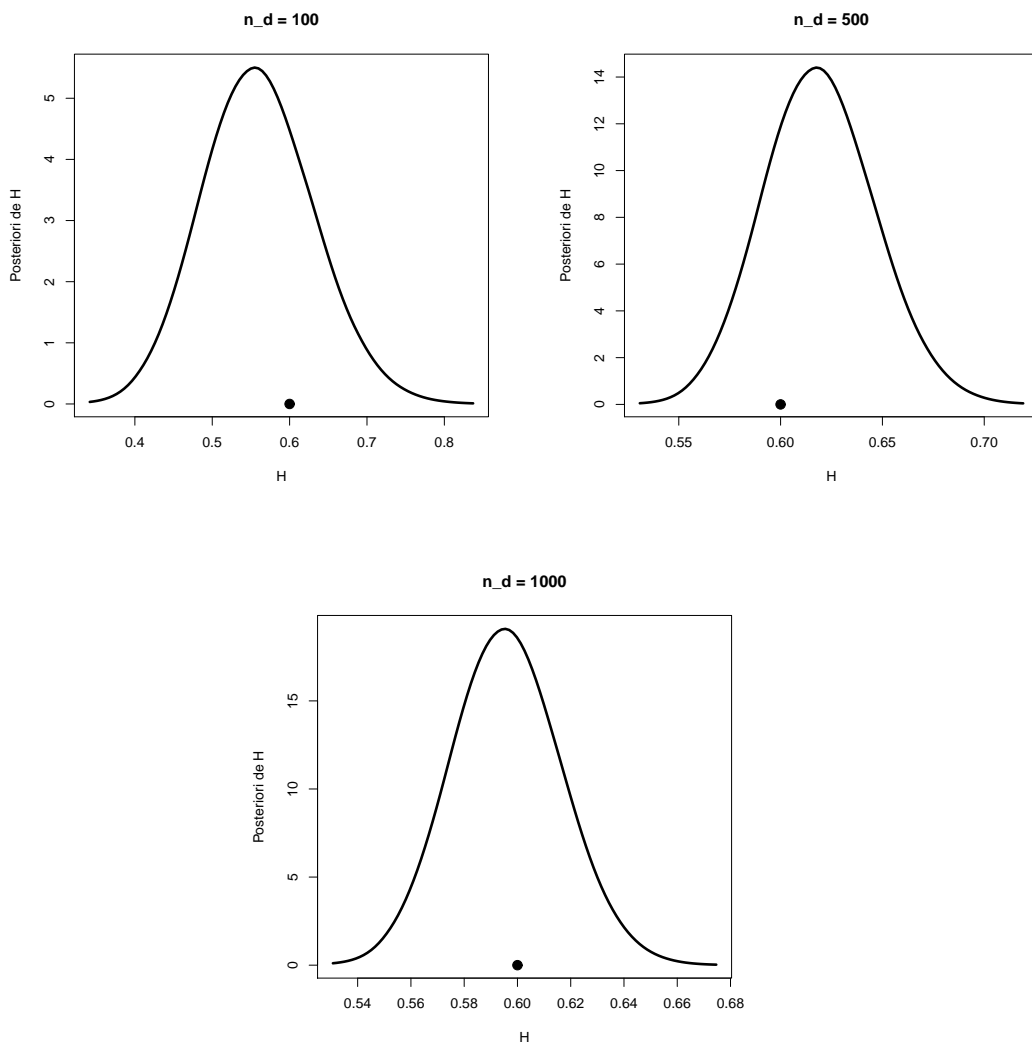


Figura 2.5: Posterioris marginais do parâmetro H para $n_d = 100, 500$ e 1000 . O ponto preto representa o valor verdadeiro do parâmetro.

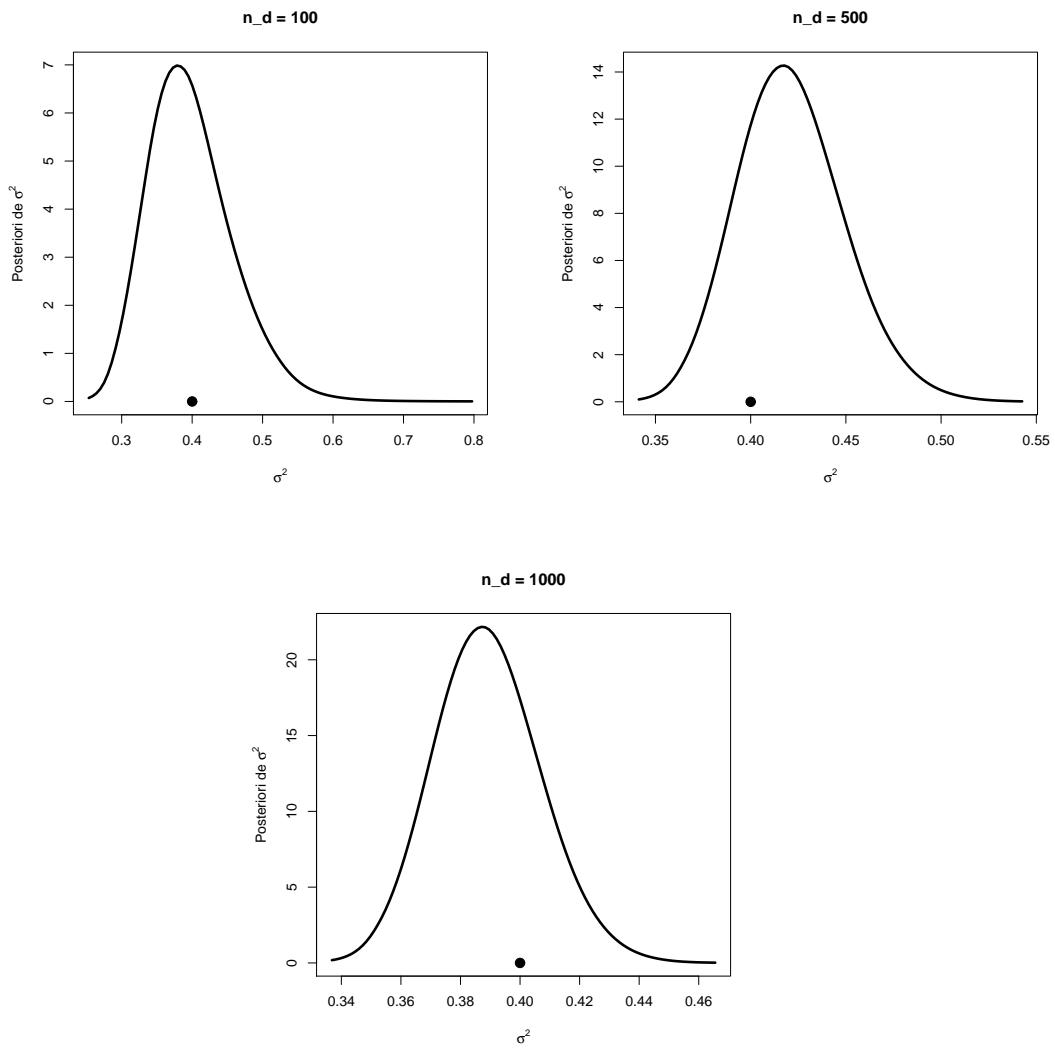


Figura 2.6: Posterioris marginais do parâmetro σ^2 para $n_d = 100, 500$ e 1000 . O ponto preto representa o valor verdadeiro do parâmetro.

Capítulo 3

INLA e aplicações

A capítulo 2 tratou da obtenção de distribuições a posteriori marginais em um contexto geral, onde dado um vetor paramétrico de interesse, $\Psi = (\psi_1, \dots, \psi_k)$, o objetivo está em obter a posteriori marginal de ψ_j , para $j = 1, \dots, k$. Este capítulo trata de um importante caso particular desse contexto geral. Suponha que se deseja obter as marginais de $\Psi = (\mathbf{x}, \boldsymbol{\theta})$, onde \mathbf{x} tem dimensão n e é denominado de campo latente Gaussiano e $\boldsymbol{\theta}$ tem dimensão m e é denominado de hiperparâmetro. Tal configuração aparece em grande parte dos modelos Bayesianos estruturados e usualmente a dimensão n é alta, de modo que os métodos apresentados no capítulo 2 não são eficientes nesse contexto, pois o custo computacional desses métodos crescem drasticamente à medida que a dimensão paramétrica aumenta. O importante artigo de Rue *et al.* (2009) propõe um método determinístico preciso e eficiente para esta classe de problemas, e o objetivo deste capítulo está em descrever e promover uma discussão qualitativa, através de exemplificação, sobre o método, de modo a esclarecer pontos importantes para seu entendimento. Além disso, a completa compreensão da metodologia apresentada ao longo deste capítulo será essencial para apreciar a extensão do método proposta no capítulo 5 da presente dissertação. A seção 3.1 descreve com mais detalhes o tipo de modelo em que o INLA é aplicável. A descrição do método será apresentada na seção 3.2 e a seção 3.3 irá discutir questões importantes sobre a metodolo-

gia do INLA, através da aplicação do método em um modelo dinâmico de primeira ordem e um modelo de volatilidade estocástica. Além disso, os resultados obtidos utilizando o INLA nesses dois exemplos serão comparados com os obtidos por métodos MCMC. O capítulo é finalizado com a seção 3.4, que oferece uma discussão sobre os motivos da eficiência e precisão do INLA, que permite a sua aplicação mesmo em situações com um grande número de parâmetros (n entre 10.000 e 100.000).

3.1 Introdução

O método *Integrated Nested Laplace Approximation* (Rue *et al.*, 2009), doravante referido como INLA, tem aplicação em modelos hierárquicos Bayesianos cuja base é um vetor aleatório não-observável \mathbf{x} de dimensão n , denominado campo latente Gaussiano, cuja densidade Normal multivariada $\pi(\mathbf{x}|\boldsymbol{\theta})$ é controlada por um vetor paramétrico $\boldsymbol{\theta}$, de dimensão m . Alguns elementos do vetor aleatório \mathbf{x} são indiretamente observados através dos dados \mathbf{y} , de dimensão n_d . Esses são assumidos condicionalmente independentes dado o campo latente \mathbf{x} , o que para este capítulo implica que $\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n_d} \pi(y_i|x_i, \boldsymbol{\theta})$, onde cada observação está conectada a somente um elemento do campo latente. No capítulo 5 será apresentada uma extensão do método, onde cada observação poderá estar relacionada com mais de um elemento do campo latente. Após atribuir uma priori para o vetor paramétrico $\boldsymbol{\theta}$, temos a seguinte distribuição a posteriori:

$$\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta}) \prod_i \pi(y_i|x_i, \boldsymbol{\theta})$$

Uma grande gama de modelos (e.g. modelos de regressão, modelos dinâmicos, modelos espaciais e espaço-temporais) amplamente utilizados na literatura se enquadra na descrição hierárquica definida acima. Refere-se ao primeiro capítulo de Rue & Held (2005) e à seção 1.2 de Rue *et al.* (2009) para uma listagem mais detalhada sobre a quantidade de aplicações que se enquadram na estrutura descrita acima.

O campo latente Gaussiano \mathbf{x} fornece uma ferramenta flexível para modelar dependência temporal e espacial entre os dados e entre os dados e as covariáveis em potencial. Na maioria dos exemplos de interesse, \mathbf{x} é um CAMG de grande dimensão.

Definição 3.1. O vetor aleatório $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathfrak{R}^n$ é um CAMG com relação ao grafo $G = (V, E)$ com média $\boldsymbol{\mu}$ e matrix de precisão $\mathbf{Q} > 0$, se e somente se sua densidade tem a forma

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

e

$$Q_{ij} \neq 0 \iff \{i, j\} \in E \quad \forall i \neq j.$$

Olhando a definição 1.1 e a propriedade 6 da seção 1.6 e comparando-as com a definição de CAMG dada acima, chegamos à conclusão que qualquer distribuição Normal é um CAMG e vice versa. Porém nosso maior interesse se encontra nos casos onde a matriz de precisão \mathbf{Q} de \mathbf{x} é esparsa, como definido na seção 1.4, de modo que algoritmos usados para aproximações de CAMG são beneficiados por essa propriedade, o que aumenta em muito sua velocidade. Foi escolhido não perseguir na presente dissertação a descrição de algoritmos específicos para matrizes esparsas utilizados no contexto de CAMG, dado que esse tópico está detalhadamente discutido em Rue (2001), Rue & Follestad (2002), Rue & Held (2005) e Rue *et al.* (2009). No entanto, ao longo desta dissertação, o pacote descrito em Bates & Maechler (2007) foi utilizado para os algoritmos envolvendo matrizes esparsas.

3.2 Método INLA

3.2.1 Objetivos

O objetivo neste capítulo é a obtenção das marginais $\pi(x_i | \mathbf{y})$, $i = 1, \dots, n$ e $\pi(\theta_j | \mathbf{y})$, $j = 1, \dots, m$, que podem ser obtidas através das duas equações abaixo:

$$\pi(x_i|\mathbf{y}) = \int \pi(x_i|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}, \quad (3.1)$$

$$\pi(\theta_j|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}_{-j}. \quad (3.2)$$

Como as quantidades e as integrais contidas em (3.1) e (3.2) não estão disponíveis analiticamente, aproximações $\tilde{\pi}(x_i|\mathbf{y})$ e $\tilde{\pi}(\theta_j|\mathbf{y})$ são necessárias, e obtidas através de

$$\tilde{\pi}(x_i|\mathbf{y}) = \sum_k \tilde{\pi}(x_i|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y})\Delta_k, \quad (3.3)$$

$$\tilde{\pi}(\theta_j|\mathbf{y}) = \sum_k \tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y})\Delta_{jk}, \quad (3.4)$$

de modo que para aplicar (3.3) e (3.4) é necessário obter as aproximações $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$, $\tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})$ e avaliar $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ em uma grade obtida conforme descrito na seção 2.3.

3.2.2 Aproximação para $\pi(\boldsymbol{\theta}|\mathbf{y})$

Temos que a identidade

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})}{\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} \quad (3.5)$$

é válida para todo \mathbf{x} . Rue & Martino (2007) propõem aproximar $\pi(\boldsymbol{\theta}|\mathbf{y})$ por

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\pi_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})} \quad (3.6)$$

onde o denominador de (3.6) é uma aproximação Gaussiana para a condicional completa de \mathbf{x} ,

$$\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2}\mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta})\mathbf{x} + \sum_{i \in \mathbf{I}} g_i(x_i, \boldsymbol{\theta}, y_i) \right\}, \quad (3.7)$$

efetuada conforme descrito na seção 1.11.3. Nesse caso $g_i(x_i, \boldsymbol{\theta}, y_i) = \log\{\pi(y_i|x_i, \boldsymbol{\theta})\}$ e $\mathbf{x}^*(\boldsymbol{\theta})$ é a moda da distribuição condicional completa $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ obtida através de algum método de otimização, como por exemplo o método de

Newton-Raphson dado pelo algoritmo 1 da seção 1.10. Nesse caso, como $g_i(x_i, \boldsymbol{\theta}, y_i)$ só depende de um elemento de \boldsymbol{x} a matriz $\mathbf{C}(\boldsymbol{x}_0)$ da equação (1.22) será diagonal, o que não acontece em um caso geral onde a verossimilhança depende de mais de um elemento de \boldsymbol{x} , assunto que será abordado no capítulo 5.

O sinal de proporcional de (3.6) vem do fato de não conhecermos a constante normalizadora de $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$. É interessante notar que, para o contexto deste capítulo, a aproximação (3.6) é equivalente a aproximação de Laplace para marginais descrita na seção 2.1.3, o que sugere que o erro de aproximação é relativo e de ordem $O(n_d^{-3/2})$. Porém, como a dimensão paramétrica não é fixa, pois depende de n_d , as suposições assintóticas usualmente usadas nas expansões de Laplace não são válidas aqui. Rue & Martino (2007) aplicaram (3.6) em diversos modelos latentes Gaussianos e verificaram que tal aproximação é muito precisa, de modo que nem longas cadeias de MCMC conseguiram detectar algum erro na aproximação.

3.2.3 Aproximação para $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$

Rue *et al.* (2009) propõem três tipos de aproximação para $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$, sendo que tais opções diferem em custo computacional e precisão conforme descrito a seguir.

Aproximação Gaussiana

A aproximação Gaussiana $\pi_G(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ é a mais simples e rápida de ser obtida pois, através da propriedade 1 da seção 1.6, a média $\mu_i(\boldsymbol{\theta})$ e a variância $\sigma_i^2(\boldsymbol{\theta})$ são obtidas de $\pi_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$, que já terá sido computada para obter $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ de acordo com a equação (3.6) para diversos valores de $\boldsymbol{\theta}$ escolhidos durante a exploração da grade, de acordo com a seção 2.3. Desse modo, o único esforço adicional necessário para obter $\pi_G(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ seria o cálculo das variâncias marginais $\sigma_i^2(\boldsymbol{\theta})$ a partir da matriz de precisão $Q^*(\boldsymbol{\theta})$ da aproximação Gaussiana para a condicional completa. Rue & Martino (2007) mostraram

que a aproximação Gaussiana oferece resultados satisfatórios em muitos casos, porém pode apresentar erros de locação e/ou erros devido à falta de assimetria da aproximação.

Aproximação de Laplace

A aproximação de Laplace é dada por

$$\tilde{\pi}_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}_{-i}=\mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta})} \quad (3.8)$$

onde $\mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta})$ é a moda de $\pi(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})$ e $\tilde{\pi}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})$ é a aproximação Gaussiana de $\pi(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})$, que difere da distribuição condicional obtida a partir de $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$. Como (3.8) é uma aproximação não paramétrica, é necessária avaliá-la em diversos pontos, digamos $\{x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(k)}\}$, para obter a densidade. A escolha desses pontos é tal que $x_i^{(j)} = \mu_i(\boldsymbol{\theta}) + \sigma_i(\boldsymbol{\theta})x_{ab}^{(j)}$, onde $x_{ab}^{(j)}$ é um ponto da abscissa dado pela quadratura de Gauss-Hermite, $\mu_i(\boldsymbol{\theta})$ e $\sigma_i(\boldsymbol{\theta})$ são a média e desvio-padrão da aproximação Gaussiana $\pi_G(x_i|\boldsymbol{\theta}, \mathbf{y})$ respectivamente. Para representar $\tilde{\pi}_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y})$ usa-se

$$\tilde{\pi}_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y}) \propto N\{x_i; \mu_i(\boldsymbol{\theta}), \sigma_i^2(\boldsymbol{\theta})\} \times \exp\{h(x_i)\} \quad (3.9)$$

A função $h(x)$ é uma função spline cúbica (Ahlberg *et al.*, 1967) ajustada à diferença $\log\{\tilde{\pi}_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y})\} - \log\{\tilde{\pi}_G(x_i|\boldsymbol{\theta}, \mathbf{y})\}$ avaliada nos pontos selecionados, e então a densidade (3.9) é normalizada utilizando integração por quadratura.

A aproximação (3.8), assim como a (3.6), oferece resultados extremamente precisos. O grande problema é que tanto a moda $\mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta})$ quanto a matrix de precisão $Q^*(x_i, \boldsymbol{\theta})$ de $\tilde{\pi}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})$ dependem do valor de x_i , o que significa que para cada ponto $x_i^{(j)}$, $i = 1, \dots, n$ e $j = 1, \dots, k$ será necessário inverter matrizes de dimensão $(n-1) \times (n-1)$ mais de uma vez.

Para amenizar esse problema, Rue *et al.* (2009) propuseram duas modificações em (3.8). A primeira seria aproximar a moda $\mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta})$ por $E_{\tilde{\pi}_G}(\mathbf{x}_{-i}|x_i)$, que é a média da distribuição condicional obtida a partir da aproximação

Gaussiana $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$. Tal aproximação seria exata caso a condicional completa (3.7) fosse Normal, o que não é o caso dado que o termo da verossimilhança $g_i(x_i, \boldsymbol{\theta}, \mathbf{y})$ é não-Gaussiano. Porém (3.7) não difere muito de uma Normal, em grande parte pelo fato de atribuirmos uma priori Gaussiana para \mathbf{x} . Além disso, a aproximação Gaussiana $\tilde{\pi}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})$ não necessita ser computada exatamente na moda da distribuição e sim em um ponto de alta densidade, sendo essa a explicação do porque a aproximação (3.8) não é degradada ao utilizar $E_{\tilde{\pi}_G}(\mathbf{x}_{-i}|x_i)$ em vez de $\mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta})$. Essa pequena modificação significa que não será necessário realizar a otimização de $\pi(\mathbf{x}_{-i}|x_i^{(j)}, \boldsymbol{\theta}, \mathbf{y})$ para $i = 1, \dots, n$ e $j = 1, \dots, k$, que é uma tarefa muito cara computacionalmente.

A segunda modificação é inspirada no fato que, devido às propriedades Markovianas do campo latente \mathbf{x} , somente alguns elementos de \mathbf{x} são afetados pelo comportamento de x_i na condicional $\pi(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})$, de modo que nem todo o vetor \mathbf{x}_{-i} precisa ser considerado. Isso significa que será necessário fatorizar matrizes de dimensão significativamente menores do que $(n-1) \times (n-1)$. Denote por $R_i(\boldsymbol{\theta})$ a "região de interesse" com relação a marginal de x_i , temos que

$$\frac{E_{\tilde{\pi}_G}(x_j|x_i) - \mu_j(\boldsymbol{\theta})}{\sigma_j(\boldsymbol{\theta})} = a_{ij}(\boldsymbol{\theta}) \frac{x_i - \mu_i(\boldsymbol{\theta})}{\sigma_i(\boldsymbol{\theta})} \quad (3.10)$$

para algum $a_{ij}(\boldsymbol{\theta})$ quando $j \neq i$. Com isso, uma simples regra para a construção de $R_i(\boldsymbol{\theta})$ é

$$R_i(\boldsymbol{\theta}) = \{j : |a_{ij}(\boldsymbol{\theta})| > 0,001\}. \quad (3.11)$$

A grande vantagem é que agora só será necessário fatorizar matrizes esparsas de dimensão $|R_i(\boldsymbol{\theta})| \times |R_i(\boldsymbol{\theta})|$.

Aproximação de Laplace simplificada

As duas modificações sugeridas acima tornaram a aproximação de Laplace (3.8) viável de ser calculada, porém ainda custosa computacionalmente, especialmente se a dimensão de \mathbf{x} for muito grande. A proposta dessa seção é

efetuar uma expansão de Taylor ao redor de $x_i = \mu_i(\boldsymbol{\theta})$ até os termos de terceira ordem no numerador e até o termo de primeira ordem no denominador de (3.8). Seja $x_i^s = (x_i - \mu_i(\boldsymbol{\theta}))/\sigma_i(\boldsymbol{\theta})$ e

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

temos que (ver apêndice A.1)

$$\log \tilde{\pi}_{SLA}(x_i^s|\boldsymbol{\theta}, \mathbf{y}) \stackrel{c}{\propto} \gamma_i^{(1)}(\boldsymbol{\theta})x_i^s - \frac{1}{2}(x_i^s)^2 + \frac{1}{6}(x_i^s)^3\gamma_i^{(3)}(\boldsymbol{\theta}) + \dots \quad (3.12)$$

onde

$$\begin{aligned} \gamma_i^{(1)}(\boldsymbol{\theta}) &= \frac{1}{2} \sum_{j \in \mathbf{I}/i} \sigma_j^2(\boldsymbol{\theta}) \{1 - \text{corr}_{\tilde{\pi}_G}(x_i, x_j)^2\} d_j^{(3)}\{\mu_i(\boldsymbol{\theta}), \boldsymbol{\theta}\} \sigma_j(\boldsymbol{\theta}) a_{ij}(\boldsymbol{\theta}) \\ \gamma_i^{(3)}(\boldsymbol{\theta}) &= \sum_{j \in \mathbf{I}/i} d_j^{(3)}\{\mu_i(\boldsymbol{\theta}), \boldsymbol{\theta}\} \{\sigma_j(\boldsymbol{\theta}) a_{ij}(\boldsymbol{\theta})\}^3. \end{aligned}$$

A equação (3.12) não define uma densidade, pois o termo de terceira ordem é ilimitado. Uma solução é ajustar uma densidade Normal assimétrica (Azalini & Capitanio, 1999) de modo que a terceira derivada avaliada na moda seja $\gamma_i^{(3)}(\boldsymbol{\theta})$, a média seja $\gamma_i^{(1)}(\boldsymbol{\theta})$ e a variância igual a 1 (ver apêndice A.2 para detalhes). Com isso, $\gamma_i^{(3)}(\boldsymbol{\theta})$ irá corrigir a aproximação Gaussiana na assimetria e $\gamma_i^{(1)}(\boldsymbol{\theta})$ irá ajustar a locação.

A forma de (3.12) só foi possível de se obter analiticamente porque a matriz $\mathbf{C}(\mathbf{x}_0)$ da aproximação Gaussiana $\pi_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ dada pela equação (1.22) é uma matrix diagonal (ver apêndice A.1). No capítulo 5 será mostrado como obter as aproximações descritas nessa seção quando a matriz $\mathbf{C}(\mathbf{x}_0)$ não for diagonal.

3.3 Aplicações

Exemplo 1: Modelo Dinâmico de Primeira Ordem

Definição do modelo

$$y_t = x_t + \epsilon_t, \quad \epsilon_t \sim N(0, V) \quad (3.13)$$

$$x_t = x_{t-1} + w_t, \quad w_t \sim N(0, W), \quad (3.14)$$

para $t = 1, \dots, n_d$. Para todo t e todo s com $t \neq s$ temos que $\epsilon_t \perp \epsilon_s$, $w_t \perp w_s$, e $\epsilon_t \perp w_s$. O modelo é completado com a distribuição a priori:

$$x_1 \sim N(0, R), \quad V \sim GI\left(\frac{n_V}{2}, \frac{n_V S_V}{2}\right) \text{ e } W \sim GI\left(\frac{n_W}{2}, \frac{n_W S_W}{2}\right)$$

Nesse caso $\mathbf{x} = (x_1, \dots, x_{n_d})$, $\boldsymbol{\theta} = (V, W)$.

Esse modelo é utilizado em diversas aplicações, principalmente em previsão de curto-prazo para planejamento de produção e controle de estoque. Por exemplo, ao modelar a demanda de mercado para um produto, x_t representa a *verdadeira* demanda intrínseca do mercado no tempo t , com ϵ_t descrevendo as flutuações aleatórias. Localmente no tempo, a demanda intrínseca x_t é caracterizada como aproximadamente constante. Mudanças significativas através de longos períodos de tempo são esperadas, mas a natureza de w_t indica que não se deseja antecipar tais mudanças, descrevendo-as como um processo puramente estocástico. Sugere-se West & Harrison (1997) para mais detalhes.

Simulação e Resultados

Um conjunto de dados com $n_d = 100$ observações, mostrado na figura 3.1, foi simulado do modelo (3.13) e (3.14) com $V = 1$, $W = 0,1$ e $x_0 = 3$. Para as distribuições a priori foi usado $R = 100$ e $n_V = n_W = n_V S_V = n_W S_W = 0,02$.

Esse não é um modelo desafiador para demonstrar o poder de um método de estimação, porém alguns pontos importantes podem ser esclarecidos analisando modelos simples como esse. Para começar, nesse caso temos que o tamanho do campo latente é igual ao tamanho dos dados, de modo que $n = n_d$, o que nem sempre é verdade, como será visto no próximo exemplo. Podemos ver a seguir a estrutura de independência condicional do campo latente \mathbf{x} codificada na matriz de precisão $Q(\boldsymbol{\theta})$ através dos elementos nulos e não-nulos como descrito na definição 3.1.

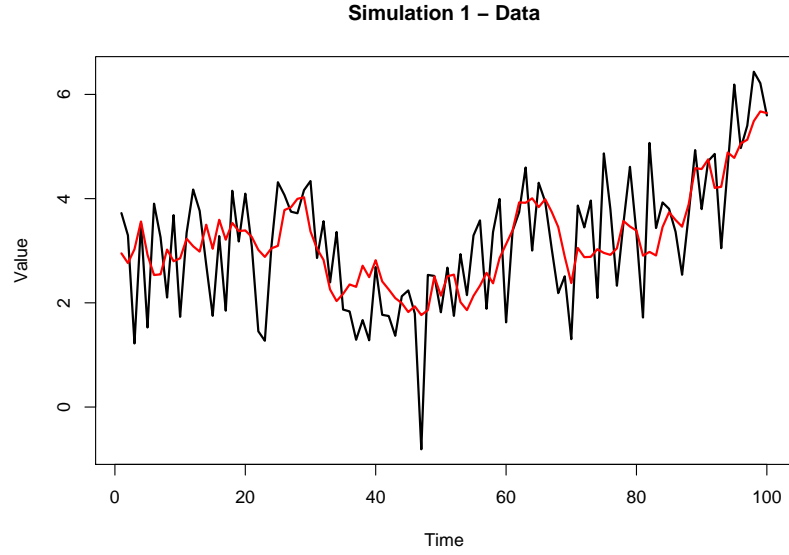


Figura 3.1: Dados simulados para o modelo dinâmico de 1a ordem: - Dados
- Campo Latente

$$\begin{aligned}
 p(\mathbf{x}|\boldsymbol{\theta}) &= p(x_1) \prod_{t=2}^n p(x_t|x_1, \dots, x_{t-1}) = p(x_1) \prod_{t=2}^n p(x_t|x_{t-1}) \\
 &\propto \exp\left\{-\frac{1}{2}R^{-1}x_1^2\right\} \prod_{t=2}^n |W|^{-1/2} \exp\left\{-\frac{W^{-1}}{2}(x_t - x_{t-1})^2\right\} \\
 &\propto \exp\left\{-\frac{1}{2}\left[W^{-1}\left(x_1^2 + x_n^2 + 2\sum_{t=2}^n x_t^2 - 2\sum_{t=2}^n x_t x_{t-1}\right) + R^{-1}x_1^2\right]\right\} \\
 &\propto \exp\left\{-\frac{1}{2}\mathbf{x}^T Q(\boldsymbol{\theta})\mathbf{x}\right\}, \text{ onde}
 \end{aligned}$$

$$Q(\boldsymbol{\theta}) = \begin{pmatrix} W^{-1} + R^{-1} & -W^{-1} & 0 & \dots & 0 & 0 \\ -W^{-1} & 2W^{-1} & -W^{-1} & \dots & 0 & 0 \\ 0 & -W^{-1} & 2W^{-1} & \dots & 0 & 0 \\ 0 & 0 & -W^{-1} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -W^{-1} & 0 \\ 0 & 0 & 0 & \dots & 2W^{-1} & -W^{-1} \\ 0 & 0 & 0 & \dots & -W^{-1} & W^{-1} \end{pmatrix}$$

Outro fato importante a ser notado é que a função de verossimilhança como função do campo latente \mathbf{x} é Normal, já que $g_i(x_i) = \log\{\pi(y_i|x_i, \boldsymbol{\theta})\}$ tem distribuição Normal, o que implica que a aproximação Gaussiana $\pi_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ obtida de acordo com a seção 1.11.3 na verdade não é uma aproximação e sim o cálculo exato de $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$. A consequência disso é que a aproximação da distribuição a posteriori de $\boldsymbol{\theta}$

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\pi_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} \quad (3.15)$$

é exata a menos de uma constante de proporcionalidade. Desse modo, a única tarefa restante a ser executada é explorar suficientemente bem a grade a ser utilizada na integração numérica das equações (3.3) e (3.4). Foi feita uma reparametrização de acordo com a seção 2.2 e a grade foi construída como na seção 2.3 utilizando-se $\delta_\pi = 7.5$ e $\delta_z = 1$ e pode ser vista na figura 3.2.

Como temos $\pi_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ calculado para cada ponto $\boldsymbol{\theta}_k$ da grade, podemos obter nossas marginais de interesse através das equações (3.3) e (3.4). A figura 3.3 mostra as distribuições a posteriori marginais de V e de W obtidas utilizando o INLA e utilizando o método Amostrador de Gibbs e, como pode ser observado, ambos os métodos fornecem o mesmo resultado. A figura 3.4 mostra as distribuições a posteriori marginais do campo latente obtidas utilizando o INLA e utilizando o Amostrador de Gibbs e mais um vez fica muito difícil detectar qualquer diferença entre os dois métodos.

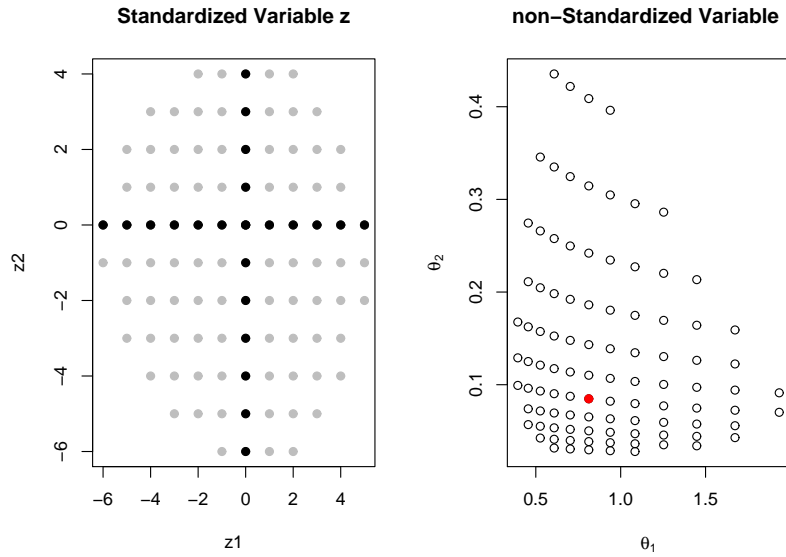


Figura 3.2: ● moda a posteriori; ● black dots; ● grey dots

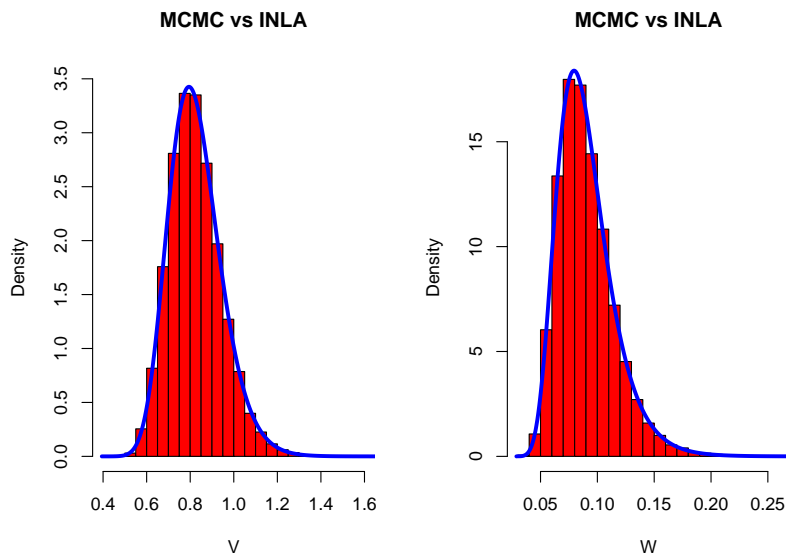


Figura 3.3: (Esquerda) posteriori marginal de V e (Direita) posteriori marginal de W : ■ (Histograma) MCMC — INLA

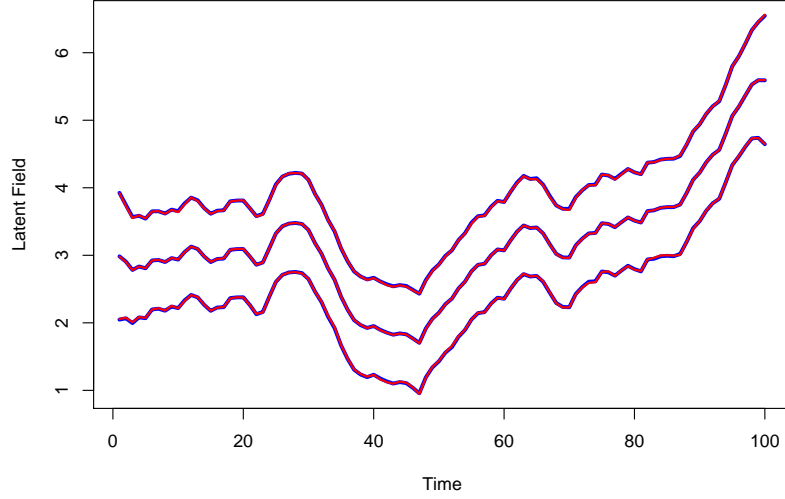


Figura 3.4: Média a posteriori e IC 95% do campo latente: - MCMC - INLA

Como visto, os resultados obtidos com o INLA e com o MCMC foram idênticos, a diferença sendo o tempo computacional. Enquanto que o INLA efetua os cálculos acima em poucos segundos, o MCMC leva minutos para chegar à precisão exibida nas figuras desse exemplo.

Exemplo 2: Modelo de Volatilidade Estocástica

O modelo de volatilidade estocástica (Taylor, 2006) definido a seguir é frequentemente usado para analisar séries temporais financeiras, onde acredita-se que a variância do processo varia ao longo do tempo.

Definição do modelo

$$y_t = \sqrt{\exp(\eta_t)}\epsilon_t, \quad \epsilon_t \sim N(0, 1) \quad (3.16)$$

$$\eta_t = \mu + \phi(\eta_t - \mu) + v_t, \quad v_t \sim N(0, \sigma_\eta^2), \quad (3.17)$$

para $t = 1, \dots, n_d$. Para todo t e todo s com $t \neq s$ temos que $\epsilon_t \perp \epsilon_s$, $v_t \perp v_s$, e $\epsilon_t \perp v_t$. O modelo é completado com as distribuições a priori:

$$\eta_1 \sim N\left(\mu, \frac{\sigma_\eta^2}{1-\phi^2}\right), \tau \sim G(\alpha_\tau, \beta_\tau), \phi^* \sim N(a_1, R_1) \text{ e } \mu \sim N(0, R_2)$$

onde $\phi^* = \log\left\{\frac{1+\phi}{1-\phi}\right\}$ e $\tau = \frac{1}{\sigma_\eta^2}$.

Nesse caso, diferente do exemplo anterior, temos duas configurações possíveis para o campo latente \mathbf{x} e hiperparâmetros $\boldsymbol{\theta}$. Podemos definir $\mathbf{x} = (\eta_1, \dots, \eta_m)$ e $\boldsymbol{\theta} = (\tau, \phi, \mu)$, o que denotamos de **Configuração 1**. Outra possibilidade é considerar $\mathbf{x} = (\eta_1, \dots, \eta_m, \mu)$ e $\boldsymbol{\theta} = (\tau, \phi)$, denotada de **Configuração 2**. A única diferença entre as duas configurações é que μ , que antes estava sendo considerado como hiperparâmetro, passou a fazer parte do campo latente. Essa abordagem será possível sempre que o parâmetro em questão tiver distribuição a priori Normal e for linear em relação aos outros elementos do campo latente, exatamente como acontece na equação (3.17). A vantagem de se utilizar a **Configuração 2** ao invés da **Configuração 1** é puramente computacional, já que o maior custo computacional do método INLA está concentrado na exploração da grade dos hiperparâmetros, sendo que, quanto menor for a dimensão de $\boldsymbol{\theta}$, mais eficiente será o método.

Além disso, nesse exemplo não é possível integrar fora o campo latente \mathbf{x} analiticamente, como feito no exemplo anterior, pois agora a verossimilhança, como função do campo latente \mathbf{x} , não é Gaussiana, sendo nesse caso aplicável os três tipos de aproximação para $\pi(x_i | \boldsymbol{\theta}, \mathbf{y})$ apresentadas na seção 3.2.3.

Resultados

O modelo de volatilidade estocástica [(3.16)-(3.17)] será ajustado nas primeiras 50 observações do conjunto de dados originalmente analisado em Durbin & Koopman (2000), mostrados na figura 3.5. Os parâmetros das prioris usados nos resultados mostrados a seguir são:

$$\alpha_\tau = 1, \beta_\tau = 0.1, a_1 = 3, R_1 = 1 \text{ e } R_2 = 1$$

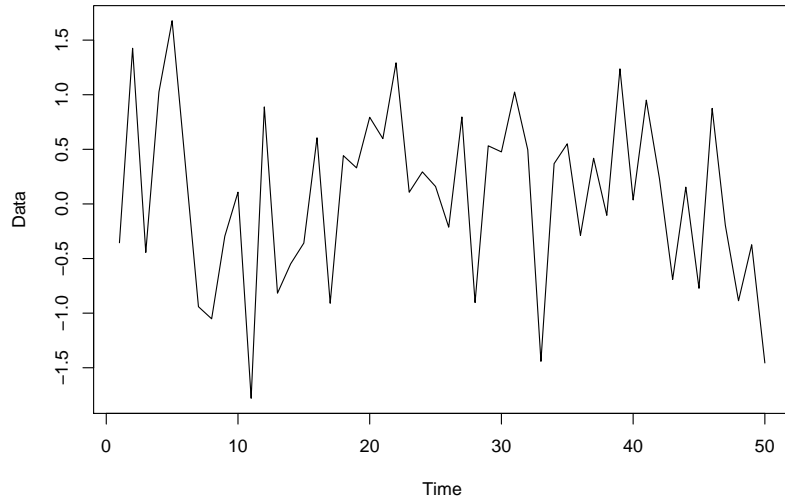


Figura 3.5

A exploração da grade de θ segundo a **Configuração 1** pode ser vista na figura 3.6, enquanto que a exploração segundo a **Configuração 2** pode ser visualizada na figura 3.7.

As marginais de ϕ e τ vão ser idênticas em ambas as configurações já que nos dois casos eles estão sendo considerados como elementos do hiperparâmetro θ . A figura 3.8 contém as marginais obtidas pelo INLA e as obtidas pelo MCMC para ϕ e τ . Para o esquema MCMC foi necessário uma cadeia realmente longa, obtida utilizando WinBugs (Lunn *et al.*, 2000), para conseguir atingir a convergência e precisão demonstradas nas figuras dessa seção, sendo que tal processo demorou horas de esforço computacional e análise. Na verdade, o motivo de só estar sendo usada as 50 primeiras observações do conjunto de dados de Durbin & Koopman (2000) é porque quanto mais dados disponíveis, maior é a correlação a posteriori entre \mathbf{x} e θ , sendo que quanto maior essa correlação, mais difícil é obter convergência no WinBUGS, enquanto que isso em nada afeta a eficácia do INLA.

Diferentemente do WinBUGS, todos os resultados dessa seção obtidos com o INLA levaram segundos para serem computados utilizando a **Con-**

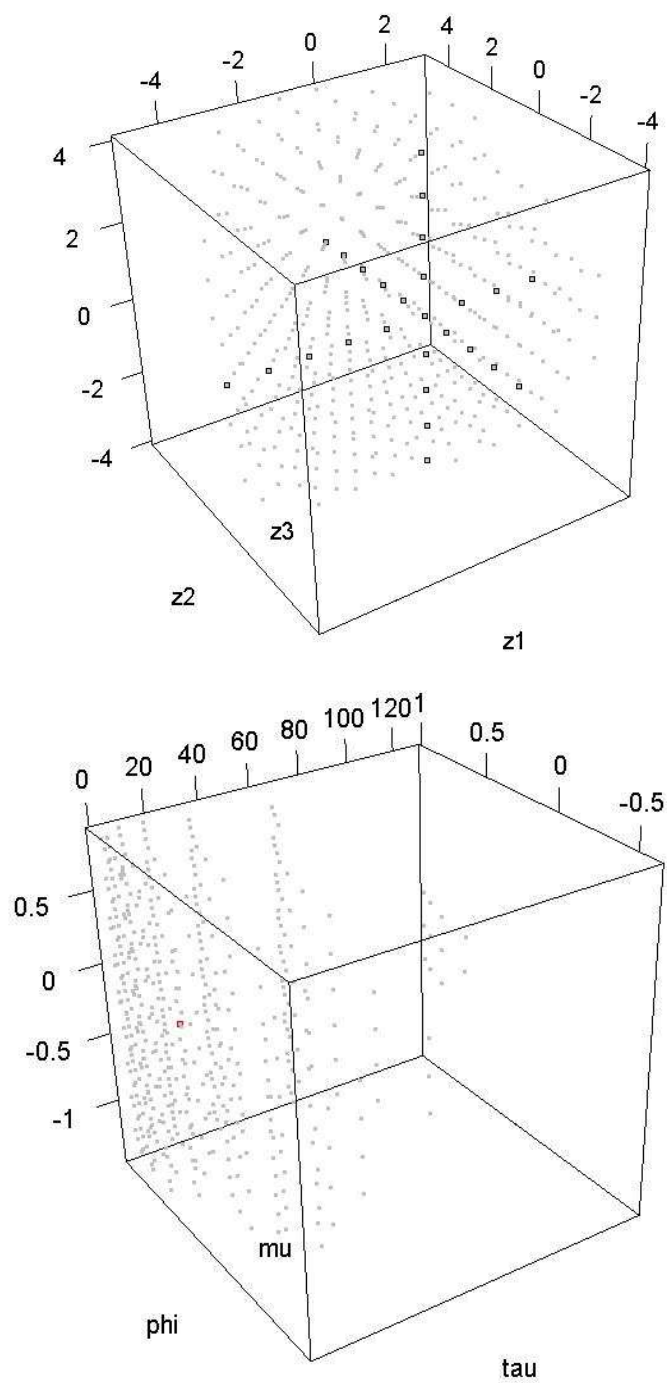


Figura 3.6: Exploração da grade com a **Configuração 1**: (Acima) θ padronizado. (Abaixo) θ na escala original

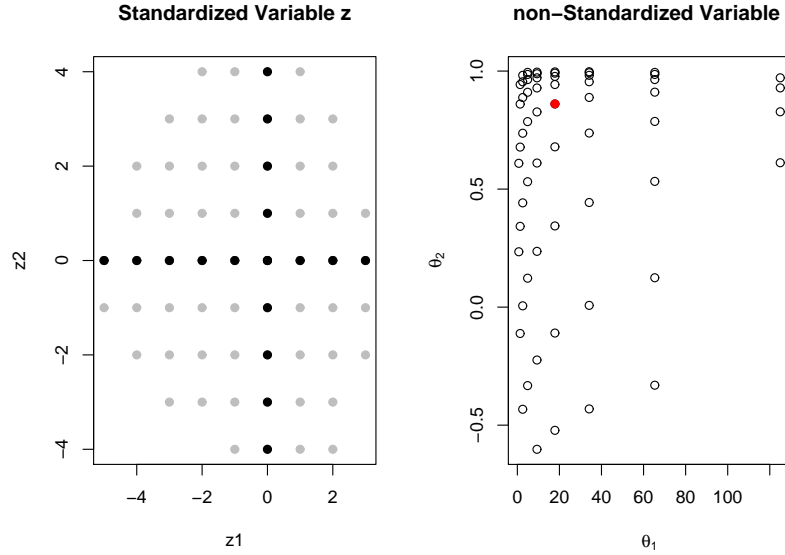


Figura 3.7: Exploração da grade com a **Configuração 2**: (Esquerda) θ padronizado. (Direita) θ na escala original

figuração 2 e minutos (um pouco mais de 2) utilizando a **Configuração 1**, sendo que esse tempo é drasticamente reduzido quando se utiliza a biblioteca de Martino & Rue (2008), onde toda a programação do INLA foi eficientemente implementada utilizando a linguagem C de programação. O motivo da diferença do tempo computacional entre as duas configurações é que, como a maior parte do custo computacional do INLA está na exploração da grade de θ , isso implica que, quanto maior a dimensão de θ , maior será o tempo necessário para a obtenção dos resultados.

A figura 3.9 contém a marginal de μ obtida pelo MCMC, a aproximação obtida pelo INLA utilizando a aproximação Gaussiana $\pi_G(x_i|\theta, \mathbf{y})$, a aproximação de Laplace Simplificada $\pi_{SLA}(x_i|\theta, \mathbf{y})$ e a aproximação de Laplace $\pi_{LA}(x_i|\theta, \mathbf{y})$, que pode ser obtida tanto com o μ incluído no θ (Configuração 1), quanto no \mathbf{x} (Configuração 2). Pela figura, nota-se que o resultado obtido com o MCMC é bem próximo do obtido pela aproximação de Laplace simplificada, que por sua vez coincidiu com a aproximação de Laplace, sendo que ambas corrigem os erros presentes ao se utilizar a aproximação Gaussiana.

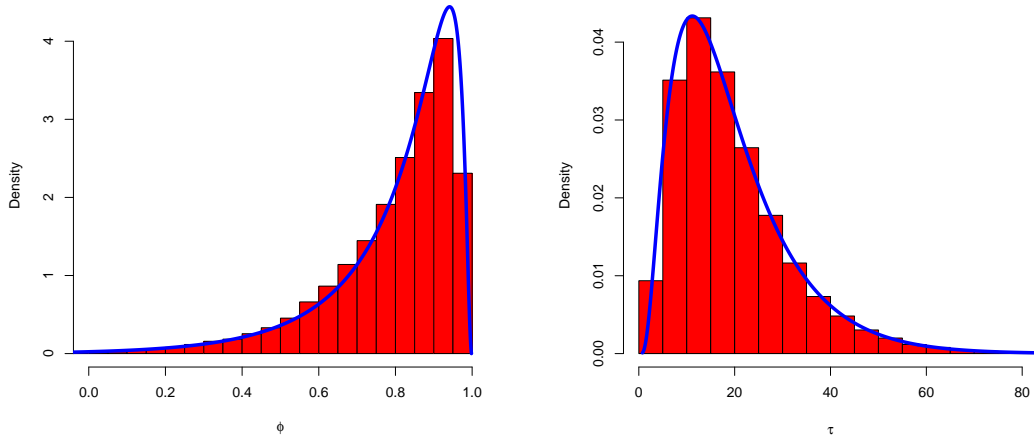


Figura 3.8: (Esquerda) $p(\phi|\mathbf{y})$ (Direita) $p(\tau|\mathbf{y})$: ■ MCMC (Histograma) — INLA

O custo computacional necessário para computar a aproximação de Laplace simplificada é baixo comparado com o custo de se obter a aproximação de Laplace, sendo que esta última só é computada quando a distância entre a aproximação Gaussiana e a de Laplace simplificado é considerada significativa, do modo como definido em Rue *et al.* (2009). A figura 3.10 mostra, para $(\eta_1, \dots, \eta_{50})$, resultados similares aos descritos para a marginal de μ , sendo que a aproximação de Laplace simplificada apresenta um excelente resultado quando comparada com a obtida pelo MCMC.

3.4 Discussão sobre o INLA

Como descrito nas seções anteriores, o INLA obtém as distribuições a posteriori marginais de cada um dos componentes do vetor paramétrico $\Psi = (\mathbf{x}, \boldsymbol{\theta})$ de forma precisa e eficiente, mesmo quando a dimensão $k = n + m$ desse vetor é muito alta (n entre 10.000 e 100.000). Sendo assim, é importante discutir o que faz o INLA ser tão bem sucedido, já que no capítulo 2 foi ressaltado que os métodos numéricos e analíticos disponíveis se tornam im-

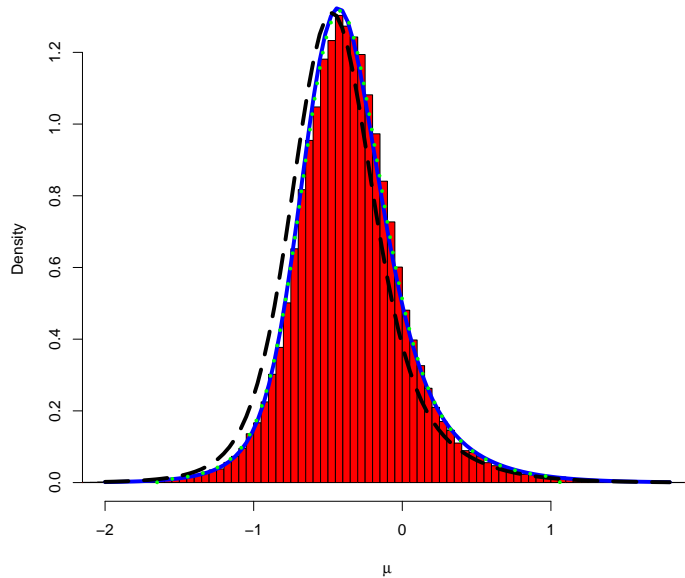


Figura 3.9: Posteriori marginal de μ : ■ (Histograma) MCMC --- Aproximação Gaussiana — Laplace simplificado ... Laplace

praticáveis quando a dimensão paramétrica é grande, devido ao alto custo computacional.

A estrutura Markoviana do campo latente traz dois grandes benefícios. O primeiro é que, de acordo com a definição 3.1, a estrutura Markoviana é codificada na matriz de precisão de \mathbf{x} , de modo que tal matriz é esparsa. Essa propriedade permite o uso de algoritmos específicos que aceleram em muito as operações que a envolvem. Um exemplo é a aproximação Gaussiana $\pi_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$, que tem papel fundamental no método e que seria inviável de se obter para n maior do que 10.000 caso a matriz de precisão fosse densa. O segundo benefício foi utilizado na aproximação de Laplace para $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$, ao considerar somente o conjunto $R_i(\boldsymbol{\theta})$ de elementos do campo latente que tem influência sobre x_i , como definido em (3.11), no denominador de (3.8). Desse modo, a estrutura Markoviana viabilizou a aproximação de Laplace para a condicional completa de x_i , e por consequência, a aproximação de

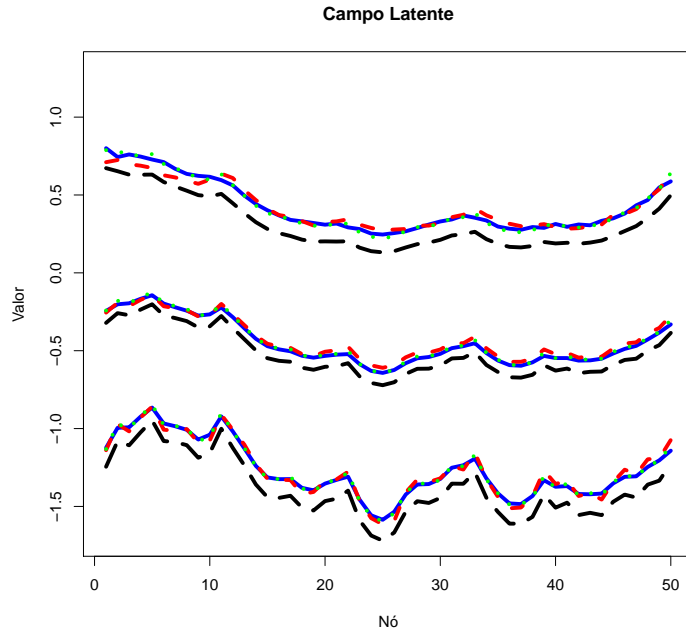


Figura 3.10: Sumário da distribuição a posteriori marginal de $\eta_1, \eta_2, \dots, \eta_{49}$ e η_{50} : média e IC 95% — MCMC — Aproximação Gaussiana — Laplace simplificado ... Laplace

Laplace simplificada também é beneficiada, já que é desenvolvida a partir de $\tilde{\pi}_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y})$.

A priori Gaussiana atribuída a $\mathbf{x}|\boldsymbol{\theta}$ tem um efeito não desprezível na distribuição da condicional completa, $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$, tornando-a ”bem comportada”. Tal característica é o principal motivo da aproximação Gaussiana $\pi_G(x_i|\boldsymbol{\theta}, \mathbf{y})$ dar resultados satisfatórios em muitos casos, como dito na seção 3.2.3. Além disso, essa característica é essencial para poder utilizar a média condicional, $E_{\tilde{\pi}_G}(\mathbf{x}_{-i}|x_i)$, em vez da moda, $\mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta})$, na aproximação de Laplace, dada pela equação (3.8), evitando assim a necessidade de realizar uma otimização para cada valor assumido por x_i .

Desse modo, chegamos a conclusão que o mérito do INLA é usufruir ao máximo das vantagens oferecidas pela estrutura Markoviana do campo latente e pela forma ”bem comportada” da condicional completa, caracterís-

tica herdada da priori Gaussiana atribuída à $\boldsymbol{x}|\boldsymbol{\theta}$.

Capítulo 4

Processos Pontuais

Espaço-Temporais: Inferência

Bayesiana Aproximada

Esse capítulo irá tratar sobre aspectos computacionais relacionados a inferência Bayesiana aproximada em processos pontuais espaço-temporais. A seção 4.1 irá introduzir aspectos teóricos e metodológicos sobre processos pontuais espaço-temporais. Sugere-se consultar Reis (2008) para um tratamento mais completo sobre o assunto. A seção 4.2 irá apresentar modelos dinâmicos Bayesianos para a modelagem dos componentes temporal, espacial e espaço-temporal de processos pontuais espaço-temporais propostos em Reis *et al.* (2010). A seção 4.3 irá apresentar cinco modelos baseados na metodologia da seção 4.2 para modelar o conjunto de dados apresentado em Faes *et al.* (2006) e propor utilizar o método INLA apresentado no capítulo 3 para realizar inferência Bayesiana aproximada nos modelos apresentados ao invés da abordagem usual que utiliza métodos baseados em amostragem como o MCMC.

4.1 Processos Pontuais Espaço-Temporais

Processos pontuais espaciais constituem a área da estatística responsável pelo estudo de observações de eventos em uma dada região, que pode ser um espaço geográfico. Estudos nessa área tem sido desenvolvidos tanto do ponto de vista teórico, onde as propriedades probabilísticas desses processos foram analisadas (ver Cox & Isham, 1980), quanto através do estudo de propriedades estatísticas, onde ênfase é dada para a estimação da intensidade dos eventos na região de interesse (Diggle *et al.*, 2003). Uma extensão relevante é analisar a variação das observações na dimensão temporal assim como na espacial. Brix & Diggle (2001) descreveram uma classe flexível de processos pontuais espaço-temporais baseados em modelos de Cox log-Gaussianos. Realizar inferência é uma tarefa difícil nesse contexto, sendo que Brix & Diggle (2001) utilizaram método dos momentos para tal propósito.

Reis *et al.* (2010) agregaram as abordagens de Brix & Diggle (2001) e Paez & Diggle (2009) para processos pontuais e a abordagem de Gelfand *et al.* (2005) para processos contínuos. O objetivo foi analisar processos pontuais espaço-temporais onde a sequência de superfícies de intensidade (variando no espaço) está ligada probabilisticamente através do tempo.

4.1.1 Formulação geral do modelo

Considere um processo pontual espaço-temporal $\{Z(s, t) : s \in S \text{ e } t \in [0, T]\}$ observado em uma região $S \subset \mathfrak{R}^d$ no espaço e $[0, T]$ no tempo. Tipicamente, $d = 2$ representando um processo pontual no plano, mas observações em espaços de outras dimensões podem ser normalmente considerados. Assuma que, para cada período fixo de tempo $t \in [0, T]$ e localização no espaço $s \in S$, a função de intensidade do processo é dada por $\lambda(s, t)$. Essa configuração faz com que a verossimilhança seja da forma

$$l(\lambda; Z) = \prod_{i=1}^r \lambda(s_i, t_i) \exp \left\{ - \int_0^T \int_S \lambda(s, t) ds dt \right\}, \quad (4.1)$$

onde r é o número de eventos observados nas localizações s_1, \dots, s_r e ocorridos nos tempos t_1, \dots, t_r .

O modelo para a intensidade é baseado na decomposição que permite efeitos temporais e espaciais específicos em adição às interações espaço-temporais. Assim, a intensidade é usualmente tratada na escala log e pode ser decomposta em

$$\log[\lambda(s, t)] = \mu(t) + \varsigma(s) + \phi(s, t), \quad (4.2)$$

onde $\mu(t)$ é a tendência temporal, comum para todas as localizações no espaço, $\varsigma(s)$ é o efeito puramente espacial, comum para todos os períodos de tempo, e $\phi(s, t)$ é o efeito espaço-temporal, que é específico para cada localização no espaço e período no tempo. Cada um desses efeitos pode ser decomposto em componentes determinísticas e estocásticas. O modelo é então completado com a especificação de distribuições a priori para cada um dos parâmetros desconhecidos de cada componente desses três efeitos.

4.1.2 Formulação do modelo discreto

O processo pontual apresentado na seção 4.1.1 está definido em um espaço contínuo. No entanto, é difícil realizar inferência através de uma verossimilhança definida em um espaço contínuo. Uma possível solução é aproximar o modelo através de uma discretização do espaço, onde os componentes espaciais ς e ϕ da função de intensidade (4.2) são assumidos constantes em cada pedaço da região de interesse. Essa abordagem foi seguida por Gamerman (1992) para inferência em processos pontuais através do tempo e é estendida aqui para o domínio espacial. Procedimento similar foi adotado, por exemplo, em Møller *et al.* (1998) e Brix (2001). Waagepetersen (2004) mostrou analiticamente que as distribuições a posteriori aproximadas das log-intensidades, calculadas através da discretização de processos Cox log-Gaussianos, convergem para as posteriores exatas quando as áreas das células da grade utilizada na discretização tendem para zero.

Será assumido que o processo da intensidade λ é constante nas N sub-regiões do espaço S , i.e., $\lambda(s, t) = \lambda_{i,j}$ para todo $s \in A_i$, para $i = 1, \dots, N$,

e $t \in (j - 1, j]$, para $j = 1, \dots, T$. $\{A_1, \dots, A_N\}$ é uma partição do espaço S . Com isso, a verossimilhança se torna

$$l(\lambda_{.,.}) \propto \prod_{i=1}^N \prod_{t=1}^T e^{-c_i \lambda_{i,t}} \lambda_{i,t}^{y_{i,t}}, \text{ onde} \quad (4.3)$$

$$\log(\lambda_{i,t}) = \varphi_{i,t} = \mu_t + \varsigma_i + \phi_{i,t} \quad (4.4)$$

onde $y_{i,t}$ contém a contagem dos eventos na i -ésima célula e no t -ésimo intervalo de tempo, c_i é a área de A_i , para $i = 1, \dots, N$ e μ_t , ς_i e $\phi_{i,t}$ são os componentes $\mu(t)$, $\varsigma(s)$ e $\phi(s, t)$ discretizados no tempo e no espaço.

4.2 Modelos Dinâmicos Bayesianos para Processos Pontuais Espaço-Temporais

A tendência temporal $\mu(t)$ em (4.2) pode, por exemplo, incorporar covariáveis variando no tempo. Um possível modelo determinístico pode ser representado por

$$\mu(t) = \mathbf{F}'_t \boldsymbol{\beta}, \quad (4.5)$$

onde $\boldsymbol{\beta}$ é o vetor de coeficientes da regressão, \mathbf{F}_t é o vetor das covariáveis variando no tempo, ou o próprio tempo, como em

$$\mu(t) = \beta_0 + \beta_1 t, \quad (4.6)$$

onde $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ e $\mathbf{F}_t = (1, t)'$. Usualmente a distribuição a priori para os coeficientes $\boldsymbol{\beta}$ é escolhida como sendo a distribuição Normal. Outro importante caso é a tendência temporal constante, onde a intensidade do processo não varia no tempo

$$\mu(t) = \beta_0. \quad (4.7)$$

A distribuição a priori Normal também é usualmente escolhida para β_0 . Em um modelo estocástico, a tendência temporal pode ter, por exemplo, a evolução dinâmica

$$\mu(t) = \mathbf{F}'_t \boldsymbol{\beta}(t), \quad (4.8)$$

$$\boldsymbol{\beta}(t) = G_t \boldsymbol{\beta}(t-1) + v(t), \quad v(t) \sim N(\mathbf{0}, \Omega_t), \quad (4.9)$$

onde $\beta(t)$ é o vetor de estados no tempo t , \mathbf{F}_t e \mathbf{G}_t são matrizes conhecidas. Um exemplo dessa especificação é o modelo dinâmico de primeira ordem

$$\mu(t) = \mu(t-1) + v(t), \quad v(t) \sim N(0, w^2), \quad t = 2, \dots, T, \quad \mu(1) \sim N(\mu_0, \tau_\mu^2), \quad (4.10)$$

Usualmente, a distribuição a priori de w^2 é especificada como uma distribuição Gama-Inversa.

O efeito espaço-temporal $\phi(s, t)$ em (4.2) é geralmente descrito como um processo espacial Gaussiano independente no tempo. Essa opção é um primeiro passo natural mas é interessante permitir dependência desses processos residuais através de períodos de tempo sucessivos. Essa possibilidade foi explorada em Reis *et al.* (2010) e incorpora dependência temporal nesses processos espaciais através de processos Gaussianos dinâmicos (Gamerman *et al.*, 2007). É assumido que $\phi(\cdot, t)$, $t = 1, \dots, T$ é um processo Gaussiano estacionário e isotrópico no espaço, e autoregressivo estacionário no tempo,

$$\phi(s, t) = \eta\phi(s, t-1) + w(s, t), \quad w(\cdot, t) \sim GP[\mathbf{0}; (1 - \eta^2)\sigma^2; \rho(\cdot; \kappa)], \quad (4.11)$$

onde $0 < \eta < 1$ é o parâmetro de correlação temporal e $\phi(\cdot, 1) \sim GP[\mathbf{0}; \sigma^2; \rho(\cdot; \kappa)]$.

O efeito puramente espacial $\varsigma(s)$ em (4.2) não será usado na aplicação da seção 4.3, mas quando usado, pode ser decomposto em componente determinístico $\varsigma_d(s)$ e componente estocástico $\varsigma_e(s)$. O componente determinístico $\varsigma_d(s)$ pode incorporar covariáveis que variem no espaço, mas que são constantes no tempo:

$$\varsigma_d(s) = \mathbf{x}(s)' \boldsymbol{\alpha}, \quad (4.12)$$

onde $\mathbf{x}(s)$ é o vetor de covariáveis observadas na localização s e $\boldsymbol{\alpha}$ é o vetor de pesos. O componente estocástico $\varsigma_e(s)$ pode ser definido, por exemplo, por um processo Gaussiano estacionário e isotrópico na região de interesse, representado por

$$\varsigma_e(\cdot) \sim GP[\mu_\varsigma, \gamma^2, \rho_\varsigma(\cdot; \tau)], \quad (4.13)$$

onde μ_ς é o vetor de médias, γ^2 é a variância do processo espacial, e $\rho_\varsigma(\cdot; \tau)$ é a função de correlação espacial.

4.3 Inferência Bayesiana Aproximada: Aspectos Computacionais

O conjunto de dados descrito a seguir demonstra o contexto ideal para aplicar a modelagem apresentada na seção 4.2. Isso fornece uma ótima oportunidade para mostrar que o método INLA, descrito no capítulo 3, é ideal para tal configuração.

Evolução Espaço-Temporal de Impulsos Elétricos no Intestino Delgado

O intestino delgado finaliza o processo de digestão, absorve os nutrientes e conduz os resíduos para o intestino grosso. As células nervosas existentes na parede do intestino delgado emitem sinais que controlam os movimentos coordenados de contração de sua parede muscular, fazendo com que o conteúdo resultante da digestão seja empurrado ao longo do trato intestinal.

Dois padrões de atividade elétrica são importantes neste processo: as ondas lentas (*slow-waves*) e os impulsos (*spike potentials*). Uma onda lenta age como um sinal de marca-passo que induz o músculo à contração. Os impulsos superimpostos às ondas lentas determinam a força e duração da contração muscular.

Uma questão de interesse sobre este processo é saber se existem áreas com incidência mais alta de impulsos, comparadas com outras áreas. Outra questão é o entendimento das características temporais e espaciais da ocorrência de impulsos durante sucessivas ondas lentas. Especificamente, deseja-se saber se as áreas com atividades elétricas mais intensas são as mesmas ao longo das ondas lentas sucessivas. Desse modo, a modelagem da distribuição espaço-temporal dos impulsos pode ajudar no entendimento do mecanismo de geração e propagação dos movimentos intestinais.

No experimento descrito em Faes *et al.* (2006), um segmento do intestino delgado foi removido de sete gatos e suas atividades elétricas espontâneas

foram observadas durante o período de um minuto, usando-se 240 eletrodos dispostos em uma grade regular (10×24) na superfície do tecido. Os autores disponibilizaram o conjunto de dados contendo as atividades elétricas de somente um gato, com a medida do número de impulsos em cada célula da grade em 13 ondas lentas sucessivas.

Três diferentes especificações para a tendência temporal μ_t foram especificadas, para $t = 1, \dots, 13$:

$$\mu_t = \mu, \quad (4.14)$$

$$\mu_t = \beta_0 + \beta_1 t, \quad (4.15)$$

$$\mu_t = \mu_{t-1} + v_t, \quad v_t \sim N(0, w^2). \quad (4.16)$$

O efeito espaço-temporal $\boldsymbol{\phi}_{[:,t]} = (\phi_{[1,t]}, \dots, \phi_{[240,t]})'$ é modelado como um processo Gaussiano autoregressivo, estacionário no tempo, $t = 2, \dots, 13$:

$$\boldsymbol{\phi}_{[:,t]} = \eta \boldsymbol{\phi}_{[:,t-1]} + \boldsymbol{w}_{[:,t]}, \quad \boldsymbol{w}_{[:,t]} \sim N(\mathbf{0}, (1 - \eta^2)\sigma^2 R_\kappa), \quad (4.17)$$

com $0 < \eta < 1$, e os elementos na matriz de correlação espacial R_κ são definidos pela função de correlação exponencial. O modelo é completado com a priori $\boldsymbol{\phi}_{[:,1]} \sim N(\mathbf{0}, \sigma^2 R_\kappa)$, e será denotado por (A).

A combinação das três especificações para μ com os efeitos espaço-temporais $\boldsymbol{\phi}$ gera três modelos, denotados 1A, 2A e 3A respectivamente. A adequabilidade desses três modelos propostos em (Reis *et al.*, 2010) para o efeito espaço-temporal é mensurado ao se comparar com alternativas usadas na literatura. A alternativa (B) assume que esses efeitos são puramente espacial, constante através do tempo ou, em outras palavras, $\phi_{[i,t]} = \varsigma_{[i]}$, $t = 1, \dots, 13$, é modelado através de

$$\varsigma_{[.]} = (\varsigma_{[1]}, \dots, \varsigma_{[240]})' \sim N(\mathbf{0}, \sigma^2 R_\kappa). \quad (4.18)$$

A alternativa (C) mantém a relevância das mudanças temporais mas não considera correlação temporal entre os efeitos ou, em outras palavras, o modelo assume que os efeitos $\phi_{[i,t]}$ são independentes no tempo:

$$\boldsymbol{\phi}_{[:,t]} \sim N(\mathbf{0}, \sigma^2 R_\kappa), \quad t = 1, \dots, 13 \quad (4.19)$$

Esse é um caso especial do modelo (A) quando o parâmetro de correlação temporal η é igual a zero. Os modelos 1B e 1C são obtidos ao combinar a tendência temporal constante (4.14) com as alternativas (B) e (C) respectivamente.

Inferência Bayesiana Aproximada utilizando INLA

Os cinco modelos definidos acima podem ser estimados utilizando MCMC, porém, mesmo quando aplicado a problemas de dimensão moderada, como o conjunto de dados analisado em Faes *et al.* (2006) descrito acima, o tempo computacional necessário para obter a convergência das cadeias de Markov pode levar semanas para cada um dos modelos (ver Reis, 2008). Tal procedimento se torna então inviável de ser realizado na prática, especialmente quando se pretende comparar diversos modelos, como ocorre agora com os modelos 1A, 2A, 3A, 1B e 1C.

Felizmente, a estrutura dos modelos aqui apresentados se enquadra na estrutura requerida para a aplicação do INLA. Neste capítulo, será demonstrado que o modelo 1A tem a requerida forma e o mesmo pode ser visto para os modelos 2A e 3A no apêndice B.1.

O modelo 1A é dado por:

$$l(\boldsymbol{\lambda}) \propto \prod_{i=1}^N \prod_{t=1}^T e^{-c_i \lambda_{i,t} \lambda_{i,t}^{y_{i,t}}}, \text{ onde}$$

$$\log(\lambda_{i,t}) = \varphi_{i,t} = \mu_t + \phi_{i,t}$$

onde μ_t é dado pela equação (4.14) e $\phi_{[.,t]}$ segue a dinâmica descrita em (4.17). Neste caso, denotando $\boldsymbol{\varphi}_{[.,t]} = (\varphi_{[1,t]}, \dots, \varphi_{[240,t]})'$, o campo latente é definido como $\boldsymbol{x} = (\boldsymbol{\varphi}_{[.,1]}, \dots, \boldsymbol{\varphi}_{[.,13]}, \boldsymbol{\mu})$ e $\boldsymbol{\theta} = (\eta, \sigma^2, \kappa)$. Será assumido a priori que os hiperparâmetros $\boldsymbol{\theta}$ são independentes e a distribuição a priori para η será uma $U[0, 1]$, e vamos atribuir uma $GI(1, 1)$ para σ^2 e para κ . Após

algumas contas, percebe-se que

$$\varphi_{[:,1]} \sim N(\mu \mathbf{1}, \sigma^2 \mathbf{R}_\kappa) \quad (4.20)$$

$$\begin{aligned} \varphi_{[:,t]} &= \eta \varphi_{[:,t-1]} + (1 - \eta) \mu \mathbf{1} + \epsilon_{[:,t]}, \\ &\text{onde } \epsilon_{[:,t]} \sim N(\mathbf{0}, (1 - \eta^2) \sigma^2 \mathbf{R}_\kappa). \end{aligned} \quad (4.21)$$

Utilizando o fato que $\mu \sim N(0, 10)$ e as equações (4.20) e (4.21), temos que a distribuição a priori $\pi(\mathbf{x}|\boldsymbol{\theta})$ é dada por:

$$\pi(\mathbf{x}|\boldsymbol{\theta}) = \pi(\mu) \pi(\varphi_{[:,1]}|\mu) \prod_{t=2}^{13} \pi(\varphi_{[:,t]}|\mu, \varphi_{[:,t-1]}) \quad (4.22)$$

$$\propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta}) \mathbf{x} \right\}, \text{ onde} \quad (4.23)$$

$$\mathbf{Q}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{Q}_{1,1} & \mathbf{Q}_{1,2} & 0 & \cdots & 0 & \mathbf{Q}_{1,\mu} \\ \mathbf{Q}_{2,1} & \mathbf{Q}_{2,2} & \mathbf{Q}_{2,3} & \cdots & 0 & \mathbf{Q}_{2,\mu} \\ 0 & \mathbf{Q}_{3,2} & \mathbf{Q}_{3,3} & \cdots & 0 & \mathbf{Q}_{3,\mu} \\ 0 & 0 & \mathbf{Q}_{4,3} & \cdots & 0 & \mathbf{Q}_{4,\mu} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbf{Q}_{12,13} & \mathbf{Q}_{12,\mu} \\ 0 & 0 & 0 & \cdots & \mathbf{Q}_{13,13} & \mathbf{Q}_{13,\mu} \\ \mathbf{Q}_{\mu,1} & \mathbf{Q}_{\mu,2} & \mathbf{Q}_{\mu,3} & \cdots & \mathbf{Q}_{\mu,13} & \mathbf{Q}_{\mu,\mu} \end{pmatrix}$$

onde,

$$\mathbf{Q}_{1,1} = [1 + \eta^2(1 - \eta^2)^{-1}] \sigma^{-2} \mathbf{Q}_\kappa;$$

$$\mathbf{Q}_{i,i} = (1 + \eta^2)(1 - \eta^2)^{-1} \sigma^{-2} \mathbf{Q}_\kappa, \quad i = 2, 3, \dots, 12;$$

$$\mathbf{Q}_{13,13} = (1 - \eta^2)^{-1} \sigma^{-2} \mathbf{Q}_\kappa;$$

$$\mathbf{Q}_{\mu,\mu} = R_1^{-1} + [1 + 12(1 - \eta)^2(1 - \eta^2)^{-1}] \sigma^{-2} \mathbf{1}' \mathbf{Q}_\kappa \mathbf{1};$$

$$\mathbf{Q}_{i,i+1} = \mathbf{Q}_{i+1,i} = -\eta(1 - \eta^2)^{-1} \sigma^{-2} \mathbf{Q}_\kappa, \quad i = 1, 2, \dots, 12;$$

$$\mathbf{Q}_{1,\mu} = [\eta(1 - \eta)(1 - \eta^2)^{-1} - 1] \sigma^{-2} \mathbf{Q}_\kappa \mathbf{1} \quad e \quad \mathbf{Q}_{\mu,1} = [\eta(1 - \eta)(1 - \eta^2)^{-1} - 1] \sigma^{-2} \mathbf{1}' \mathbf{Q}_\kappa;$$

$$\mathbf{Q}_{i,\mu} = (1 - \eta)(\eta - 1)(1 - \eta^2)^{-1} \sigma^{-2} \mathbf{Q}_\kappa \mathbf{1}, \quad i = 2, 3, \dots, 12;$$

$$\mathbf{Q}_{\mu,i} = (1 - \eta)(\eta - 1)(1 - \eta^2)^{-1} \sigma^{-2} \mathbf{1}' \mathbf{Q}_\kappa, \quad i = 2, 3, \dots, 12;$$

$$\mathbf{Q}_{13,\mu} = -(1 - \eta)(1 - \eta^2)^{-1} \sigma^{-2} \mathbf{Q}_\kappa \mathbf{1} \quad e \quad \mathbf{Q}_{\mu,13} = -(1 - \eta)(1 - \eta^2)^{-1} \sigma^{-2} \mathbf{1}' \mathbf{Q}_\kappa,$$

sendo \mathbf{Q}_κ o inverso de \mathbf{R}_κ

Além disso, temos que a verossimilhança na localização i e no tempo t só depende do campo latente \mathbf{x} através de um elemento, $\varphi_{i,t}$. Essa última característica, junto com o fato de $\mathbf{x}|\boldsymbol{\theta}$ ser um CAMG com a matriz de precisão $\mathbf{Q}(\boldsymbol{\theta})$ esparsa mostra que o modelo 1A pode ser estimado com o INLA. O mesmo acontece para os outros quatro modelos e a tabela 4.1 mostra as médias a posteriori e o intervalo de 95% de credibilidade para os hiperparâmetros, assim como o DIC (Spiegelhalter *et al.*, 2002) para cada um dos modelos. O DIC mostra forte evidência em favor do modelo 1A, com a tendência temporal constante e correlação temporal entre os efeitos espaciais.

As posteriores marginais dos hiperparâmetros $\boldsymbol{\theta}$ e do μ (considerado no campo latente) podem ser visualizadas na figura 4.1. Para a posteriori marginal de μ foi usada a aproximação Gaussiana e a aproximação de Laplace simplificada.

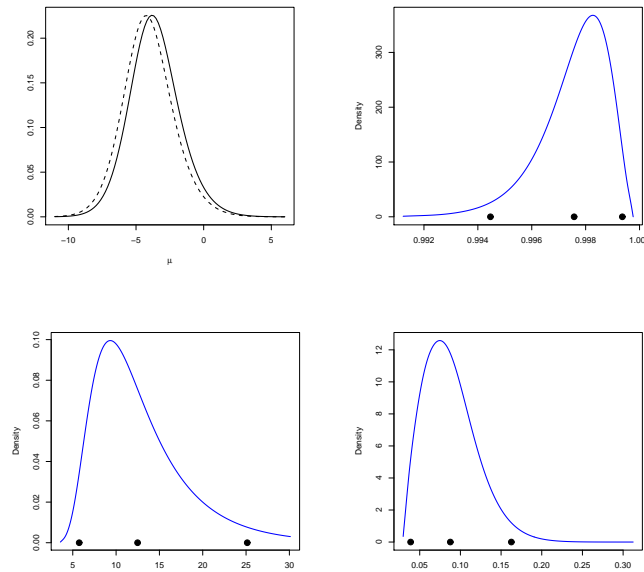


Figura 4.1: Posteriores marginais: (Acima a esquerda) μ com $--$ Aproximação Gaussiana $-$ Laplace simplificado; (Acima a direita) η ; (Abaixo a esquerda) σ^2 ; (Abaixo a direita) κ .

A concentração de η ao redor de valores perto de 1 mostra forte persistência dos efeitos espaciais ao longo do tempo, o que descredita o modelo 1C, que é igual ao modelo 1A com $\eta = 0$. Valores de σ^2 estão concentrados em valores distantes de zero, sendo essa mais uma evidência contra o modelo 1B, cujo os efeitos espaciais são constantes ao longo do tempo. Esses resultados reforçam a importância dos efeitos espaciais variarem de forma suave ao longo do tempo.

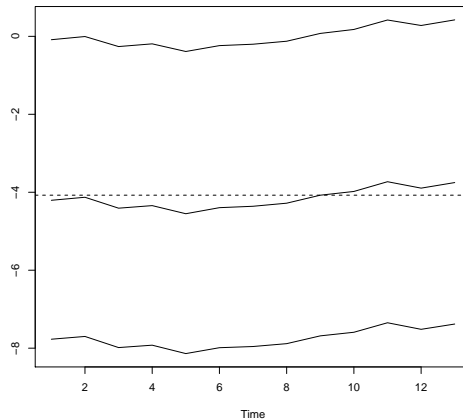


Figura 4.2: Médias a posteriori para μ_t no modelo 3A (linha sólida) e para μ do modelo 1A (linha tracejada) e intervalos de credibilidade de 95% para μ_t .

Desse modo, o modelo proposto com efeitos espaciais específicos para cada período de tempo e estrutura autoregressiva, e com a componente de tendência temporal constante se mostrou o mais apropriado para esse conjunto de dados. A figura 4.3 mostra, sob esse modelo, os mapas dos padrões espaciais da log intensidade e sua variabilidade. Percebe-se claramente que a variabilidade dos efeitos ϕ também possui uma estrutura espacial.

A intensidade dos impulsos através de sucessivas ondas lentas é caracterizado pelo efeito temporal μ_t . Todos os modelos estimam μ_t ao redor de -4 . Sob o modelo 1A, a constante μ foi estimada em $-4,1$. Sob o modelo 2A, o intervalo de credibilidade do coeficiente linear β_1 é muito amplo e centrado ao redor de zero, indicando um efeito não significativo do tempo. Portanto,

não é justificável considerar uma tendência linear para o componente temporal do modelo quando aplicado a esse conjunto de dados, e o modelo 1A é preferível ao modelo 2A. Similarmente, as estimativas de μ_t mostradas na figura 4.2 apresentam um padrão estável bem concentrado aos redor do valor constante μ obtido com o modelo 1A. Esse padrão fornece mais evidências sobre a preferência do modelo 1A quando comparado com o modelo 3A.

A estimação de cada um dos modelos aqui apresentados levou horas (18 horas para o modelo 1A) utilizando o INLA, sendo que seria necessário semanas para obter os resultados utilizando um esquema MCMC, sendo que, mesmo após semanas de simulação ainda é difícil diagnosticar a convergência das cadeias, dando possibilidade de se utilizar resultados que não são totalmente confiáveis.

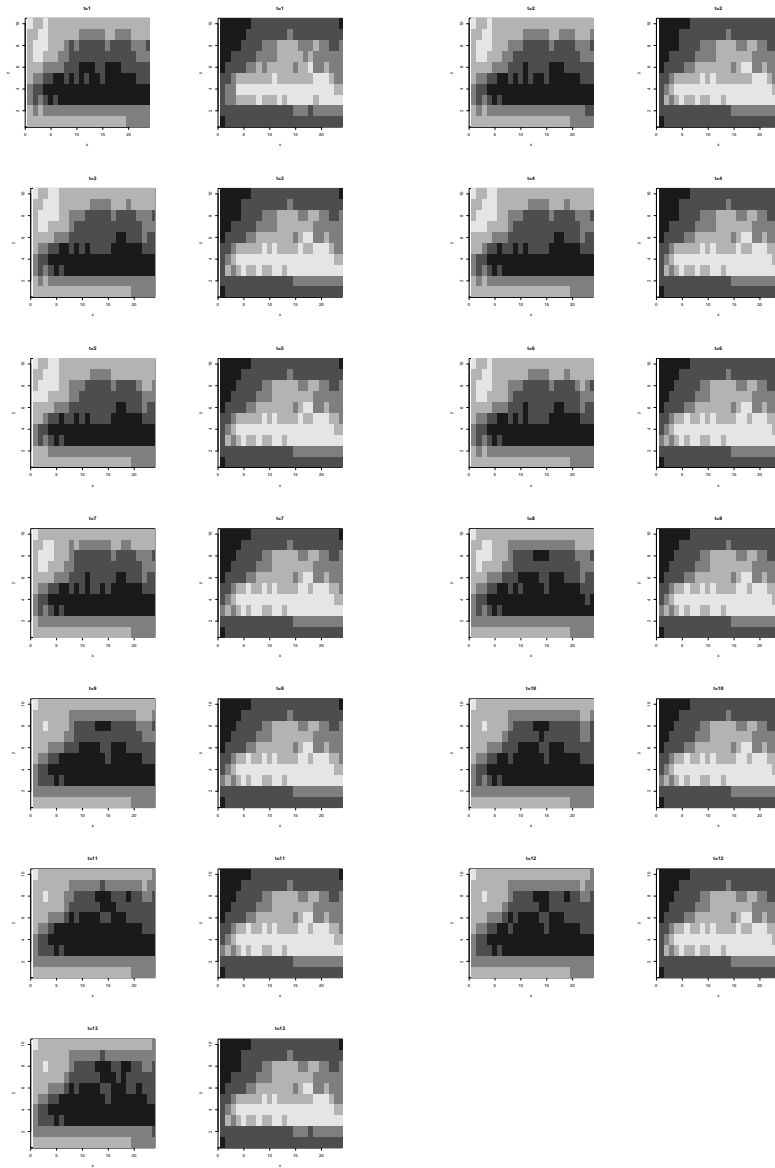


Figura 4.3: log intensidade $\varphi_{[i,t]}$, $t = 1, \dots, 13$ no modelo 1A: (Esquerda) Mapas das médias a posterioris; (Direita) Mapas das variâncias a posteriori.
 Médias: $\bullet < -5.93$ $\bullet [-5.93, -4,17)$ $\bullet [-4,17, -2,42)$ $\bullet [-2,42, -0,66)$ $\bullet \geq -0,66$
 Variâncias: $\bullet < -0,10$ $\bullet [0,10, 0,25)$ $\bullet [0,25, 0,66)$ $\bullet [0,66, 1,72)$ $\bullet \geq 1,72$

Tabela 4.1: Médias a posteriori e intervalos de credibilidade de 95% para os hiperparâmetros do modelo.

	Model				
	1A: (4.14)-(4.17)	2A: (4.15)-(4.17)	3A: (4.16)-(4.17)	1B: (4.14)-(4.18)	1C: (4.14)-(4.19)
μ	-4.08 [-7.69 ; -0.05]			-3.70 [-4.66; -2.92]	3.91 [-7.59; 0.94]
β_0		-4.24 [-7.91 ; 0.02]			
β_1		0.042 [-0.957; 1.042]			
σ^2	12.48 [5.74 ; 25.14]	13.15 [5.67 ; 30.60]	12.29 [5.16 ; 27.54]	4.80 [3.24; 7.37]	15.71 [4.69; 58.21]
κ	0.087 [0.038; 0.162]	0.086 [0.031; 0.165]	0.093 [0.036; 0.184]	0.11 [0.07; 0.16]	0.09 [0.02; 0.21]
η	0.997 [0.994; 0.999]	0.997 [0.994; 0.999]	0.997 [0.993; 0.999]		
ω^2			0.237 [0.106; 0.520]		
<i>DIC</i>	2537.268	2570.283	2570.597	3207.595	2560.839

Capítulo 5

INLA - Extensões

Como dito anteriormente, para aplicar a metodologia descrita no capítulo 3 é necessário que cada elemento dos dados, y_i , esteja conectado a somente um elemento do campo latente, de modo que a verossimilhança possa ser escrita como

$$\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \prod_{t=1}^{n_d} \pi(y_t|x_t, \boldsymbol{\theta}). \quad (5.1)$$

Este capítulo fará extensões na metodologia apresentada no capítulo 3 de modo a eliminar essa restrição.

A seção 5.1 irá apresentar a estrutura de modelo mais geral que, antes das extensões propostas neste capítulo, não podia ser analisada utilizando o método INLA como descrito no capítulo 3. A seção 5.2 irá de fato analisar e propor soluções para os problemas computacionais e metodológicos que surgem ao tentar aplicar o INLA na estrutura de modelo apresentada na seção 5.1. A seção 5.3 irá apresentar os resultados da aplicação da metodologia descrita na seção 5.2 para um modelo específico que se enquadra na estrutura descrita na seção 5.1, denominado modelo de volatilidade estocástica assimétrica.

5.1 Introdução

O capítulo 3 mostrou a estrutura de modelos em que o método INLA (Rue *et al.*, 2009) é aplicável, que pode ser resumida do seguinte modo:

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}), \quad (5.2)$$

$$\mathbf{x}|\boldsymbol{\theta} \sim \pi(\mathbf{x}|\boldsymbol{\theta}), \quad (5.3)$$

$$\mathbf{y}|\boldsymbol{\theta}, \mathbf{x} \sim \pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}), \quad (5.4)$$

onde $\pi(\mathbf{x}|\boldsymbol{\theta})$ tem distribuição Normal multivariada e $\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$ tem propriedades de independência condicional de modo que

$$\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n_d} \pi(y_i|x_i, \boldsymbol{\theta}). \quad (5.5)$$

Como dito anteriormente, a estrutura descrita acima engloba muitos dos modelos Bayesianos estruturados utilizados atualmente e será denominada de **Estrutura 1** ao longo deste capítulo. No entanto, exemplos importantes não satisfazem a condição (5.5), onde cada dado y_i só pode estar conectado com um elemento do campo latente x_i . Esse fato é melhor visualizado através do seguinte exemplo.

5.1.1 Modelo de Volatilidade Estocástica Assimétrica

Na seção 3.3 foi definido o modelo de volatilidade estocástica do seguinte modo:

$$y_t = \sqrt{\exp(\eta_t)}\epsilon_t, \quad \epsilon_t \sim N(0, 1) \quad (5.6)$$

$$\eta_t = \mu + \phi(\eta_{t-1} - \mu) + v_t, \quad v_t \sim N(0, \sigma^2), \quad (5.7)$$

para $t = 1, \dots, n_d$.

O Modelo é completado com distribuições a priori independentes:

$$\eta_1 \sim N\left(\mu, \frac{\sigma^2}{1 - \phi^2}\right), \quad \tau \sim G(\alpha_\tau, \beta_\tau), \quad \phi^* \sim N(a_1, R_1) \text{ e } \mu \sim N(0, R_2) \quad (5.8)$$

onde $\phi^* = \log \left\{ \frac{1+\phi}{1-\phi} \right\}$ and $\tau = \frac{1}{\sigma^2}$.

Uma característica frequentemente observada em estudos financeiros é que a volatilidade responde de forma assimétrica a choques positivos e negativos nos retornos. Vários motivos tem sido propostos na literatura para explicar a presença dessa relação assimétrica entre volatilidade e retornos. Dois dos mais citados, Black (1976) e Christie (1982), sugerem que, quando o retorno do ativo é positivo (negativo), ele se torna menos (mais) arriscado, com isso diminuindo (aumentando) a volatilidade. Em outras palavras, há uma correlação negativa entre retornos e volatilidade.

O modelo de volatilidade estocástica assimétrica univariado foi introduzido pela primeira vez em Harvey & Shephard (1996), permitindo que os erros das equações (5.6) e (5.7) sejam negativamente correlacionados. Formalmente,

$$\text{Corr}(\epsilon_t, v_{t+1}) = \rho, \text{ com } \rho < 0. \quad (5.9)$$

Tal modelo é estimado com o método da quase-verossimilhança em Harvey & Shephard (1996), e com o MCMC em Meyer & Yu (2000). Esse parece ser também um modelo ideal para a aplicação do INLA como descrito no capítulo 3, porém, a partir de (5.6), (5.7) e (5.9) temos que

$$\begin{pmatrix} y_t \\ \eta_{t+1} \end{pmatrix} \Bigg| \eta_t, \mu, \theta \sim N \left(\begin{pmatrix} 0 \\ \mu + \phi(\eta_t - \mu) \end{pmatrix}, \begin{pmatrix} \exp(\eta_t) & S_t \\ S_t & \sigma^2 \end{pmatrix} \right)$$

onde

$$S_t = \text{Cov}(y_t, \eta_{t+1} | \eta_t, \mu, \theta) = \exp(\eta_t/2) \sigma \rho$$

e usando a propriedade 3 da seção 1.6 temos a seguinte função de verossimilhança:

$$y_t | \eta_{t+1}, \eta_t, \mu, \theta \sim N(m_t, C_t), \quad (5.10)$$

com

$$m_t = \rho \exp(\eta_t/2) [\eta_{t+1} - \mu - \phi(\eta_t - \mu)] / \sigma \quad (5.11)$$

$$C_t = \exp(\eta_t) (1 - \rho^2). \quad (5.12)$$

para $t = 1, \dots, n_d - 1$ e

$$y_{n_d} | \eta_{n_d}, \mu, \theta \sim N(0, \exp(\eta_{n_d})). \quad (5.13)$$

Com isso, vemos pela equação (5.10) que cada dado y_t depende de mais de um elemento do campo latente $\mathbf{x} = (\eta_1, \dots, \eta_{n_d}, \mu)$. Mais especificamente, cada y_t depende de η_t, η_{t+1} e μ para $t = 1, \dots, n_d - 1$. Desse modo a condição (5.5) não é satisfeita neste caso, assim como em tantos outros, sendo necessário ajustes, tanto metodológicos quanto computacionais, para poder realizar inferência Bayesiana aproximada determinística de forma precisa e eficiente como feito no capítulo 3. Tais ajustes, inéditos na literatura, serão o assunto da próxima seção.

5.2 Extensões

Essa seção irá descrever como aplicar o método INLA quando a propriedade de independência condicional de $\pi(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta})$ é representada por

$$\pi(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n_d} \pi(y_i | x_{\mathbf{I}_i}, \boldsymbol{\theta}). \quad (5.14)$$

ou seja, diferentemente do representado em (5.5), cada dado y_i pode estar conectado a um conjunto de elementos \mathbf{I}_i do campo latente \mathbf{x} , para $i = 1, \dots, n_d$. Esse novo contexto será denominado **Estrutura 2**. Os desafios ao lidar com esse novo contexto ocorre em duas frentes, descritas nas seções 5.2.1 e 5.2.2, ambas relacionadas ao fato da matriz $\mathbf{C}(\mathbf{x})$ da aproximação Gaussiana dada pela equação (1.22) não ser mais uma matriz diagonal, como pode ser visto a seguir com a matriz $\mathbf{C}(\mathbf{x})$ (triângulo superior omitido dado a simetria) do modelo de volatilidade estocástica assimétrica da seção 5.1.1.

2. A aproximação Gaussiana no ponto \mathbf{x}_0 terá matriz de precisão $\mathbf{Q}^*(\mathbf{x}_0) = (\mathbf{Q} + \mathbf{C}(\mathbf{x}_0))$ e média $\mathbf{x}_1 = \mathbf{Q}^{*-1}(\mathbf{x}_0)\mathbf{b}(\mathbf{x}_0)$.
3. Efetue uma aproximação Gaussiana no ponto \mathbf{x}_1 obtido no item anterior, que terá uma matriz de precisão $\mathbf{Q}^*(\mathbf{x}_1) = (\mathbf{Q} + \mathbf{C}(\mathbf{x}_1))$ e média $\mathbf{x}_2 = \mathbf{Q}^{*-1}(\mathbf{x}_1)\mathbf{b}(\mathbf{x}_1)$.
4. Repita o passo 3 até obter convergência.

No entanto, ao trabalhar com a **Estrutura 2**, há casos onde, para certos valores iniciais $\{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_u\}$ do procedimento acima, a matriz de precisão $\mathbf{Q}^*(\mathbf{x}_t)$, $t = 1, \dots, u$ não é positiva-definida. Isso acontece porque agora, diferente do que ocorre no capítulo 3, a matriz $\mathbf{C}(\mathbf{x}_t)$ não é mais diagonal, como pode ser visto em (5.15). Com isso, a chance de, em algum ponto \mathbf{x}_t , essa matriz não ser positiva-definida aumenta bastante quando comparada com uma matriz diagonal, onde ter os elementos da diagonal maiores ou iguais a zero é condição suficiente para evitar essa situação. Como o importante para o algoritmo de otimização é que $\mathbf{Q}^*(\mathbf{x}_t) = (\mathbf{Q} + \mathbf{C}(\mathbf{x}_t))$ seja positiva-definida, há casos em que mesmo $\mathbf{C}(\mathbf{x}_t)$ não satisfazendo essa condição, a combinação com a matriz \mathbf{Q} de precisão a priori de \mathbf{x} , faz com que $\mathbf{Q}^*(\mathbf{x}_t)$ seja positiva-definida. Além disso, esse é um problema que ocorre somente no estágio inicial da otimização de $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$, quando \mathbf{x}_t está distante da moda \mathbf{x}^* e mesmo assim somente quando $\boldsymbol{\theta}$ está situado em uma região de baixa densidade $\pi(\boldsymbol{\theta}|\mathbf{y})$. De todo modo, é necessário tomar providências para que, caso essa situação ocorra, o algoritmo de otimização não seja interrompido com um erro.

Desse modo, o primeiro passo seria tornar o algoritmo mais robusto do modo como descrito na seção 1.10, ao utilizar algum método de *busca em linha* dentro do algoritmo de otimização, como representado no algoritmo 2 da mesma seção. Para os exemplos deste capítulo foi usado o método *busca em linha retroativo*, representado no algoritmo 3 da seção 1.10, por sua simplicidade e eficiência.

Mesmo tornando o algoritmo de otimização mais robusto, ainda temos que resolver os casos onde a matriz $\mathbf{Q}^*(\mathbf{x}_t)$ não é positiva-definida. A idéia é transformar uma matriz que não é positiva-definida em uma que seja sem alterar muito sua estrutura. Há vários métodos capazes de realizar essa tarefa (por exemplo, Knol & ten Berge, 1989; Cheng & Higham, 1998; Higham, 2002), porém nossa abordagem é simples e eficiente, bastando aumentar os elementos da diagonal de $\mathbf{C}(\mathbf{x}_t)$ até que essa matriz se torne positiva-definida (o mesmo resultado pode ser obtido ao ir reduzindo os elementos fora da diagonal). Uma vez atingido o ponto onde a matriz $\mathbf{C}(\mathbf{x}_t)$ se torna positiva-definida basta continuar com a otimização normalmente. Esse método funciona porque esse problema só acontece nos estágios iniciais da maximização. Então, ao aumentar os elementos da diagonal, a matriz de precisão torna-se positiva-definida sem alterar as relações de dependência codificadas na matriz. Isso permite que o método de otimização continue operando até chegar em uma região suficientemente perto da moda, onde esse problema de não ser positiva-definida pare de ocorrer. A partir daí o método passa a funcionar normalmente sem a intervenção na matriz $\mathbf{C}(\mathbf{x})$.

Após essas alterações, o algoritmo para a otimização de $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ pode ser descrito da seguinte forma:

1. Escolha um ponto inicial \mathbf{x}_0 .
2. Verifique se $\mathbf{Q}^*(\mathbf{x}_0) = (\mathbf{Q} + \mathbf{C}(\mathbf{x}_0))$ é positiva-definida, caso afirmativo, prossiga para o passo 4, caso contrário, execute o passo 3.
3. Multiplique os elementos da diagonal de $\mathbf{C}(\mathbf{x}_0)$ por um fator F maior que 1, digamos $F = 2$, e então retorne para o passo 2.
4. A aproximação Gaussiana no ponto \mathbf{x}_0 terá matriz de precisão $\mathbf{Q}^*(\mathbf{x}_0) = (\mathbf{Q} + \mathbf{C}(\mathbf{x}_0))$ e média $\mathbf{x}_1 = \mathbf{Q}^{*-1}(\mathbf{x}_0)\mathbf{b}(\mathbf{x}_0)$.
5. Faça $\mathbf{x}_0 = \mathbf{x}_0 + t(\mathbf{x}_1 - \mathbf{x}_0)$, onde t é obtido pelo método *busca em linha retroativo* e retorne para o passo 2 até obter convergência.

Ao utilizar o algoritmo de otimização acima, os problemas de instabilidade numérica na maximização de $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ são resolvidos e com isso, a aproximação Gaussiana pode ser executada na **Estrutura 2** do mesmo modo como na **Estrutura 1**. No entanto, será visto que a aproximação de Laplace simplificada não pode ser obtida como descrito no capítulo 3. A seção 5.2.2 irá propor um modo de obter tal aproximação no contexto da **Estrutura 2**.

5.2.2 Aproximação de Laplace Simplificado

No capítulo 3 a aproximação de Laplace simplificada foi obtida efetuando-se uma expansão de Taylor ao redor da média $\mu_i(\boldsymbol{\theta})$ de $\pi_G(x_i|\boldsymbol{\theta}, \mathbf{y})$ até os termos de terceira ordem no numerador e até o termo de primeira ordem no denominador de

$$\tilde{\pi}_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}_{-i}=E_{\pi_G}(\mathbf{x}_{-i}|x_i)} \quad (5.18)$$

Temos que o denominador de (5.18) pode ser escrito como

$$\begin{aligned} \log \tilde{\pi}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y}) \Big|_{\mathbf{x}_{-i}=E_{\pi_G}(\mathbf{x}_{-i}|x_i)} &\stackrel{c}{\propto} \frac{1}{2} \log |\mathbf{H} + \mathbf{C}_{-i}(\mathbf{x}_{-i}(x_i))| \\ &= f(x_i) \end{aligned} \quad (5.19)$$

onde \mathbf{H} é a matriz de precisão a priori do CAMG com a i -ésima coluna e linha deletada e $\mathbf{C}_{-i}(\mathbf{x}_{-i}(x_i))$, daqui em diante denotada por $\mathbf{C}(x_i)$ para não carregar a notação, é a matriz (não mais diagonal) da aproximação Gaussiana com a i -ésima linha e coluna excluídas e avaliada no ponto $\mathbf{x}_{-i}(x_i) = E_{\pi_G}(\mathbf{x}_{-i}|x_i)$.

A equação (5.19) significa que, para cada valor de x_i , teríamos que fatorar uma matriz de dimensão $(n-1) \times (n-1)$, o que é muito caro, já que temos que avaliar (5.18) em diversos pontos $\{x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(k)}\}$ para cada elemento do campo latente \mathbf{x} , conforme descrito no item que fala sobre a aproximação de Laplace na seção 3.2.3. Desse modo, teríamos que fatorar $k \times n$ vezes uma matriz de dimensão $(n-1) \times (n-1)$. Para evitar esse custo computacional,

a abordagem do capítulo 3 foi realizar uma expansão de Taylor ao redor de $\mu_i(\boldsymbol{\theta})$ até o termo de primeira ordem. Isso significa que é necessário calcular a derivada da função $f(x_i)$ dada pela equação (5.19). No entanto, como salientado no apêndice A.1, o cálculo dessa derivada é possível de se obter analiticamente porque a matriz $\mathbf{C}(\mathbf{x})$ é diagonal, o que não é mais verdade ao se trabalhar com a **Estrutura 2**.

Porém, a derivada de $f(x_i)$ pode ser calculada de forma numérica

$$\delta_f^i = \frac{f(x_i + h) - f(x_i)}{h}. \quad (5.20)$$

Além disso, no ponto $x_i = \mu_i(\boldsymbol{\theta})$ o log do determinante de $(\mathbf{H} + \mathbf{C}(x_i))$ pode ser calculado com praticamente custo zero, do seguinte modo:

$$f(\mu_i(\boldsymbol{\theta})) = \frac{1}{2} \log |\mathbf{H} + \mathbf{C}(\mu_i(\boldsymbol{\theta}))| = \frac{1}{2} \log |\mathbf{Q}^*| + \frac{1}{2} \log \sigma_i^2(\boldsymbol{\theta}) \quad (5.21)$$

onde \mathbf{Q}^* é a matriz de precisão da aproximação Gaussiana $\pi_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ e $\sigma_i^2(\boldsymbol{\theta})$ é a variância de $\pi_G(x_i|\boldsymbol{\theta}, \mathbf{y})$, ambas já calculadas. A explicação da equação (5.21) consta no apêndice A.3.

Desse modo, uma aproximação linear de (5.19) exige que $f(x_i)$ seja avaliada somente no ponto $x_i + h$, digamos $h = 0,0001$, para cada elemento do campo latente, o que significa que essa operação irá custar a fatoração de uma matriz de dimensão $(n - 1) \times (n - 1)$ para cada elemento do campo latente, o que totaliza n fatorações em vez das $k \times n$ necessárias para a aproximação de Laplace (5.18). A equação (5.19) seria constante caso os dados fossem normais, de modo que o primeiro termo da expansão de Taylor já é uma correção para a aproximação Gaussiana.

Com relação a expansão de Taylor até o termo de terceira ordem do numerador de (5.18), temos que, conforme descrito no Apêndice A.1, o cálculo do termo de terceira ordem da expansão só foi possível de se obter porque a matriz $\mathbf{C}(\mathbf{x})$ era diagonal. Como o terceiro termo é o que corrige o erro de assimetria da aproximação Gaussiana $\pi_G(x_i|\boldsymbol{\theta}, \mathbf{y})$, uma vez não calculado, a expansão de Taylor deixa de ter utilidade. Portanto, ao se trabalhar com a **Estrutura 2**, não será feito uma expansão do numerador de (5.18). Sendo

assim, a equação (5.18), com o denominador substituído pela aproximação linear descrita acima, deverá ser avaliada em diversos pontos, do mesmo modo como feito para a aproximação de Laplace da seção 3.2.3. Para representar essa aproximação de Laplace simplificada vamos usar

$$\tilde{\pi}_{SLA}(x_i|\boldsymbol{\theta}, \mathbf{y}) \propto N\{x_i; \mu_i(\boldsymbol{\theta}), \sigma_i^2(\boldsymbol{\theta})\} \times \exp\{h(x_i)\} \quad (5.22)$$

A função $h(x)$ é uma função spline cúbica (Ahlberg *et al.*, 1967) ajustada à diferença $\log\{\tilde{\pi}_{SLA}(x_i|\boldsymbol{\theta}, \mathbf{y})\} - \log\{\tilde{\pi}_G(x_i|\boldsymbol{\theta}, \mathbf{y})\}$ avaliada nos pontos selecionados, e então a densidade (5.22) é normalizada utilizando integração por quadratura.

5.3 Modelo de Volatilidade Estocástica Assimétrica - Resultados

Para demonstrar a metodologia descrita na seção 5.2, foi simulado 30 observações do modelo de volatilidade estocástica assimétrica apresentado na seção 5.1.1, representado pelas equações [(5.6)-(5.7)-(5.9)]. Para a simulação, foram utilizados os seguintes valores para os parâmetros do modelo:

$$\rho = -0,5; \quad \phi = 0,8; \quad \mu = 1; \quad \sigma^2 = 1 \quad e \quad \eta_1 = 0$$

e para os parâmetros das distribuições a priori (5.8):

$$\alpha_\tau = 1; \quad \beta_\tau = 0.01; \quad a_1 = 0; \quad R_1 = 10; \quad R_2 = 100$$

sendo que para o parâmetro ρ foi utilizado uma $U[0, 1]$ como priori.

As distribuições marginais a posteriori de σ^2 , ϕ , ρ e μ se encontram na figura 5.1 e percebe-se que os resultados fornecidos tanto pelo INLA quanto pelo MCMC parecem estar de acordo. No gráfico da distribuição marginal a posteriori de μ (gráfico da parte de baixo à direita da figura 5.1) se encontra a aproximação fornecida pelo MCMC assim como as aproximações Gaussiana e de Laplace simplificado fornecida pelo INLA, já que o μ foi considerado no

campo latente para diminuir o tempo computacional do INLA (ver exemplo do modelo de volatilidade estocástica do capítulo 3). Neste gráfico nota-se que tanto a aproximação Gaussiana quanto a aproximação de Laplace simplificado estão bem próximas da aproximação fornecida pelo MCMC.

Na figura 5.2 estão a média e o intervalo de credibilidade de 95% de cada uma das distribuições marginais a posteriori de $(\eta_1, \eta_2, \dots, \eta_{29}, \eta_{30})$. Mais uma vez os resultados são satisfatórios quando comparados com o fornecido pelo MCMC. Nota-se que a aproximação de Laplace simplificado corrigiu levemente os erros da aproximação Gaussiana em alguns pontos e que ainda há pequenas diferenças entre a resultado obtido pela aproximação de Laplace simplificado e o obtido pelo MCMC. No entanto, fica difícil diagnosticar se essa diferença é devido ou não a erros de Monte Carlo, já que a convergência do modelo de volatilidade estocástica assimétrica ocorreu (foi assumido convergência através de análise gráfica das cadeias, omitidas neste texto) de forma extremamente lenta.

O motivo de se utilizar uma pequena quantidades de observações neste exercício, assim como no exemplo do capítulo 3, foi de facilitar a convergência das cadeias ao diminuir a correlação a posteriori entre \mathbf{x} e $\boldsymbol{\theta}$. Os resultados apresentados nesta seção levaram um pouco mais de 5 minutos utilizando a extensão do INLA e um pouco mais de 10 horas utilizando WinBUGS. O tempo necessário para obter resultados para o modelo de volatilidade estocástica assimétrica utilizando MCMC pode ser reduzido ao elaborar programação específica para este modelo em alguma outra linguagem mais eficiente. No entanto, esta tarefa está longe de ser trivial e irá requerer um tempo considerável por parte do programador. Enquanto isso, as extensões propostas neste capítulo podem ser programadas de forma eficiente uma única vez e, a partir daí, serem utilizadas para a estimação desse e outros modelos diferentes a um custo muito baixo para o usuário.

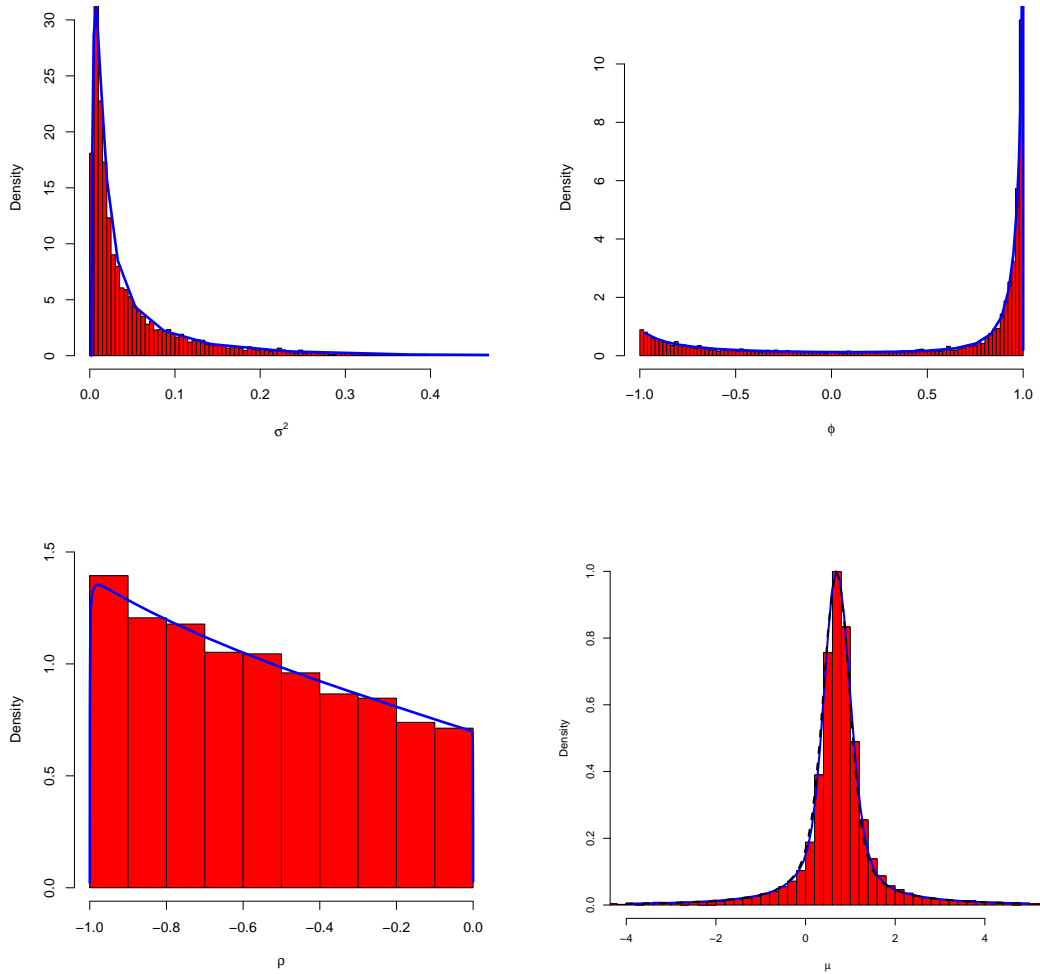


Figura 5.1: (Acima à Esquerda) $p(\sigma^2 | \mathbf{y})$; (Acima à Direita) $p(\phi | \mathbf{y})$; (Abaixo à Esquerda) $p(\rho | \mathbf{y})$; (Abaixo à Direita) $p(\mu | \mathbf{y})$: ■ MCMC (Histograma) — INLA

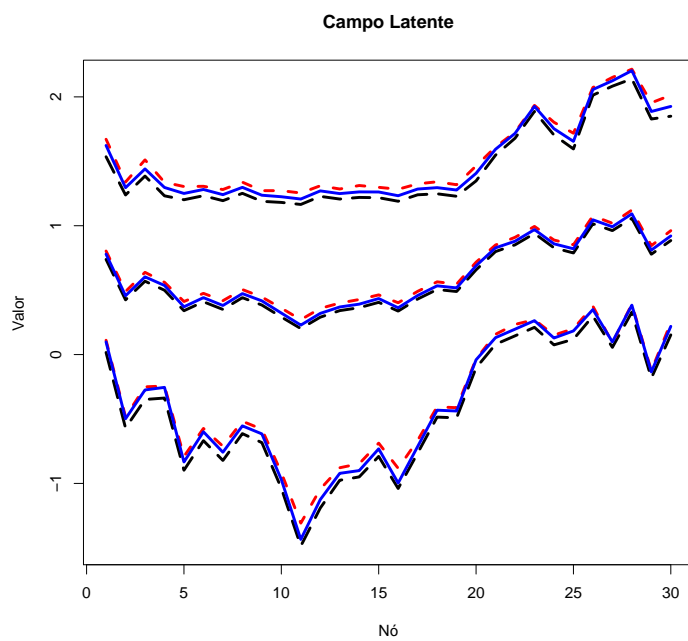


Figura 5.2: Sumário da distribuição a posteriori marginal de $\eta_1, \eta_2, \dots, \eta_{29}$ e η_{30} : média e IC 95% — MCMC — Aproximação Gaussiana — Laplace simplificado

Capítulo 6

Discussão Final

Essa dissertação começa abordando aproximações numéricas e analíticas de distribuições a posteriori marginais em um contexto geral, onde o interesse está em obter a posteriori de cada componente do vetor paramétrico $\Psi = (\psi_1, \dots, \psi_k)$. O capítulo 2 faz uma revisão da literatura, discutindo alguns dos mais importantes métodos utilizados nesse contexto e demonstra que há modelos onde o uso desses métodos é preferível aos métodos de simulação estocástica, tanto com relação a tempo quanto precisão dos resultados. No entanto, é salientado que tais métodos sofrem do problema de dimensionalidade, pois o custo computacional cresce drasticamente à medida que a dimensão k do vetor paramétrico aumenta.

O método INLA é apresentado e discutido no capítulo 3, onde é aplicado ao modelo dinâmico de primeira ordem e ao modelo de volatilidade estocástica e tem seus resultados comparados com os obtidos pelo MCMC. Através desses dois exemplos foi possível ilustrar detalhes que podem, à primeira vista, passar despercebidos para os interessados no método, como a possibilidade de incluir alguns parâmetros "fixos" no campo latente a fim de obter uma redução no tempo computacional do método. O capítulo 3 é finalizado com uma discussão sobre os motivos do INLA apresentar tamanha eficiência quando comparado com os métodos mais gerais apresentados no capítulo 2. A conclusão dessa análise foi que o mérito do INLA está em

usufruir ao máximo das vantagens oferecidas pela estrutura Markoviana do campo latente e pela forma "bem comportada" da condicional completa, característica herdada da priori Gaussiana atribuída a $\mathbf{x}|\boldsymbol{\theta}$. Essa conclusão mostra de forma inegável o peso que a priori desempenha no resultado final da análise, o que é muitas vezes negligenciado pelos que pensam estar atribuindo uma priori vaga ao parâmetro ao simplesmente escolher uma priori Normal com variância grande. Tal opção sem dúvida permite que a verossimilhança tenha um peso muito grande no resultado final, mas a forma da posteriori terá sido influenciada de maneira não desprezível pela priori Gaussiana, o que garante o sucesso do INLA nos modelos latentes Gaussianos.

No capítulo 4 foi proposta a utilização do INLA para inferência Bayesiana aproximada em modelos dinâmicos Bayesianos para processos pontuais espaço-temporais em vez da abordagem que utiliza métodos de simulação estocástica, pois o último pode demorar semanas para fornecer resultados que o INLA fornece em horas, sem contar a dificuldade de diagnosticar convergência quando se usa métodos iterativos.

Uma extensão do método INLA é proposta no capítulo 5, flexibilizando as hipóteses necessárias para sua aplicação, o que permite que uma gama ainda mais ampla de modelos possam ser estimados utilizando a metodologia apresentada neste capítulo. O modelo de volatilidade estocástica assimétrica foi escolhido para ilustrar a nova metodologia.

Como dito anteriormente, o INLA é um método poderoso, pois sabe usar as vantagens oferecidas pelo CAMG. No entanto, ele compartilha um ponto fraco com os métodos apresentados no capítulo 2. O custo computacional do INLA também cresce drasticamente à medida que o tamanho m do vetor $\boldsymbol{\theta}$ de hiperparâmetros aumenta. Isso ocorre porque a posteriori desse vetor paramétrico não é necessariamente tão bem "comportada" quanto a condicional completa do campo latente, o que não permite o uso das modificações apresentadas na seção 3.2.3 para tornar as aproximações mais rápidas de serem obtidas. Rue *et al.* (2009) sugerem o uso de *central composite design* (Box & Wilson, 1951) para amenizar o problema quando o tamanho de $\boldsymbol{\theta}$

é muito alto, porém o custo dessa abordagem é a redução na precisão das estimativas. Desse modo, o desenvolvimento de métodos para a obtenção de posterioris marginais em um contexto geral, como os apresentados no capítulo 2, volta a ganhar importância, sendo necessário ter métodos capazes de fornecer resultados satisfatórios, ao mesmo tempo que o custo para obter tais resultados não cresça exponencialmente no tempo. Essa parece, sem dúvida, ser uma importante área para trabalhos futuros.

Apêndice A

Apêndice A.1: Obtendo a aproximação de Laplace simplificada

A aproximação de Laplace, conforme definida na seção 3.2.3 é dada por:

$$\tilde{\pi}_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}_{-i}=E_{\pi_G}(\mathbf{x}_{-i}|x_i)} \quad (6.1)$$

onde $E_{\pi_G}(\mathbf{x}_{-i}|x_i)$ foi usada no lugar da moda $\mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta})$. Na seção 3.12 foi apresentada a aproximação de Laplace simplificada, obtida ao se efetuar a expansão de Taylor ao redor de $x_i = \mu_i(\boldsymbol{\theta})$ até os termos de terceira ordem no numerador e até o termo de primeira ordem no denominador. A seguir vamos apresentar essas operações com mais detalhes: Seja $x_i^s = (x_i - \mu_i(\boldsymbol{\theta}))/\sigma_i(\boldsymbol{\theta})$, onde $\mu_i(\boldsymbol{\theta})$ e $\sigma_i(\boldsymbol{\theta})$ são a média e a variância marginal de x_i obtida através da aproximação Gaussiana $\pi_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ e defina

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

O seguinte Lema será útil.

Lema 6.1. *Seja $\mathbf{x} = (x_1, \dots, x_n)^T \sim N(0, \boldsymbol{\Sigma})$, então para todo x_1*

$$-\frac{1}{2}(x_1, E(\mathbf{x}_{-1}|x_1)^T)\boldsymbol{\Sigma}^{-1} \begin{pmatrix} x_1 \\ E(\mathbf{x}_{-1}|x_1) \end{pmatrix} = -\frac{1}{2}x_1^2/\Sigma_{11}.$$

Expandindo o numerador ao redor de $x_i = \mu_i(\boldsymbol{\theta})$ até os termos de terceira ordem resulta em

$$\begin{aligned} \log \pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y}) \Big|_{\mathbf{x}_{-i}=E_{\pi_G}(\mathbf{x}_{-i}|x_i)} &= -\frac{1}{2}(x_i^{(s)})^2 \\ &+ \frac{1}{6}(x_i^{(s)})^3 \sum d_j^{(3)}(\mu_i(\boldsymbol{\theta}), \boldsymbol{\theta}) \{\sigma_j(\boldsymbol{\theta}) a_{ij}(\boldsymbol{\theta})\}^3 + \dots \end{aligned} \quad (6.2)$$

onde o primeiro e segundo termos seriam a aproximação Gaussiana enquanto que o terceiro termo corrige a assimetria. O terceiro termo só tem essa forma porque a função de verossimilhança no ponto y_i só depende de um elemento do campo latente. Caso isso não seja verdade a expressão da equação (6.2) se torna bem mais complicada.

O denominador de (6.1) se reduz a

$$\begin{aligned} \log \tilde{\pi}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y}) \Big|_{\mathbf{x}_{-i}=E_{\pi_G}(\mathbf{x}_{-i}|x_i)} &\stackrel{c}{\propto} \frac{1}{2} \log |\mathbf{H} + \mathbf{C}_{-i}(\mathbf{x}_{-i}(x_i))| \\ &= f(x_i) \end{aligned} \quad (6.3)$$

onde \mathbf{H} é a matrix de precisão a priori do CAMG com a i -ésima coluna e linha deletada e $\mathbf{C}_{-i}(\mathbf{x}_{-i}(x_i))$, daqui em diante denotada por $\mathbf{C}(x_i)$ para não carregar a notação, é a matriz diagonal da aproximação Gaussiana contendo menos as segundas derivadas da log verossimilhança na diagonal, mas com a i -ésima linha e coluna excluídas e avaliada no ponto $\mathbf{x}_{-i}(x_i) = E_{\pi_G}(\mathbf{x}_{-i}|x_i)$. A equação (6.3) mostra explicitamente que o denominador de (6.1) é uma função de x_i e como queremos efetuar uma expansão de Taylor até o termo de primeira ordem, tempo que calcular a derivada dessa função.

Como para qualquer matriz \mathbf{M} nós temos que $\partial \log |\mathbf{M}| = \text{traco}(\mathbf{M}^{-1} \partial \mathbf{M})$, então

$$\begin{aligned} f^{(1)}(x_i) &= d \log |\mathbf{H} + \mathbf{C}(x_i)| / dx_i \\ &= \text{traco}\{[\mathbf{H} + \mathbf{C}(x_i)]^{-1} d[\mathbf{H} + \mathbf{C}(x_i)] / dx_i\} \end{aligned} \quad (6.4)$$

$$= \text{traco}\{[\mathbf{H} + \mathbf{C}(x_i)]^{-1} \text{diag}[\mathbf{d}^3(x_i, \boldsymbol{\theta})]\} \quad (6.5)$$

$$= \sum_j \text{Var}(x_j|x_i) d_j^3(x_i, \boldsymbol{\theta})$$

$$= \sum_j \sigma_j(\boldsymbol{\theta}) [1 - \text{Corr}_{\pi_G}^2(x_i, x_j|\boldsymbol{\theta})] d_j^3(x_i, \boldsymbol{\theta})$$

Repare que a transição da equação (6.4) para a (6.5) só foi possível porque a matrix $\mathbf{C}(x_i)$ é diagonal, o que simplifica o cálculo de $d[\mathbf{H} + \mathbf{C}(x_i)]/dx_i$. Desse modo a expansão de Taylor do denominador é

$$\begin{aligned} \log \tilde{\pi}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y}) \Big|_{\mathbf{x}_{-i}=E\pi_G(\mathbf{x}_{-i}|x_i)} &\stackrel{c}{\propto} \\ -\frac{1}{2}x_i^{(s)} \sum_{j \in I/i} \sigma_j(\boldsymbol{\theta}) [1 - Corr_{\tilde{\pi}_G}^2(x_i, x_j|\boldsymbol{\theta})] d_j^3(\mu_i(\boldsymbol{\theta}), \boldsymbol{\theta}) \times & \quad (6.6) \\ \times \sigma_j(\boldsymbol{\theta}) a_{ij}(\boldsymbol{\theta}) + \dots & \end{aligned}$$

Como a equação (6.3) seria uma constante caso os dados fossem Normais, o termo de primeira ordem da expansão de Taylor acima já é uma correção para observações não-Gaussianas.

A partir das equações (6.1), (6.2) e (6.6), temos que

$$\log \tilde{\pi}_{SLA}(x_i^s|\boldsymbol{\theta}, \mathbf{y}) \stackrel{c}{\propto} \gamma_i^{(1)}(\boldsymbol{\theta})x_i^s - \frac{1}{2}(x_i^s)^2 + \frac{1}{6}(x_i^s)^3 \gamma_i^{(3)}(\boldsymbol{\theta}) + \dots \quad (6.7)$$

onde

$$\begin{aligned} \gamma_i^{(1)}(\boldsymbol{\theta}) &= \frac{1}{2} \sum_{j \in I/i} \sigma_j^2(\boldsymbol{\theta}) \{1 - corr_{\tilde{\pi}_G}(x_i, x_j)^2\} d_j^{(3)}\{\mu_i(\boldsymbol{\theta}), \boldsymbol{\theta}\} \sigma_j(\boldsymbol{\theta}) a_{ij}(\boldsymbol{\theta}) \\ \gamma_i^{(3)}(\boldsymbol{\theta}) &= \sum_{j \in I/i} d_j^{(3)}\{\mu_i(\boldsymbol{\theta}), \boldsymbol{\theta}\} \{\sigma_j(\boldsymbol{\theta}) a_{ij}(\boldsymbol{\theta})\}^3. \end{aligned}$$

como definido na seção 3.12.

Apêndice A.2: Ajustando uma densidade Normal assimétrica

A seguir será explicado como ajustar uma distribuição Normal assimétrica (Azzalini & Capitanio, 1999) a uma expansão da forma

$$\log \pi(x) \stackrel{c}{\propto} -\frac{1}{2}x^2 + \gamma^{(1)}x + \frac{1}{6}\gamma^{(3)} + \dots \quad (6.8)$$

A distribuição Normal assimétrica é dada por:

$$\pi(z) = \frac{2}{w} \phi\left(\frac{z - \xi}{w}\right) \Phi\left(a \frac{z - \xi}{w}\right) \quad (6.9)$$

onde $\phi(\cdot)$ e $\Phi(\cdot)$ são a densidade e a função de distribuição da Normal padrão, e $\xi, w > 0$ e a são os parâmetros de locação, escala e assimetria, respectivamente.

Até segunda ordem, (6.8) é Gaussiana com média $\gamma^{(1)}$ e variância 1. A média e a variância da distribuição Normal Assimétrica são $\xi + w\delta\sqrt{2/\pi}$ e $w^2(1 - 2\delta^2/\pi)$, respectivamente, onde $\delta = a/\sqrt{1+a^2}$. Mantêm-se essas quantidades fixas em $\gamma^{(1)}$ e 1, respectivamente, e ajusta-se a de modo que a terceira derivada na moda de (6.9) seja $\gamma^{(3)}$. Esse procedimento fornece três equações para determinar (ξ, w, a) . A configuração modal de (6.9) não é disponível analiticamente, mas uma expansão em série do log da densidade Normal assimétrica ao redor de $x = \xi$ fornece

$$x^* = \left(\frac{a}{w}\right) \frac{\sqrt{2\pi} + 2\xi\left(\frac{a}{w}\right)}{\pi + 2\left(\frac{a}{w}\right)^2} + \dots$$

Agora, computa-se a terceira derivada do log da densidade da Normal assimétrica em x^* . Para obter um ajuste de forma analítica (e rápido de computar), expande-se a derivada até os termos de terceira ordem com respeito a a/w :

$$\frac{\sqrt{2}(4-\pi)}{\pi^{3/2}} \left(\frac{a}{w}\right)^3 + \dots \quad (6.10)$$

e se faz com que (6.10) seja igual a $\gamma^{(3)}$. Esse procedimento fornece uma forma explícita para a obtenção dos 3 parâmetros da Normal assimétrica.

Apêndice A.3: Determinante de $\mathbf{Q}_{[-i,-i]}$

Para qualquer CAMG \mathbf{x} , com matriz de precisão \mathbf{Q} , tem-se que

$$\pi(\mathbf{x}) \propto |\mathbf{Q}|^{1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} \right\}. \quad (6.11)$$

Através das propriedades básicas da distribuição Gaussiana, temos que, para qualquer índice $i = 1, \dots, n$ a matriz de precisão de $\mathbf{x}_{-i}|x_i$ é $\mathbf{Q}_{[-i,-i]}$. Além

disso, temos que

$$\pi(\mathbf{x}) = \pi(x_i)\pi(\mathbf{x}_{-i}|x_i) \propto \text{Var}(x_i)^{-1/2} |\mathbf{Q}_{[-i,-i]}|^{1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} \right\} \quad (6.12)$$

Comparando (6.11) e (6.12) temos que

$$\frac{1}{2} \log |\mathbf{Q}_{[-i,-i]}| = \frac{1}{2} \log |\mathbf{Q}| + \frac{1}{2} \log \text{Var}(x_i)$$

Apêndice B

Apêndice B.1: Modelos Dinâmicos Bayesianos para Processos Pontuais Espaço-Temporais

Será mostrado a seguir que, para os modelos 2A, 3A, 1B e 1C do capítulo 4, a distribuição a priori de $\mathbf{x}|\boldsymbol{\theta}$ é um CAMG com matriz de precisão $Q(\boldsymbol{\theta})$ esparsa e que a verossimilhança na localização i e no tempo t só depende do campo latente \mathbf{x} através de um elemento. Com isso fica evidenciado que o INLA pode ser aplicado a cada um desses modelos de forma eficiente.

Para todos os modelos a seguir a verossimilhança é dada por:

$$l(\boldsymbol{\lambda}) \propto \prod_{i=1}^N \prod_{t=1}^T e^{-c_i \lambda_{i,t}} \lambda_{i,t}^{y_{i,t}}, \text{ onde}$$
$$\log(\lambda_{i,t}) = \varphi_{i,t} = \mu_t + \phi_{i,t}$$

sendo que o que vai variar serão as especificações dos componentes μ_t e $\phi_{i,t}$.

Modelo 2A

No modelo 2A, temos que μ_t é dado por

$$\mu_t = \beta_0 + \beta_1 t, \tag{6.13}$$

e $\phi_{[:,t]}$ segue a dinâmica

$$\phi_{[:,t]} = \eta\phi_{[:,t-1]} + \mathbf{w}_{[:,t]}, \quad \mathbf{w}_{[:,t]} \sim N(\mathbf{0}, (1 - \eta^2)\sigma^2 R_\kappa). \quad (6.14)$$

Neste caso, denotando $\varphi_{[:,t]} = (\varphi_{[1,t]}, \dots, \varphi_{[240,t]})'$, o campo latente é definido como $\mathbf{x} = (\varphi_{[:,1]}, \dots, \varphi_{[:,13]}, \beta_0, \beta_1)$ e $\boldsymbol{\theta} = (\eta, \sigma^2, \kappa)$. Será assumido a priori que os hiperparâmetros $\boldsymbol{\theta}$ são independentes e a distribuição a priori para η será uma uniforme em $[0, 1]$, e vamos atribuir uma $GI(1, 1)$ para σ^2 e para κ . Após algumas contas, percebe-se que

$$\varphi_{[:,1]} \sim N((\beta_0 + \beta_1)\mathbf{1}, \sigma^2 R_\kappa) \quad (6.15)$$

$$\varphi_{[:,t]} = \eta\varphi_{[:,t-1]} + [(1 - \eta)\beta_0 + (t - \eta(t - 1))\beta_1]\mathbf{1} + \varepsilon_{[:,t]}, \quad (6.16)$$

onde $\varepsilon_{[:,t]} \sim N(0, (1 - \eta^2)\sigma^2 R_\kappa)$.

Utilizando o fato que $\beta_0, \beta_1 \sim N(0, 10)$ e as equações (6.15) e (6.16), temos que a distribuição a priori $\pi(\mathbf{x}|\boldsymbol{\theta})$ é dada por:

$$\begin{aligned} \pi(\mathbf{x}|\boldsymbol{\theta}) &= \pi(\mu)\pi(\varphi_{[:,1]}|\mu) \prod_{t=2}^{13} \pi(\varphi_{[:,t]}|\mu, \varphi_{[:,t-1]}) \\ &\propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta}) \mathbf{x} \right\}, \text{ onde} \\ \mathbf{Q}(\boldsymbol{\theta}) &= \begin{pmatrix} \mathbf{Q}_{1,1} & \mathbf{Q}_{1,2} & 0 & \dots & 0 & \mathbf{Q}_{1,\beta_0} & \mathbf{Q}_{1,\beta_1} \\ \mathbf{Q}_{2,1} & \mathbf{Q}_{2,2} & \mathbf{Q}_{2,3} & \dots & 0 & \mathbf{Q}_{2,\beta_0} & \mathbf{Q}_{2,\beta_1} \\ 0 & \mathbf{Q}_{3,2} & \mathbf{Q}_{3,3} & \dots & 0 & \mathbf{Q}_{3,\beta_0} & \mathbf{Q}_{3,\beta_1} \\ 0 & 0 & \mathbf{Q}_{4,3} & \dots & 0 & \mathbf{Q}_{4,\beta_0} & \mathbf{Q}_{4,\beta_1} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \mathbf{Q}_{T-1,T} & \mathbf{Q}_{T-1,\beta_0} & \mathbf{Q}_{T-1,\beta_1} \\ 0 & 0 & 0 & \dots & \mathbf{Q}_{T,T} & \mathbf{Q}_{T,\beta_0} & \mathbf{Q}_{T,\beta_1} \\ \mathbf{Q}_{\beta_0,1} & \mathbf{Q}_{\beta_0,2} & \mathbf{Q}_{\beta_0,3} & \dots & \mathbf{Q}_{\beta_0,T} & \mathbf{Q}_{\beta_0,\beta_0} & \mathbf{Q}_{\beta_0,\beta_1} \\ \mathbf{Q}_{\beta_1,1} & \mathbf{Q}_{\beta_1,2} & \mathbf{Q}_{\beta_1,3} & \dots & \mathbf{Q}_{\beta_1,T} & \mathbf{Q}_{\beta_1,\beta_0} & \mathbf{Q}_{\beta_1,\beta_1} \end{pmatrix} \end{aligned}$$

onde,

$$\mathbf{Q}_{1,1} = [1 + \eta^2(1 - \eta^2)^{-1}]\sigma^{-2}\mathbf{Q}_\kappa;$$

$$\mathbf{Q}_{i,i} = (1 + \eta^2)(1 - \eta^2)^{-1}\sigma^{-2}\mathbf{Q}_\kappa, \quad i = 2, 3, \dots, T - 1;$$

$$\begin{aligned}
\mathbf{Q}_{T,T} &= (1 - \eta^2)^{-1} \sigma^{-2} \mathbf{Q}_\kappa; \\
\mathbf{Q}_{\beta_0, \beta_0} &= R_1^{-1} + [1 + (T - 1)(1 - \eta)^2(1 - \eta^2)^{-1}] \sigma^{-2} \mathbf{1}' \mathbf{Q}_\kappa \mathbf{1}; \\
\mathbf{Q}_{\beta_1, \beta_1} &= R_2^{-1} + [1 + (1 - \eta^2)^{-1} \sum_{t=2}^T (t - (t - 1)\eta)^2] \sigma^{-2} \mathbf{1}^T \mathbf{Q}_\kappa \mathbf{1}; \\
\mathbf{Q}_{i, i+1} &= \mathbf{Q}_{i+1, i} = -\eta(1 - \eta^2)^{-1} \sigma^{-2} \mathbf{Q}_\kappa, \quad i = 1, 2, \dots, T - 1; \\
\mathbf{Q}_{1, \beta_0} &= [\eta(1 - \eta)(1 - \eta^2) - 1] \sigma^{-2} \mathbf{Q}_\kappa \mathbf{1} \quad e \quad \mathbf{Q}_{\beta_0, 1} = \mathbf{Q}_{1, \beta_0}^T; \\
\mathbf{Q}_{i, \beta_0} &= (1 - \eta)(\eta - 1)(1 - \eta^2)^{-1} \sigma^{-2} \mathbf{Q}_\kappa \mathbf{1}, \quad i = 2, 3, \dots, T - 1; \\
\mathbf{Q}_{\beta_0, i} &= \mathbf{Q}_{i, \beta_0}^T, \quad i = 2, 3, \dots, T - 1; \\
\mathbf{Q}_{T, \beta_0} &= -(1 - \eta)(1 - \eta^2)^{-1} \sigma^{-2} \mathbf{Q}_\kappa \mathbf{1} \quad e \quad \mathbf{Q}_{\beta_0, T} = \mathbf{Q}_{T, \beta_0}^T; \\
\mathbf{Q}_{1, \beta_1} &= [\eta(2 - \eta)(1 - \eta^2)^{-1} - 1] \sigma^{-2} \mathbf{Q}_\kappa \mathbf{1} \quad e \quad \mathbf{Q}_{\beta_1, 1} = \mathbf{Q}_{1, \beta_1}^T; \\
\mathbf{Q}_{i, \beta_1} &= [\eta((i + 1) - i\eta) - (i - (i - 1)\eta)](1 - \eta^2)^{-1} \sigma^{-2} \mathbf{1}, \quad i = 2, \dots, T - 1; \\
\mathbf{Q}_{\beta_1, i} &= \mathbf{Q}_{i, \beta_1}^T, \quad i = 2, \dots, T - 1; \\
\mathbf{Q}_{T, \beta_1} &= -(T - (T - 1)\eta)(1 - \eta^2)^{-1} \sigma^{-2} \mathbf{Q}_\kappa \mathbf{1} \quad e \quad \mathbf{Q}_{\beta_1, T} = \mathbf{Q}_{T, \beta_1}^T; \\
\mathbf{Q}_{\beta_0, \beta_1} &= \mathbf{Q}_{\beta_1, \beta_0} = [1 + (1 - \eta)(1 - \eta^2)^{-1} \sum_{t=2}^T (t - (t - 1)\eta)] \sigma^{-2} \mathbf{1}^T \mathbf{Q}_\kappa \mathbf{1},
\end{aligned}$$

sendo \mathbf{Q}_κ o inverso de \mathbf{R}_κ

Modelo 3A

No modelo 3A, temos que μ_t é dado por

$$\mu_t = \mu_{t-1} + v_t, \quad v_t \sim N(0, w^2). \quad (6.17)$$

e $\phi_{[:,t]}$ segue a dinâmica

$$\phi_{[:,t]} = \eta \phi_{[:,t-1]} + \mathbf{w}_{[:,t]}, \quad \mathbf{w}_{[:,t]} \sim N(\mathbf{0}, (1 - \eta^2) \sigma^2 R_\kappa). \quad (6.18)$$

Neste caso, denotando $\boldsymbol{\varphi}_{[:,t]} = (\varphi_{[1,t]}, \dots, \varphi_{[240,t]})'$, o campo latente é definido como $\mathbf{x} = (\boldsymbol{\varphi}_{[:,1]}, \dots, \boldsymbol{\varphi}_{[:,13]}, \mu_1, \mu_2, \dots, \mu_T)$ e $\boldsymbol{\theta} = (\eta, \sigma^2, \kappa, w^2)$. Será assumido a priori que os hiperparâmetros $\boldsymbol{\theta}$ são independentes e a distribuição a priori para η será uma uniforme em $[0, 1]$, e vamos atribuir uma $GI(1, 1)$ para σ^2 , w^2 e κ . Após algumas contas, percebe-se que

$$\boldsymbol{\varphi}_{[:,1]} \sim N(1\mu_1, \sigma^2 R_\kappa) \quad (6.19)$$

$$\boldsymbol{\varphi}_{[:,t]} = \eta \boldsymbol{\varphi}_{[:,t-1]} + \mathbf{1}[\mu_t - \eta \mu_{t-1}] + \boldsymbol{\varepsilon}_{[:,t]}, \quad (6.20)$$

$$\text{onde } \boldsymbol{\varepsilon}_{[:,t]} \sim N(0, (1 - \eta^2) \sigma^2 R_\kappa).$$

Utilizando o fato que $\mu_1 \sim N(0, 10)$ e as equações (6.19) e (6.20), temos que a distribuição a priori $\pi(\mathbf{x}|\boldsymbol{\theta})$ é dada por:

$$\begin{aligned} \pi(\mathbf{x}|\boldsymbol{\theta}) &= \pi(\mu_1) \prod_{t=2}^{13} \pi(\mu_t|\mu_{t-1}) \pi(\varphi_{[.,1]}|\mu_1) \prod_{t=2}^{13} \pi(\varphi_{[.,t]}|\mu_{t-1}, \mu_t, \varphi_{[.,t-1]}) \\ &\propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta}) \mathbf{x} \right\}, \text{ onde} \end{aligned}$$

$$\mathbf{Q}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{Q}_{\varphi_1, \varphi_1} & \mathbf{Q}_{\varphi_1, \varphi_2} & \cdots & 0 & \mathbf{Q}_{\varphi_1, \mu_1} & \mathbf{Q}_{\varphi_1, \mu_2} & \cdots & 0 \\ \mathbf{Q}_{\varphi_2, \varphi_1} & \mathbf{Q}_{\varphi_2, \varphi_2} & \cdots & 0 & \mathbf{Q}_{\varphi_2, \mu_1} & \mathbf{Q}_{\varphi_2, \mu_2} & \cdots & 0 \\ 0 & \mathbf{Q}_{\varphi_3, \varphi_2} & \cdots & 0 & 0 & \mathbf{Q}_{\varphi_3, \mu_2} & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{Q}_{\varphi_{T-1}, \varphi_T} & 0 & 0 & \cdots & \mathbf{Q}_{\varphi_{T-1}, \mu_T} \\ 0 & 0 & \cdots & \mathbf{Q}_{\varphi_T, \varphi_T} & 0 & 0 & \cdots & \mathbf{Q}_{\varphi_T, \mu_T} \\ \hline \mathbf{Q}_{\mu_1, \varphi_1} & \mathbf{Q}_{\mu_1, \varphi_2} & \cdots & 0 & \mathbf{Q}_{\mu_1, \mu_1} & \mathbf{Q}_{\mu_1, \mu_2} & \cdots & 0 \\ \mathbf{Q}_{\mu_2, \varphi_1} & \mathbf{Q}_{\mu_2, \varphi_2} & \cdots & 0 & \mathbf{Q}_{\mu_2, \mu_1} & \mathbf{Q}_{\mu_2, \mu_2} & \cdots & 0 \\ 0 & \mathbf{Q}_{\mu_3, \varphi_2} & \cdots & 0 & 0 & \mathbf{Q}_{\mu_3, \mu_2} & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{Q}_{\mu_{T-1}, \varphi_T} & 0 & 0 & \cdots & \mathbf{Q}_{\mu_{T-1}, \mu_T} \\ 0 & 0 & \cdots & \mathbf{Q}_{\mu_T, \varphi_T} & 0 & 0 & \cdots & \mathbf{Q}_{\mu_T, \mu_T} \end{pmatrix}$$

onde,

$$\begin{aligned} \mathbf{Q}_{\varphi_1, \varphi_1} &= [1 + \eta^2(1 - \eta^2)^{-1}] \sigma^{-2} \mathbf{Q}_\kappa; \\ \mathbf{Q}_{\varphi_i, \varphi_i} &= (1 + \eta^2)(1 - \eta^2)^{-1} \sigma^{-2} \mathbf{Q}_\kappa, \quad i = 2, 3, \dots, T-1; \\ \mathbf{Q}_{\varphi_T, \varphi_T} &= (1 - \eta^2)^{-1} \sigma^{-2} \mathbf{Q}_\kappa; \\ \mathbf{Q}_{\varphi_i, \varphi_{i+1}} &= \mathbf{Q}_{\varphi_{i+1}, \varphi_i} = -\eta(1 - \eta^2)^{-1} \sigma^{-2} \mathbf{Q}_\kappa, \quad i = 1, 2, \dots, T-1; \\ \mathbf{Q}_{\mu_1, \mu_1} &= \tau_0^{-2} + w^{-2} + \sigma^{-2} \mathbf{1}^T \mathbf{Q}_\kappa \mathbf{1} [1 + \eta^2(1 - \eta^2)^{-1}]; \\ \mathbf{Q}_{\mu_i, \mu_i} &= 2w^{-2} + \sigma^{-2} \mathbf{1}^T \mathbf{Q}_\kappa \mathbf{1} (1 + \eta^2)(1 - \eta^2)^{-1}, \quad i = 2, 3, \dots, T-1; \\ \mathbf{Q}_{\mu_T, \mu_T} &= w^{-2} + \sigma^{-2} \mathbf{1}^T \mathbf{Q}_\kappa \mathbf{1} (1 - \eta^2)^{-1}; \\ \mathbf{Q}_{\mu_i, \mu_{i+1}} &= -w^{-2} - \sigma^{-2} \mathbf{1}^T \mathbf{Q}_\kappa \mathbf{1} \eta (1 - \eta^2)^{-1} = \mathbf{Q}_{\mu_{i+1}, \mu_i}, \quad i = 1, 2, \dots, T-1; \\ \mathbf{Q}_{\mu_1, \varphi_1} &= -\sigma^{-2} \mathbf{1}^T \mathbf{Q}_\kappa [1 + \eta^2(1 - \eta^2)^{-1}]; \\ \mathbf{Q}_{\mu_i, \varphi_i} &= -\sigma^{-2} \mathbf{1}^T \mathbf{Q}_\kappa (1 + \eta^2)(1 - \eta^2)^{-1}, \quad i = 2, 3, \dots, T-1; \end{aligned}$$

$$\begin{aligned}
\mathbf{Q}_{\mu_T, \varphi_T} &= -\sigma^{-2} \mathbf{1}^T \mathbf{Q}_\kappa (1 - \eta^2)^{-1}; \\
\mathbf{Q}_{\mu_i, \varphi_{i+1}} &= \sigma^{-2} \mathbf{1}^T \mathbf{Q}_\kappa \eta (1 - \eta^2)^{-1} = \mathbf{Q}_{\mu_{i+1}, \varphi_i}, \quad i = 1, 2, \dots, T - 1
\end{aligned}$$

sendo \mathbf{Q}_κ o inverso de \mathbf{R}_κ

Referências Bibliográficas

- Ahlberg, J.H., Nilson, E.N., & Walsh, J.L. 1967. *The theory of splines and their applications*. Academic Pr.
- Azzalini, A., & Capitanio, A. 1999. Statistical applications of the multivariate skew normal distribution. *Journal of the royal statistical society. series b (statistical methodology)*, **61**(3), 579–602.
- Bates, D., & Maechler, M. 2007. Matrix: A Matrix package for R. *R package version 0.99875-2*, url <http://cran.r-project.org>, **15**.
- Bates, D., Chambers, J., Dalgaard, P., Gentleman, R., Hornik, K., Iacus, S., Ihaka, R., Leisch, F., Lumley, T., Maeschler, M., *et al.* 2004. The R project for statistical computing. Url <http://www.r-project.org>.
- Black, F. 1976. Studies of stock market volatility changes. *Page 181 of: Proceedings of the american statistical association, business and economic statistics section*, vol. 177.
- Box, GEP, & Wilson, KB. 1951. On the experimental attainment of optimum conditions. *Journal of the royal statistical society. series b (methodological)*, **13**(1), 1–45.
- Boyd, S.P., & Vandenberghe, L. 2004. *Convex optimization*. Cambridge Univ Pr.
- Brix, A. 2001. Space-time multi type log Gaussian Cox processes with a view to modelling weeds. *Scandinavian journal of statistics*, **28**(3), 471–488.

- Brix, A., & Diggle, P.J. 2001. Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the royal statistical society. series b (statistical methodology)*, **63**(4), 823–841.
- Cheng, S.H., & Higham, N.J. 1998. A modified Cholesky algorithm based on a symmetric indefinite factorization. *Siam journal on matrix analysis and applications*, **19**(4), 1097–1112.
- Cheridito, P. 2003. Arbitrage in fractional Brownian motion models. *Finance and stochastics*, **7**(4), 533–553.
- Christie, A.A. 1982. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial economics*, **10**(4), 407–432.
- Cox, D.R., & Isham, V. 1980. *Point processes*. Chapman & Hall/CRC.
- De Bruijn, NG. 1981. *Asymptotic methods in analysis*. Dover Pubns.
- Diggle, P., Knorr-Held, L., Rowlingson, B., Su, T., Hawtin, P., & Bryant, T.N. 2003. On-line monitoring of public health surveillance data. *Monitoring the health of populations: Statistical principles and methods for public health surveillance*, 233–66.
- Duff, I.S., Erisman, A.M., & Reid, J.K. 1989. *Direct methods for sparse matrices*. Oxford University Press, USA.
- Durbin, J., & Koopman, SJ. 2000. 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 (statistical methodology)*, **62**(1), 3–56.
- Faes, C., Aerts, M., Geys, H., Bijmens, L., Ver Donck, L., & Lammers, W.J.E.P. 2006. GLMM approach to study the spatial and temporal evolution of spikes in the small intestine. *Statistical modelling*, **6**(4), 300.

- Gamerman, D. 1992. A dynamic approach to the statistical analysis of point processes. *Biometrika*, **79**(1), 39.
- Gamerman, D., & Lopes, H.F. 2006. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman & Hall/CRC.
- Gamerman, D., Salazar, & Reis, E. 2007. Dynamic Gaussian process priors, with applications to the analysis of space-time data (with discussion). *Bayesian statistics 8*, 149–174.
- Gelfand, A.E., & Smith, A.F.M. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the american statistical association*, **85**(410), 398–409.
- Gelfand, A.E., Banerjee, S., & Gamerman, D. 2005. Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*, **16**(5), 465–480.
- George, A., & Liu, JW. 1981. Computer solution of large sparse positive definite systems. *Prentice-hall, inc., englewood cliffs, nj 07632*.
- Gordon, N.J., Salmond, D.J., & Smith, A.F.M. 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Pages 107–113 of: Iee proceedings*, vol. 140.
- Harvey, A.C., & Shephard, N. 1996. Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of business & economic statistics*, **14**(4), 429–434.
- Higham, N.J. 2002. Computing the nearest correlation matrix—a problem from finance. *Ima journal of numerical analysis*, **22**(3), 329.
- Hills, SE, & Smith, AFM. 1992. Parameterization issues in Bayesian inference. *Bayesian statistics*, **4**, 227–246.
- Hills, S.E., & Smith, A.F.M. 1993. Diagnostic plots for improved parameterization in Bayesian inference. *Biometrika*, **80**(1), 61–74.

- Knol, D.L., & ten Berge, J.M.F. 1989. Least-squares approximation of an improper correlation matrix by a proper one. *Psychometrika*, **54**(1), 53–61.
- Lunn, D.J., Thomas, A., Best, N., & Spiegelhalter, D. 2000. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, **10**(4), 325–337.
- Martino, S., & Rue, H. 2008. *Implementing approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations: A manual for the inla-program*. Tech. rept. Citeseer.
- Meyer, R., & Yu, J. 2000. BUGS for a Bayesian analysis of stochastic volatility models. *The econometrics journal*, **3**(2), 198–215.
- Møller, J., Syversveen, A.R., & Waagepetersen, R.P. 1998. Log Gaussian Cox processes. *Scandinavian journal of statistics*, 451–482.
- Naylor, JC, & Smith, AFM. 1982. Applications of a method for the efficient computation of posterior distributions. *Journal of the royal statistical society. series c (applied statistics)*, **31**(3), 214–225.
- Paez, M., & Diggle, P. 2009. Cox processes in time for point patterns and their aggregations. *To appear in environmetrics*.
- Reilly, P.M. 1976. The numerical computation of posterior distributions in Bayesian statistical inference. *Journal of the royal statistical society. series c (applied statistics)*, **25**(3), 201–209.
- Reis, E. 2008. *Bayesian Dynamic Models for Space-Time Point Processes*. Unpublished Dr.Sc. thesis (in Portuguese), Universidade Federal do Rio de Janeiro, Brazil.
- Reis, E., Gamerman, D., Paez, M., & Guerrero, T. 2010. Bayesian Dynamic Models for Space-Time Point Processes. *Submitted for publication*.
- Robert, C.P., & Casella, G. 2004. *Monte Carlo statistical methods*. Springer Verlag.

- Rue, H. 2001. Fast sampling of Gaussian Markov random fields. *Journal of the royal statistical society. series b (statistical methodology)*, **63**(2), 325–338.
- Rue, H., & Follstad, T. 2002. GMRFlib: a C-library for fast and exact simulation of Gaussian Markov random fields. *Statistics report*, **1**.
- Rue, H., & Held, L. 2005. *Gaussian Markov random fields: theory and applications*. Chapman & Hall.
- Rue, H., & Martino, S. 2007. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of statistical planning and inference*, **137**(10), 3177–3192.
- Rue, H., Martino, S., & Chopin, N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (*with discussion*). *Journal of the royal statistical society: Series b (statistical methodology)*, **71**(2), 319–392.
- Salzer, H.E., Zucker, R., & Capuano, R. 1952. Table of the zeros and weight factors of the first twenty Hermite polynomials. *Journal of research of the national bureau of standards*, **48**(2), 111–116.
- Smith, AFM, Skene, AM, Shaw, JEH, Naylor, JC, & Dransfield, M. 1985. Implementation of the Bayesian paradigm. *Commun. stat. theory methods.*, **14**(5), 1079–1102.
- Smith, AFM, Skene, AM, Shaw, JEH, & Naylor, JC. 1987. Progress with numerical and graphical methods for practical Bayesian statistics. *The statistician*, 75–82.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & van der Linde, A. 2002. Bayesian measures of model complexity and fit. *Journal of the royal statistical society. series b (statistical methodology)*, **64**(4), 583–639.

- Taylor, S.J. 2006. Modeling stochastic volatility: A review and comparative study. *Mathematical finance*, **4**(2), 183–204.
- Tibshirani, R., & Wasserman, L. 1994. Some aspects of the reparametrization of statistical models. *The canadian journal of statistics/la revue canadienne de statistique*, **22**(1), 163–173.
- Tierney, L., & Kadane, J.B. 1986. Accurate approximations for posterior moments and marginal densities. *Journal of the american statistical association*, 82–86.
- Tierney, L., Kass, R.E., & Kadane, J.B. 1989. Approximate marginal densities of nonlinear functions. *Biometrika*, **76**(3), 425.
- Waagepetersen, R. 2004. Convergence of posteriors for discretized log Gaussian Cox processes. *Statistics and probability letters*, **66**(3), 229–235.
- West, M., & Harrison, J. 1997. *Bayesian forecasting and dynamic models*. Springer Verlag.