

# Revisitando MLD para família exponencial p-dimensional

Raíra Marotta Bastos Vieira



Universidade Federal do Rio de Janeiro  
Instituto de Matemática  
Departamento de Métodos Estatísticos

2018

# **Revisitando MLD para família exponencial p-dimensional**

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

**Raíra Marotta Bastos Vieira**

Orientadores:

Mariane B. Alves

Hélio dos Santos Migon

Rio de Janeiro, RJ - Brasil

2018

# Revisitando MLD para família exponencial p-dimensional

**Raíra Marotta Bastos Vieira**

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Estatística do 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:

---

Mariane Branco Alves

DME/IM - UFRJ - Orientadora.

---

Dani Gamerman

DME/IM - UFRJ.

---

Heudson Mirandola

IM - UFRJ

Rio de Janeiro, RJ - Brasil

2018

*Dei sub numine viget*

# Agradecimentos

À minha família, por todo amor, suporte e compreensão.

Aos meus amigos, por tornarem esta jornada mais leve e cheia de alegrias.

Ao Matheus Rebelo, por todo carinho, cuidado e apoio.

Aos meus orientadores Mariane e Migon, por serem fonte inesgotável de inspiração.

Além da paciência e suporte durante essa jornada.

Ao professor Heudson Mirandola, por acreditar no trabalho e ser extramente solícito, nos ajudando a desenvolvê-lo.

# Resumo

Os modelos lineares generalizados dinâmicos são uma extensão de modelos lineares dinâmicos (no sentido de considerar respostas não gaussianas) e para modelos lineares generalizados, que consideram respostas na família exponencial, mas presumem efeitos fixos ao longo do tempo. O presente trabalho revisita métodos de inferência para essa classe de modelos não só no que concerne à família exponencial uniparamétrica como  $k$ -paramétrica. Métodos como *Conjugate Updating*, *Conjugate Updating Estendido* e modelos não estruturados são apresentados. Neste trabalho apresenta-se uma extensão do método *Conjugate Updating Estendido* para o caso Multinomial. Além do mais, uma nova abordagem via geometria da informação para estimação de modelos lineares dinâmicos generalizados é proposta. Conceitos como Divergência de Kullback-Leibler e Teorema da Projeção são utilizados no desenvolvimento do método. Estudos simulados bem como aplicação a dados reais foram feitos e apresentaram resultados satisfatórios.

*Palavras-Chaves:* inferência bayesiana, modelos lineares dinâmicos generalizados, geometria da informação, conjugate updating.

# Abstract

Dynamic generalized linear models are an extension of dynamic linear models (in the sense of considering non-Gaussian responses) and for generalized linear models, which consider responses in the exponential family, but assume fixed effects over time. The present work revisits inference methods for this class of models, not only for the one-parameter exponential family but also for the k-parameter. Methods such as *Conjugate Updating*, *Extended Conjugate Updating*, and unstructured models are presented. In this work we extend the Extended Conjugate Updating method for the Multinomial case. Moreover, a new approach via information geometry is proposed for estimating dynamic generalized linear models. Concepts such as Kullback-Leibler Divergence and Projection Theorem are used in the development of the method. Simulation studies as well as application to real data were made and satisfactory results were obtained.

*Keywords:* bayesian inference, dynamic generalized linear models, information geometry, conjugate updating.

# Sumário

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introdução</b>   | <b>1</b>  |
| <b>2</b> | <b>Conceitos Básicos</b>                                  | <b>5</b>  |
| 2.1      | Família Exponencial . . . . .                             | 5         |
| 2.1.1    | Conjugação na Família Exponencial . . . . .               | 7         |
| 2.2      | Geometria da Informação . . . . .                         | 10        |
| 2.2.1    | Divergências . . . . .                                    | 10        |
| 2.2.1.1  | Divergência de Bregman . . . . .                          | 10        |
| 2.2.1.2  | Divergência de Kullback-Leibler . . . . .                 | 11        |
| 2.2.2    | Teoremas . . . . .  | 12        |
| 2.2.3    | Exemplos de Projeções . . . . .                           | 15        |
| 2.2.3.1  | Projetando a Distribuição Log-Normal na Gama . . . . .    | 15        |
| 2.2.3.2  | Projetando a Distribuição Logito-Normal na Beta . . . . . | 18        |
| <b>3</b> | <b>Modelos Dinâmicos Generalizados</b>                    | <b>20</b> |
| 3.1      | Modelos Lineares Dinâmicos . . . . .                      | 22        |
| 3.2      | Modelos Lineares Dinâmicos Generalizados . . . . .        | 25        |
| 3.2.1    | <i>Conjugate Updating</i> . . . . .                       | 26        |
| 3.2.2    | Modelos Dinâmicos de Nível Local . . . . .                | 30        |
| <b>4</b> | <b>MLDG via Geometria da Informação</b>                   | <b>35</b> |
| 4.1      | Introdução . . . . .                                      | 35        |
| 4.2      | Método proposto . . . . .                                 | 36        |



|          |  |           |
|----------|--|-----------|
| 4.2.1    | Modelo Linear Dinâmico Generalizado Poisson . . . . .  | 38        |
| 4.2.2    | Modelo Linear Dinâmico Generalizado Bernoulli/Binomial . . . . .                                       | 41        |
| 4.2.3    | Estudo de Simulação . . . . .  | 42        |
| 4.2.3.1  | Modelo Poisson . . . . .   | 43        |
| 4.3      | Compatibilizando diferentes modelos . . . . .  | 44        |
| 4.3.1    | Modelo Linear Dinâmico Generalizado Poisson . . . . .  | 46        |
| 4.3.2    | Modelo Linear Dinâmico Generalizado Bernoulli/Binomial . . . . .                                       | 49        |
| 4.3.3    | Aplicação . . . . .  | 50        |
| <b>5</b> | <b><i>Conjugate Updating</i> estendido</b>   | <b>58</b> |
| 5.1      | Distribuição Gama . . . . .  | 61        |
| 5.2      | Distribuição Normal . . . . .  | 63        |
| 5.3      | Distribuição Multinomial . . . . .   | 65        |
| 5.3.1    | Estudo Simulado . . . . .  | 72        |
| 5.4      | Aplicações a dados reais . . . . .   | 75        |
| <b>6</b> | <b>Conclusão</b>   | <b>77</b> |
| <b>A</b> | <b>Resultados secundários referentes à série de vendas</b>   | <b>79</b> |
| A.1      | Tabela de resultados do modelo Poisson via Geometria da Informação que compatibiliza modelos . . . . . | 79        |
| A.2      | Cadeias obtidas pelo MCMC . . . . .  | 80        |

# Lista de Tabelas

|     |   |    |
|-----|---|----|
| 4.1 | Resultados da estimação dos parâmetros do modelo de nível local . . . . .   | 54 |
| 4.2 | Tempo computacional gasto em cada um dos métodos . . . . .  | 54 |
| 4.3 | Comparação dos modelos via geometria da informação compatibilizando modelos e sem compatibilizar, <i>Conjugate Updating</i> e modelo de nível local | 57 |
| A.1 | Tabela com os resultados obtidos de acordo com os parâmetros $S$ e $I$ para os 10 melhores modelos. . . . .   | 79 |

# Lista de Figuras

|     |   |    |
|-----|---|----|
| 2.1 | Exemplo do Teorema da Projeção . . . . .  | 13 |
| 2.2 | Exemplo de Compatibilização de Variáveis . . . . .  | 14 |
| 2.3 | Distribuições Log-Normal e Gama aproximadas através da minimização da divergência de Kullback-Leibler quando utilizados diferentes parâmetros para as distribuições Log-Normal. . . . .   | 17 |
| 2.4 | Distribuições Logito-Normal e Beta aproximadas através da minimização da divergência de Kullback-Leibler quando utilizados diferentes parâmetros para as distribuições Logito-Normal. . . . .   | 19 |
| 3.1 | Ilustração da evolução do sistema . . . . .   | 23 |
| 4.1 | Boxplots resultantes das réplicas para a média a <i>posteriori</i> do parâmetro natural e seus valores verdadeiros representados pela linha vermelha . . . .  | 44 |
| 4.2 | Média a <i>posteriori</i> da distribuição preditiva a um passo para diferentes tamanhos de amostra junto ao seu intervalo de credibilidade. As escalas dos gráficos diferem para facilitar a visualização. . . . .  | 44 |
| 4.3 | Série Observada . . . . .   | 51 |
| 4.4 | Predição 1 passo à frente via geometria da informação compatibilizando modelos e sem compatibilizar, <i>Conjugate Updating</i> e modelo de nível local junto a série observada. Considerou-se que após 10 passos todos os modelos já tinham aprendido o comportamento da série. . . . . | 55 |
| 4.5 | $E(\int_0^T D_T)$ estimado por diferentes métodos . . . . .   | 55 |
| 4.6 | Média à <i>posteriori</i> suavizada para o par de harmônicos. . . . .   | 56 |

|     |  |    |
|-----|--|----|
| 5.1 | Boxplots para as médias a <i>posteriori</i> dos parâmetros $\mu_{1t}$ e $\mu_{2t}$ junto aos seus valores verdadeiros (linha vermelha). . . . .        | 72 |
| 5.2 | $E(\mu_{1t} D_t)$ e $E(\mu_{2t} D_t)$ (linha pontilhada azul) junto aos seus valores verdadeiros (linha contínua preta). . . . .                       | 73 |
| 5.3 | $E(Y_{1t+1} D_t)$ , $E(Y_{2t+1} D_t)$ e $E(Y_{3t+1} D_t)$ (linha pontilhada azul) junto a série observada (linha contínua preta). . . . .              | 74 |
| 5.4 | Série mensal do log-retorno da SP500 . . . . .   | 75 |
| 5.5 | Série observada junto a predição 1 passo a frente e seu respectivo intervalo de credibilidade. . . . .   | 76 |
| 5.6 | Média a <i>posteriori</i> filtrada e suavizada junto ao intervalo de credibilidade da média suavizada para o parâmetro da média e de precisão. . . . . | 76 |
| A.1 | Cadeias resultantes obtidas pelo método MCMC para cada um dos parâmetros e para o parâmetro $\lambda$ . . . . .  | 80 |

# Capítulo 1

## Introdução

Propostos por [Nelder and Wedderburn \(1972\)](#), os modelos lineares generalizados são uma extensão dos modelos lineares que permitem que a variável resposta tenha uma distribuição de probabilidade diferente da gaussiana, desde que sua distribuição pertença à família exponencial de distribuições. Entretanto, no contexto de séries temporais seu uso não é o mais adequado, pois é assumida a independência entre as observações, o que não ocorre em séries temporais.

[Harrison and Stevens \(1976\)](#) definiram a classe dos modelos lineares dinâmicos e desenvolveram uma abordagem bayesiana para tal tipo de modelagem e previsão. A ideia é que os parâmetros possam variar no tempo. Desta forma, a relação entre as covariáveis e variável resposta é dinâmica, mantendo, condicionalmente aos estados a independência dos observáveis.

[Smith \(1979\)](#) propõe uma extensão da metodologia de [Harrison and Stevens \(1976\)](#) na qual considera a modelagem de dados com distribuição diferente da gaussiana. O autor busca encontrar formas alternativas para atualizar o nível da série, uma vez que a simples adição de termos de erro para inflação de incerteza acarreta intratabilidade para a *priori*, em contextos mais genéricos que o normal. É utilizado então um processo de tomada de decisão bayesiano, de forma que sejam satisfeitas as seguintes condições: as decisões relacionadas a uma determinada função perda deverão permanecer constantes enquanto não houver informações futuras, já perda esperada associada a essas decisões aumentará no decorrer deste intervalo. A ideia é que a *priori* para os estados no tempo

$t$  é proporcional à *priori* do tempo  $t - 1$  elevada a uma determinada potência  $\beta$ . Dessa forma, mantém-se a média da distribuição e a incerteza aumenta. Uma explicação mais detalhada sobre tal modelo poderá ser encontrada na Seção 3.2.2.

Nessa linha de pensamento, [Harvey and Fernandes \(1989\)](#) apresentam uma forma de modelar a média da série para dados qualitativos e de contagem. Um hiperparâmetro é introduzido no modelo de forma que as observações passadas são descontadas ao se fazer previsões. O método de estimação escolhido para este hiperparâmetro foi o de máxima verossimilhança. Variáveis explicativas podem ser inseridas através da função de ligação do Modelo Linear Generalizado (MLG). Também é permitido que efeitos sazonais e de tendência sejam considerados na modelagem, contudo não é permitido que estes sejam estocásticos.

[Gamerman et al. \(2013\)](#) introduzem a classe dos Modelos Dinâmicos da Família Gama. Também é apresentada uma especificação expandida para a evolução do sistema que, assim como em [Harvey and Fernandes \(1989\)](#), utiliza um hiperparâmetro na distribuição da componente do nível do modelo que atua como um fator de desconto. Como métodos de inferência para tal, são utilizados o Método de Máxima Verossimilhança e estimadores Bayesianos. Uma vantagem do método é que a forma analítica da verossimilhança é mantida.

Uma outra alternativa de inferência para modelos lineares dinâmicos generalizados é o método via compatibilização de momentos proposto por [West et al. \(1985\)](#). Este utiliza métodos de aproximação para obter as estimativas do modelo, uma vez que somente as distribuições do parâmetro canônico e da preditiva têm forma analítica fechada. Uma das dificuldades encontradas é que há duas distribuições a *priori* para o parâmetro natural da família exponencial: uma induzida pelo vetor de estados e outra obtida por conjugação da família exponencial. A solução encontrada por [West et al. \(1985\)](#), foi compatibilizar os momentos de tais prioris.

No que concerne à compatibilização de prioris, [Poole and Raftery \(2000\)](#) e referências citadas neste artigo propõem uma metodologia denominada *Bayesian Melding*. São sugeridas duas formas de compatibilização: *linear pooling* e *logarithm pooling*. No primeiro caso, sugere-se uma combinação convexa de *prioris*, isto é, atribui-se pesos para cada

uma delas. Já no segundo caso, é proposta uma multiplicação de *prioris*, de forma que cada uma é elevada a um determinado peso. Argumenta-se que o segundo caso é mais eficiente no sentido de que pode-se compatibilizar e depois atualizar a *priori* via Teorema de Bayes, ou atualizar e depois compatibilizar, obtendo-se a mesma *posteriori*, o que não acontece quando utilizada a metodologia *linear pooling*.

Nesta dissertação, será proposta uma outra forma de compatibilização de *prioris* que utiliza conceitos de Geometria da Informação tais como Teorema da Projeção e Divergência de Kullback-Leibler. A ideia é projetar a *priori* induzida para o parâmetro canônico pelo vetor de estados no espaço da *priori* conjugada para este parâmetro e depois compatibilizá-las. Uma vez que as *prioris* são compatibilizadas, torna-se simples atualizá-las. Duas distribuições a *posteriori* são obtidas para o preditor linear: uma induzida pela conjugação do parâmetro canônico e outra induzida pela distribuição assumida para os estados. A solução encontrada foi projetar a distribuição induzida pela conjugação do parâmetro natural no espaço da distribuição induzida pelo vetor de estados. Uma vez que tal compatibilização seja feita, utilizamos o método Linear Bayes para obter os momentos do vetor de estados.

Souza et al. (2016) expandem a metodologia proposta por West et al. (1985) para a família exponencial biparamétrica. Tal expansão não é simples de ser feita, pois há de se resolver um sistema que conta, eventualmente, com mais equações, referentes aos momentos relacionados a média e precisão do vetor de preditores lineares, do que parâmetros provenientes da *priori* conjugada da família exponencial. Como solução, foi utilizado o Método de Momentos Generalizado. Serão apresentados neste trabalho aplicações de tal metodologia, bem como uma solução para o caso Multinomial, sendo esta uma extensão do trabalho de Souza et al. (2016).

Este trabalho está organizado em 6 capítulos, incluindo este. O segundo capítulo contém conceitos básicos de família exponencial e geometria da informação para desenvolvimento das metodologias a serem propostas. O capítulo 3 revisita os métodos de estimação para MLDG sugeridos por West et al. (1985) e Gamerman et al. (2013). O quarto capítulo traz uma proposta de inferência para os MLDG da família exponencial uniparamétrica via geometria da informação. O capítulo 5 apresenta uma abordagem via

método *Conjugate Updating* estendido para a família exponencial k-paramétrica, sugerida por Souza et al. (2016). Por fim, o último capítulo apresenta as conclusões e trabalhos futuros desta pesquisa.



# Capítulo 2

## Conceitos Básicos

Ao longo deste capítulo serão apresentados conceitos básicos que serão utilizados no desenvolvimento do modelo proposto. As seções deste capítulo estão divididas entre Família Exponencial e Geometria da Informação.

### 2.1 Família Exponencial

Uma distribuição é pertencente à família exponencial se a sua função densidade de probabilidade,  $p(y; \eta)$ , pode ser escrita da seguinte forma (Bernardo and Smith (2001)):

$$p(y; \eta) = f(y)g(\eta) \exp \left( \sum_{i=1}^k c_i \eta_i h_i(y) \right) \quad (2.1)$$

onde  $\eta = (\eta_1, \dots, \eta_k)$ ,  $\eta = (\eta_1, \dots, \eta_k)$  e dadas as funções  $f; h_i$  e as constantes  $c_i$ ,

$$\frac{1}{g(\eta)} = \int_Y f(y) \exp \left( \sum_{i=1}^k c_i \eta_i h_i(y) \right) dy < \infty \quad (2.2)$$

Ademais, se  $y_1, \dots, y_n \in Y$  é uma sequência permutável tal que, dada a família exponencial regular k-paramétrica,

$$p(y_1, \dots, y_n) = \int_H \prod_{i=1}^n p(y_i; \eta) dQ(\eta); \quad (2.3)$$

para algum  $dQ(\cdot)$ , então,

$$T_n = T_n(y_1, \dots, y_n) = \left( \prod_{i=1}^n h_1(y_i), \dots, \prod_{i=1}^n h_k(y_i) \right); n = 1, 2, \dots \quad (2.4)$$

é uma sequência de estatísticas suficientes.

Usualmente, a família exponencial é apresentada em sua forma canônica. Tal representação pode ser obtida após algumas transformações na definição apresentada, resultando na seguinte expressão:

$$p(x_j) = c(x) \exp \left( \sum_{i=1}^k \eta_i h_i(x) \right) \quad (2.5)$$

Em que  $x = (x_1, \dots, x_k)$ ,  $x_i = h_i(y)$ ;  $\eta_i = c_i(\theta)$  e  $\theta = (\theta_1, \dots, \theta_k)$  onde  $i = 1, \dots, k$ .

Utilizando elementos de sua representação na forma canônica, podemos encontrar média e variância da distribuição em análise, isto é,

$$E(x_j) = \eta_j; V(x_j) = \eta_j^2 \quad (2.6)$$

A seguir serão apresentados exemplos de distribuições pertencentes à família exponencial.

**Exemplo 2.1.1** *Poisson*( $\theta$ ):

$$\begin{aligned} p(y_j) &= \frac{\theta^y e^{-\theta}}{y!} \\ &= \frac{1}{y!} \exp \left( y \log(\theta) - \theta \right) \quad g: \end{aligned}$$

Dessa forma, a distribuição Poisson pertence a família exponencial uniparamétrica com:  $x = h(y) = y$ ,  $c(y) = 1/y!$ ,  $\eta = \log(\theta)$  e  $b(\theta) = \theta$ .

**Exemplo 2.1.2**  $Normal(\mu; \sigma^2)$ :

$$\begin{aligned} p(y; \mu, \sigma^2) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right] \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y^2 - 2y\mu + \mu^2)\right] \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right] \frac{1}{\sigma\sqrt{2\pi}} \end{aligned}$$

Portanto a distribuição Normal pertence a família exponencial biparamétrica com vetor paramétrico  $\eta = (\mu; \sigma^2)$ , onde:  $x_1 = h_1(y) = y$ ,  $x_2 = h_2(y) = y^2$ ,  $c(\eta) = \frac{1}{\sigma\sqrt{2\pi}}$ ,  $\eta = [(\mu; \sigma^2); (\mu = 2\mu^2)]$  e  $b(\eta) = (1/2\sigma^2) \mu^2 + \log(\sigma\sqrt{2\pi})$ .

**Exemplo 2.1.3**  $Bernoulli(p)$ :

$$\begin{aligned} p(y; p) &= p^y(1-p)^{1-y} \\ &= \exp\left[y \log p + (1-y) \log(1-p)\right] \end{aligned}$$

Assim, a distribuição Bernoulli pertence a família exponencial uniparamétrica onde:  $x = h(y) = y$ ,  $c(p) = 1$ ,  $\eta = \log p$  e  $b(\eta) = \log(1-p)$ .

**Exemplo 2.1.4** Multinomial com  $p$  categorias, assumindo  $\sum_{i=1}^p y_i = n$  e  $\sum_{i=1}^p p_i = 1$ :

$$\begin{aligned} p(\mathbf{y}; \mathbf{p}) &= \frac{n!}{\prod_{i=1}^p y_i!} \prod_{i=1}^p p_i^{y_i} \\ &= \frac{n!}{\prod_{i=1}^p y_i!} \exp\left[\sum_{i=1}^p y_i \log(p_i)\right] \end{aligned}$$

Portanto a distribuição Multinomial pertence a família exponencial  $p$ -paramétrica com vetor paramétrico  $\eta = (p_1; \dots; p_k)$ , onde:  $x_1 = h_1(y) = y_1$ ,  $x_2 = h_2(y) = y_2; \dots; x_p = h_p(y) = y_p$ ,  $c(\eta) = \frac{n!}{\prod_{i=1}^p y_i!}$  e  $\eta = (\log(p_1); \dots; \log(p_p))$ .

## 2.1.1 Conjugação na Família Exponencial

Uma característica importante da família exponencial é que existe uma estatística suficiente para o vetor paramétrico com dimensão fixa. Assim, torna-se simples encontrar

a distribuição a *priori* conjugada à distribuições pertencentes a tal família. A ideia é que, uma vez atribuída uma distribuição a *priori* para os parâmetros, e sendo esta pertencente à família exponencial, a distribuição a *posteriori* pertencerá à mesma classe de distribuições da *priori* havendo apenas uma mudança nos hiperparâmetros.

Seja  $y = (y_1, \dots, y_n)$  uma amostra de uma distribuição pertencente à família exponencial tal que

$$p(y_j) = \prod_{j=1}^n f(y_j) [g(\theta)]^n \exp \left( \sum_{i=1}^k c_i(\theta) \sum_{j=1}^n y_j \right);$$

então a conjugada para  $\theta$  terá a seguinte forma (Bernardo and Smith (2001)):

$$p(\theta | y) = [K(\theta)]^{-1} [g(\theta)]^0 \exp \left( \sum_{i=1}^k c_i(\theta) \sum_{j=1}^n y_j \right); \quad \theta \in H;$$

onde  $K(\theta) = \int_H [g(\theta)]^0 \exp \left( \sum_{i=1}^k c_i(\theta) \sum_{j=1}^n y_j \right) d\theta < \infty$ ;

**Exemplo 2.1.5** *Poisson*( $\theta$ ):

$$\begin{aligned} p(y_j) &= \frac{\prod_{i=1}^n y_i! e^{-n}}{\prod_{i=1}^n y_i!} \exp \left( \sum_{i=1}^n y_i \log(\theta) \right) \\ &= \prod_{i=1}^n \frac{y_i! e^{-n}}{y_i!} \exp \left( \sum_{i=1}^n y_i \log(\theta) \right); \end{aligned}$$

Logo,

$$\begin{aligned} p(\theta | y; \theta_0, \theta_1) &\propto \exp \left( \sum_{i=1}^n y_i \log(\theta) \right) \\ &= \frac{1}{K(\theta_0, \theta_1)} \exp \left( \sum_{i=1}^n y_i \log(\theta) \right) \end{aligned} \quad (2.7)$$

Sendo este o núcleo da distribuição Gama, pois é da forma  $a e^{-b}$ . Assim, a *priori* conjugada será  $G(\alpha; \beta)$ , onde  $\alpha = \sum_{i=1}^n y_i + 1$  e  $\beta = \theta_0$ .

**Exemplo 2.1.6** *Bernoulli*( $p$ ):

$$\begin{aligned} p(\mathbf{y}|p) &= p^{\sum_{i=1}^n y_i} (1-p)^{n - \sum_{i=1}^n y_i} \\ &= (1-p)^n \exp \left( \sum_{i=1}^n y_i \log(p) + \sum_{i=1}^n (1-y_i) \log(1-p) \right) \end{aligned}$$

Assim,

$$\begin{aligned} p(j_0; \alpha_0, \beta_0) &\propto (1-p)^{\alpha_0} p^{\beta_0} \exp \left( -\sum_{i=1}^n y_i \log \frac{p}{1-p} \right) \\ &= \frac{1}{K(\alpha_0, \beta_0)} p^{\beta_0} (1-p)^{\alpha_0} \end{aligned}$$

Dessa forma, o núcleo encontrado remete a distribuição Beta.

**Exemplo 2.1.7** *Multinomial* com  $k$  categorias, assumindo  $\sum_{i=1}^k y_i = n$ ,  $\sum_{i=1}^k \alpha_i = 1$  e  $\sum_{i=1}^k \beta_i = 1$ :

$$\begin{aligned} p(\mathbf{y}) &= \frac{n!}{\prod_{i=1}^k y_i!} \prod_{i=1}^k \left( \frac{\alpha_i}{k} \right)^{y_i} \binom{n}{y_1, \dots, y_k} \\ &= \frac{n!}{\prod_{i=1}^k y_i!} \exp \left( \sum_{i=1}^k y_i \log \frac{\alpha_i}{k} + n \log \left( \frac{1}{k} \right) \right) \end{aligned}$$

Deste modo,

$$\begin{aligned} p(\mathbf{j}_1; \alpha_1, \dots, \alpha_k) &\propto \exp \left( \sum_{i=1}^k y_i \log \frac{\alpha_i}{k} + n \log \left( \frac{1}{k} \right) \right) \\ &= \frac{1}{K(\alpha_1, \dots, \alpha_k)} \exp \left( \sum_{i=1}^k y_i \log \frac{\alpha_i}{k} + n \log \left( \frac{1}{k} \right) \right) \end{aligned}$$

que é o núcleo da distribuição Dirichlet onde cada  $\alpha_i = \beta_i + 1; i = 1, \dots, k$ :

## 2.2 Geometria da Informação

### 2.2.1 Divergências

No presente trabalho, a noção de divergência será aplicada da seguinte forma: admitiremos que o vetor de estados que governam o preditor linear em um MDLG siga, *a priori*, distribuição gaussiana. Como veremos adiante, tal hipótese induz uma distribuição de probabilidade para o parâmetro canônico da família exponencial. Entretanto, como visto na subseção anterior, é possível obter distribuição *a priori* conjugada para tal parâmetro. Buscaremos compatibilizar a distribuição conjugada e a distribuição induzida pelos estados por meio da minimização da divergência entre estas.

A seguir, serão apresentadas as condições para que uma determinada divergência seja válida .

Sejam duas distribuições  $p(y)$  e  $q(y)$  em um espaço  $\mathcal{Y}$ , que possuem como parâmetros  $\rho$  e  $\omega$ , respectivamente. A divergência entre estas,  $D[\rho : \omega]$ , é uma função de tais parâmetros que satisfazem os seguintes critérios:

1.  $D[\rho : \omega] \geq 0$ .
2.  $D[\rho : \omega] = 0$ , se e somente se  $\rho = \omega$ .
3. Se  $p(y)$  e  $q(y)$  são distribuições suficientemente próximas, denotando suas coordenadas como  $\omega = \rho + d$ , a expansão de Taylor de  $D$  pode ser escrita como:

$$D[\rho : \omega + d] = \frac{1}{2} \sum_{i,j} g_{ij}(\rho) d_i d_j + O(d^3);$$

e a matrix  $\mathbf{G} = (g_{ij})$  é positiva definida, dependendo de  $\rho$ .

#### 2.2.1.1 Divergência de Bregman

Sejam duas distribuições de probabilidade  $p(y)$  e  $q(y)$ , parametrizadas por  $\rho$  e  $\omega$ , respectivamente. A divergência de  $P$  para  $Q$ , derivada de uma função convexa  $\psi$ , pode ser escrita como:

$$D[p(y) : q(y)] = (p - q)^T r(q); \quad (2.8)$$

onde  $r$  representa o gradiente da função.

Se  $(u)$  for definida como  $\sum_i u_i \log(u_i)$ , obtém-se a divergência de Kullback-Leibler a ser apresentada a seguir.

### 2.2.1.2 Divergência de Kullback-Leibler

Uma das métricas mais utilizadas para se medir dissimilaridades entre distribuições é a divergência de Kullback-Leibler. Podemos defini-la da seguinte forma: sejam  $p(y)$  e  $q(y)$  duas distribuições de probabilidade da variável aleatória  $Y$ , a divergência de Kullback-Leibler será dada por:

$$D_{KL}[p(y) : q(y)] = \int p(y) \log \frac{p(y)}{q(y)} dy \quad (2.9)$$

Se  $Y$  for uma variável discreta, a integral é substituída pelo somatório. A seguir encontra-se a prova de que a divergência de KL é de fato uma divergência.

#### Prova das condições:

(a)  $D_{KL}[p(y) : q(y)] \geq 0$ :

Seja  $D = \{y : p(y) > 0\}$  o espaço de  $p(y)$ . Admita  $Y$  variável aleatória contínua,

$$D_{KL}[p(y) : q(y)] = \int p(y) \log \frac{p(y)}{q(y)} dy \quad (2.10)$$

$$= \int p(y) \log \frac{q(y)}{p(y)} dy \quad (2.11)$$

$$= \int \log \frac{q(y)}{p(y)} p(y) dy \quad (2.12)$$

$$= 0 \quad (2.13)$$

Logo,  $D_{KL}[p(y) : q(y)] \geq 0$ . A equação (2.12) é resultado obtido da Desigualdade de Jensen.

(b) Se  $p(y) = q(y)$ :

$$\begin{aligned} D_{KL}[p(y) : p(y)] &= \int p(y) \log \frac{p(y)}{p(y)} dy \\ &= 0 \end{aligned}$$

(c) Por expansão de Taylor:

$$f(\theta) = f(\theta_0) + f_i(\theta_0)(\theta - \theta_0)^i + \frac{1}{2} f_{ij}(\theta_0)(\theta - \theta_0)^i(\theta - \theta_0)^j + O(\|\theta - \theta_0\|^3) \dots$$

Onde  $f_i(\theta)$  representa a primeira derivada e  $f_{ij}(\theta)$  a segunda derivada da função a ser expandida em série de Taylor; no caso em tela, a própria Divergência de KL. Conforme observado anteriormente, temos que  $f(\theta_0) = 0$  pois  $D_{KL}[p(y) : p(y)]$  é igual a zero por (a). Temos inclusive que  $f_i(\theta_0)$  também é igual a zero, pois é o ponto mínimo da função e a derivada do mínimo é zero.

Sejam as distribuições  $p(y)$  e  $q(y)$  parametrizadas por  $\mu$  e  $\mu + d$ , respectivamente. Assim,

$$D_{KL}[p : q + d] = \frac{1}{2} \sum g_{ij}(\mu) d_i d_j + O(\|d\|^3);$$

onde,  $g_{ij}(\mu) = f_{ij}(\theta_0)(\theta - \theta_0)^i(\theta - \theta_0)^j$

### 2.2.2 Teoremas

Os teoremas apresentados aqui servirão como base para o desenvolvimento da metodologia proposta para a classe dos Modelos Lineares Dinâmicos Generalizados. O teorema da projeção nos indica que a menor distância entre duas distribuições vai ser equivalente a minimizar a Divergência de KL entre elas, o que também equivale a igualar os momentos na família exponencial. Já o teorema seguinte nos auxilia no que concerne à abordagem que compatibiliza modelos.



**Teorema 2.2.1 (Teorema da projeção)** *Seja  $p(y)$  uma distribuição de probabilidade em um espaço  $Y$ . Considere  $S$  um modelo da família exponencial em  $Y$ . O Teorema de Projeção afirma que a distribuição  $q(y)$  que minimiza a divergência  $D_{KL}[p : q] ; q(y) \in S$  é tal que  $E_q(h_i(y)) = E_p(h_i(y)) \quad \forall i$ .*

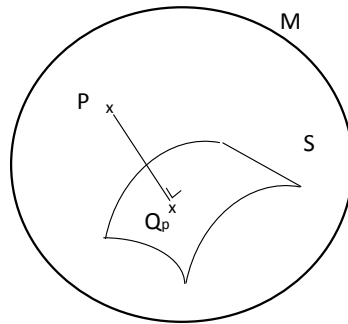


Figura 2.1: Exemplo do Teorema da Projeção

**Prova:**

Seja  $q^* = \operatorname{argmin}_{q \in S} D_{KL}[p : q] = \int p \log p - \int p \log q$ . Observe que a primeira integral não depende dos parâmetros em  $q(y)$ , portanto focaremos na segunda integral. Temos que

$$\begin{aligned} \int p(y) \log q(y) &= \int p(y) f h(y) - b(\eta) g \\ \frac{\partial}{\partial \eta_i} &= \int p(y) h_i(y) - \frac{\partial}{\partial \eta_i} b(\eta) \\ &= E_p(h_i(y)) - \frac{\partial}{\partial \eta_i} b(\eta) = 0 \\ \frac{\partial}{\partial \eta_i} b(\eta) &= E_p(h_i(y)) \end{aligned} \tag{2.14}$$

**Teorema 2.2.2** *Sejam  $R, Q$  e  $U$  distribuições de probabilidade equivalentes e pertencentes a família exponencial, parametrizadas respectivamente por  $\eta_r, \eta_q$  e  $\eta_u$ . Por definição, estas estão sob a mesma geodésica e portanto uma combinação convexa pode ser feita nos*

parâmetros de forma que  $x$  pode ser escrito da seguinte forma:

$$u = S q + (1 - S) r \quad (2.15)$$

De forma que o uso dos teoremas (2.2.1) e (2.2.2) simultaneamente, resultam na compatibilização das distribuições em uma das abordagens propostas.

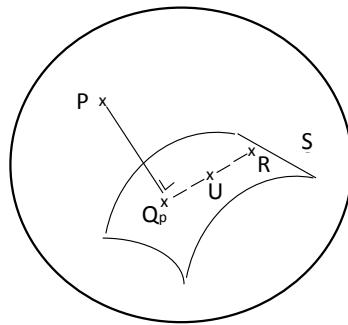


Figura 2.2: Exemplo de Compatibilização de Variáveis

## 2.2.3 Exemplos de Projeções

### 2.2.3.1 Projetando a Distribuição Log-Normal na Gama

Seja  $p(y)$  a distribuição Log-Normal. Queremos encontrar a distribuição Gama que melhor aproxima tal distribuição segundo a divergência de Kullback-Leibler. Tal distribuição está representada aqui por  $q(y)$ .

Pelo Teorema (2.2.1) apresentado, sabe-se que devemos encontrar a Gama que minimiza a divergência de Kullback Leibler.

Assim, deseja-se encontrar os parâmetros da Gama que minimizam a distância de Kullback-Leibler. Ou seja, tem-se por objetivo encontrar o

$$\min D_{KL}(p : q) = \min \int p(y) \log \frac{p(y)}{q(y)} dy$$

onde,  $p(y)$  é a distribuição Log-Normal( ,  $\sigma^2$ ) e  $q(y)$  é a distribuição Gama( $a; b$ ).

Podemos reescrever a divergência como

$$D_{KL}(p : q) = \int p(y) \log p(y) dy - \int p(y) \log q(y) dy \quad (2.16)$$

Observe que minimizar a divergência de Kullback-Leibler é equivalente a maximizar a segunda integral, pois  $\int p(y) \log p(y) dy$  é uma constante que não depende dos parâmetros  $a$  e  $b$ .

Como,

$$\log q(y) = a \log b + (a - 1) \log y - by - \log \Gamma(a)$$

$$\begin{aligned} l(a; b) &= \int p(y) \log q(y) dy = a \log b + (a - 1) E_p(\log y) - b E_p(y) - \log \Gamma(a) \\ &= a \log b + (a - 1) \left[ -\frac{1}{2} \left( \frac{1}{\sigma^2} + (1-2) \frac{1}{\sigma^2} \right) \right] - \log \Gamma(a) \end{aligned} \quad (2.17)$$

Derivando e igualando a equação (2.17) a zero é possível encontrar os parâmetros da Gama que melhor aproxima uma distribuição Log-Normal.

Dessa forma,

$$\frac{\partial l(a; b)}{\partial a} = \log b + \psi(a) \quad (2.18)$$

$$\frac{\partial l(a; b)}{\partial b} = \frac{a}{b} \exp\left(-\frac{a}{b}\right) \quad (2.19)$$

onde  $\psi(a)$  é a função *digamma* de  $a$ .

Igualando as derivadas acima a zero temos que:

$$b = \exp\left(-\psi(a)\right) \quad (2.20)$$

$$a = b \exp\left(-\frac{a}{b}\right) \quad (2.21)$$

Nas equações (2.20) e (2.21), os parâmetros  $a$  e  $b$  ainda não puderam ser isolados de forma que um não dependesse mais do outro. Assim, o *log* foi aplicado nas equações como forma de contornar esse problema. Além disso, aplicou-se uma aproximação para  $\psi(a)$  como proposto por Abramovitz:

$$\psi(a) = \log a - \frac{1}{2a} + O\left(\frac{1}{a^2}\right); \text{ se } a \rightarrow \infty$$

Dessa forma,

$$\begin{aligned} \log b &= \log a - \frac{1}{2a} \\ \log a &= \log b + \frac{a}{b} \exp\left(-\frac{a}{b}\right) \end{aligned} \quad (2.22)$$

Resolvendo esse sistema de equações,

$$\begin{aligned} \frac{1}{2a} &= \frac{a}{b} \exp\left(-\frac{a}{b}\right) \\ a &= \frac{1}{2} \end{aligned} \quad (2.23)$$

$$\begin{aligned}
 b &= a \exp\left(\frac{1}{2} \sigma^2\right) \\
 &= \frac{1}{2} \exp\left(\frac{1}{2} \sigma^2\right):
 \end{aligned}
 \tag{2.24}$$

Foi possível, portanto, encontrar os valores de  $a$  e  $b$

$$a = \frac{1}{\sigma} e^{-\frac{1}{2\sigma^2}} \text{ e } b = \frac{1}{2} \exp\left(\frac{1}{2} \sigma^2\right) :
 \tag{2.25}$$

A Figura (2.3) exibe exemplos de aproximações Gama para distribuições Log-Normal obtidas por meio da minimização da divergência de Kullback-Leibler.

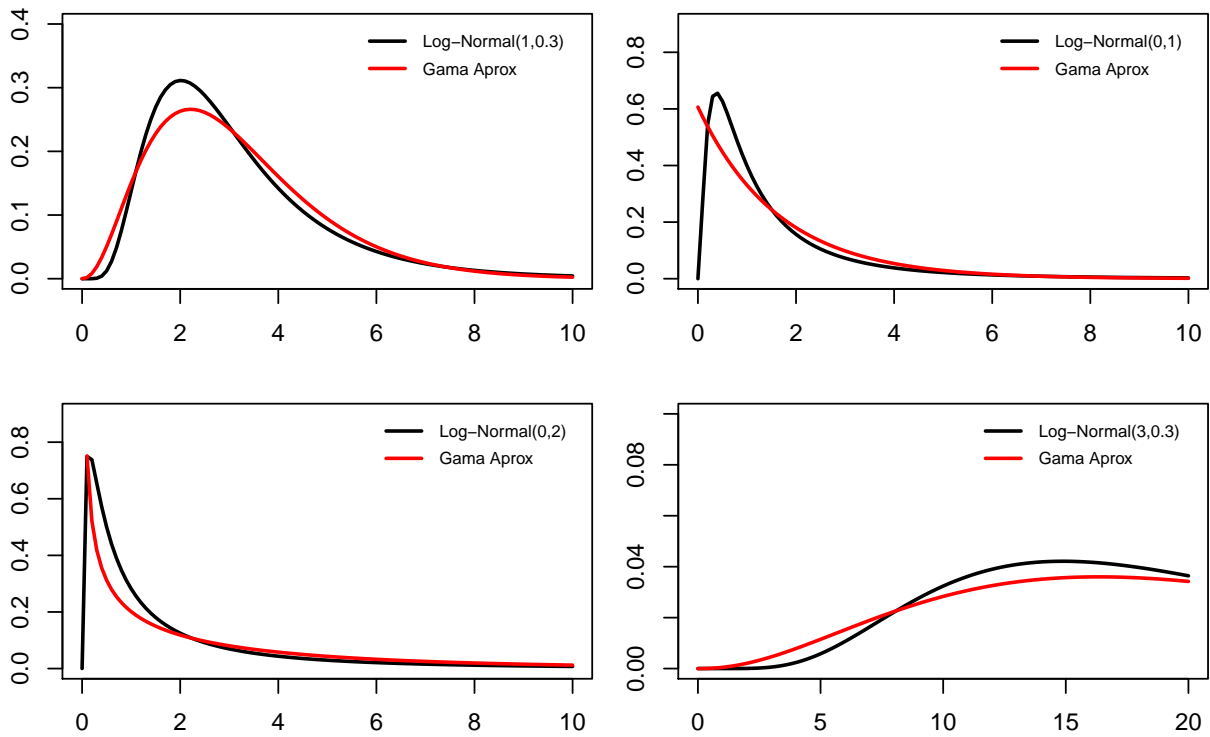


Figura 2.3: Distribuições Log-Normal e Gama aproximadas através da minimização da divergência de Kullback-Leibler quando utilizados diferentes parâmetros para as distribuições Log-Normal.

### 2.2.3.2 Projetando a Distribuição Logito-Normal na Beta

Seja  $p(y)$  a distribuição Logito-Normal( ; 2). Queremos encontrar a distribuição Beta( $a; b$ ) que melhor aproxima tal distribuição segundo a distância de Kullback-Leibler. Tal distribuição Beta está representada aqui por  $q(y)$ .

$$D_{KL}(p: q) = \int p(y) \log p(y) dy - \int p(y) \log q(y) dy: \quad (2.26)$$

Assim, basta maximizar a segunda integral, pois a primeira não depende dos parâmetros da Beta que se deseja obter.

$$\log q(y) = (a - 1) \log(y) + (b - 1) \log(1 - y) + \log \frac{(a + b)}{(a) (b)} :$$

$$l(a; b) = \int p(y) \log q(y) dy = (a - 1) E_p(\log(y)) + (b - 1) E_p(\log(1 - y)) + \log \frac{(a + b)}{(a) (b)}$$

Os momentos da distribuição Logito-Normal não têm solução analítica. Para contornar esse problema, foi utilizado o método de Monte-Carlo para encontrar  $E(\log(y))$  e  $E(\log(1 - y))$ . Uma vez encontrados tais valores esperados, segue-se para o próximo passo, que é encontrar os parâmetros  $a$  e  $b$  da Beta.

Dessa forma,

$$\frac{\partial l(a; b)}{\partial a} = E(\log(y)) + \psi(a + b) - \psi(a) \quad (a) \quad (2.27)$$

$$\frac{\partial l(a; b)}{\partial b} = E(\log(1 - y)) + \psi(a + b) - \psi(b) \quad (b) \quad (2.28)$$

onde,  $\psi(\cdot)$  representa a função *digamma* de  $\Gamma$ . Aqui também utilizou-se como aproximação de  $\psi(\cdot)$  como proposto por Abramovitz:

$$\psi(x) = \log(x) - \frac{1}{2x}:$$

Dessa forma,

$$\begin{aligned} \log b \frac{1}{2b} &= E(\log(1 - y)) + (a + b); \\ \log a \frac{1}{2a} &= E(\log(y)) + (a + b): \end{aligned} \quad (2.29)$$

Observe que não é possível isolar os parâmetros  $a$  e  $b$  analiticamente. Para que tais resultados pudessem ser encontrados, foi utilizado o método de Newton-Raphson.

A Figura(2.4) exibe exemplos de aproximações Beta para distribuições Logito-Normal obtidas por meio da minimização da divergência de Kullback-Leibler.

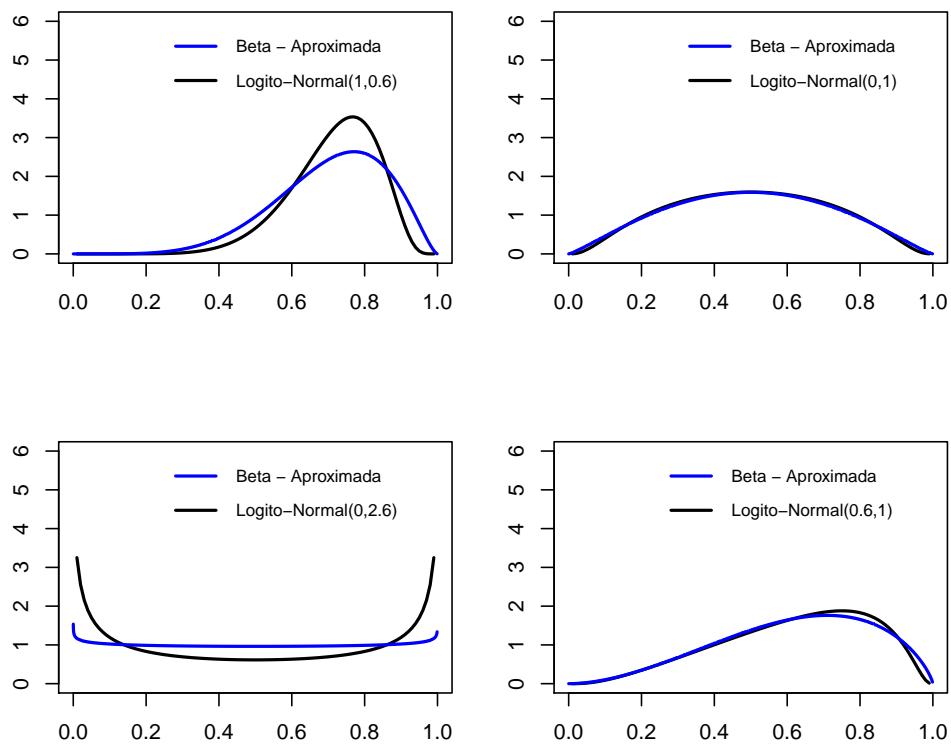


Figura 2.4: Distribuições Logito-Normal e Beta aproximadas através da minimização da divergência de Kullback-Leibler quando utilizados diferentes parâmetros para as distribuições Logito-Normal.

# Capítulo 3

## Modelos Dinâmicos Generalizados

É bastante usual ter-se interesse na modelagem de variáveis que não são acomodadas por uma distribuição normal. Por exemplo, variáveis binárias ou de contagem e variáveis com comportamento assimétrico. Este capítulo tem por objetivo apresentar a classe dos modelos lineares generalizados tanto na sua forma estática quanto dinâmica. No que se refere à inferência na classe dos modelos lineares dinâmicos generalizados, serão apresentados os métodos de inferência via *Conjugate Updating* e via modelos dinâmicos que admitem evolução somente da média o qual denominaremos de modelos dinâmicos não estruturados.

Os modelos dinâmicos lineares generalizados podem ser vistos como uma generalização dos modelos lineares generalizados, no sentido de aliar à flexibilidade de respostas não gaussianas, efeitos dinâmicos associados a covariáveis e outras componentes estruturais. Tal mecanismo de dinâmica naturalmente acomoda a autocorrelação inerente a dados temporalmente observados.

Os modelos lineares generalizados são uma extensão dos modelos lineares para resposta gaussiana. Propostos por Nelder e Wedderburn (1972), estes permitem que a variável que se deseja analisar tenha uma distribuição de probabilidade diferente da gaussiana, uma vez que sua distribuição pertença à família exponencial de distribuições.

Os modelos lineares generalizados contêm três componentes, sendo:

- (a) a variável resposta  $Y$ , que é a componente aleatória do modelo e para o qual assume-se



distribuição pertencente à família exponencial na forma canônica.

- (b) componente sistemática dada por um preditor linear  $\eta$  construído com base nas variáveis explicativas do modelo,  $\mathbf{x} = (x_1; \dots; x_p)$ ;
- (c) uma função  $g$  monótona e diferenciável que liga o valor esperado de  $T$ ,  $\mu = E(Y)$ , ao preditor linear  $\eta$  denominada função de ligação. Embora a função de ligação possa ser especificada arbitrariamente, uma forma de se determinar  $g$  é através da parametrização natural (canônica) da família exponencial. Tem-se então:

$$\mu = g(\eta) \tag{3.1}$$

onde  $\eta$  é o preditor linear,  $\mu$  é a média e  $g$  a função de ligação.

É assumido que a média transformada segue um modelo linear, de forma que:

$$\mu = \mathbf{x}^T \boldsymbol{\beta} \tag{3.2}$$

Como a função  $g$  é um-a-um,

$$\eta = g^{-1}(\mu): \tag{3.3}$$

Há diversas formas de se escolher a função de ligação a ser utilizada, sendo o método mais usual é considerar  $\eta = \mu$ , onde  $\mu$  é chamado parâmetro canônico da distribuição. Dessa forma, esse tipo de função de ligação é chamada de *função de ligação canônica*.

### 3.1 Modelos Lineares Dinâmicos

Um caso particular dos modelos dinâmicos lineares generalizados são os modelos lineares dinâmicos, para os quais pressupõe-se resposta gaussiana. Harrison and Stevens (1976) definiram a classe dos modelos lineares dinâmicos e desenvolveram uma abordagem Bayesiana para tal tipo de modelagem e previsão. A ideia é que os parâmetros possam variar no tempo. Desta forma, a relação entre as covariáveis e a variável resposta é dinâmica. É importante ressaltar que os MLG (e os Modelos Lineares normais, que são um de seus casos particulares) não dão tratamento à dependência temporal inerente a dados observados ao longo do tempo, o que ocorre no caso dos Modelos Lineares Dinâmicos.

Sejam:

$x_t$  o vetor de estados de dimensão  $p \geq 1$ ;

$Y_t$  o vetor de observações de dimensão  $r \geq 1$ ;

$F_t$  a matriz que contém as covariáveis no tempo  $t$  de dimensão  $p \times r$ ;

$G_t$  a matriz de evolução paramétrica no tempo  $t$  de dimensão  $p \times p$ ;

$V_t$  e  $W_t$  matrizes de covariância, sendo a primeira relacionada ao erro observacional e a segunda ao erro evolutivo.

Um modelo dinâmico é caracterizado pela quádrupla  $(F_t; G_t; V_t; W_t)$  e sua estrutura mais geral pode ser definida da seguinte forma:

$$\text{Equação Obs.: } Y_t = F_t' x_t + \epsilon_t \text{ onde } \epsilon_t \sim N(0; V_t) \quad (3.4)$$

$$\text{Equação Evol.: } x_t = G_t x_{t-1} + \eta_t \text{ onde } \eta_t \sim N(0; W_t) \quad (3.5)$$

onde a primeira equação é a **observacional** e a segunda é a de **evolução**.

Seja  $D_t$  o histórico das informações até o tempo  $t$  e admita  $D_t = D_{t-1} \cup \{Y_t\}$ , com  $D_0$  denotando a informação que se tem antes da primeira observação. Assuma a informação inicial como  $(x_0 | D_0) \sim N(m_0; C_0)$ .

Condicionalmente a  $x_t, y_t$  independe de  $y_{t-1}; \dots; y_1$ . Observe-se, entretanto, a asso-

ciação entre  $y_t$  e  $y_{t-1}$ , explicitada na equação de evolução em 3.5. O tratamento da autocorrelação temporal entre as quantidades observáveis  $y_1, \dots, y_t$  é viabilizado por tal relação, como ilustra a figura 3.1:

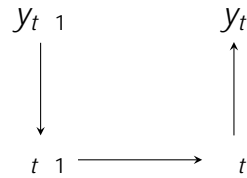


Figura 3.1: Ilustração da evolução do sistema

Os erros observacionais e evolutivos são independentes ao longo do tempo e uns com os outros. Os erros  $\epsilon_t$  controlam a evolução através de sua variância, elementos da diagonal principal de  $W_t$ . Quanto maior (menor) seu valor, mais errática (suave) a variação da média será e a média zero dos erros garante constância local. A relação entre  $W_t$  e  $V_t$  fornece indícios para avaliar de onde vem o movimento da série. No caso  $W/V$  pequeno teremos que boa parte deste movimento se dá devido às observações. Enquanto  $W/V$  grande implica que o movimento se dá devido às observações, mas também às variações do nível da série, sendo este  $F_t^0$ .

Considere o caso  $Y$  escalar e a variância observacional constante e conhecida igual a  $V_t = V = 1 = \sigma^2$ . Assim, as distribuições *a priori* para o vetor de estados, *posteriori* e previsão um passo a frente serão dadas por:

(a) *priori*:

$$(y_t | D_{t-1}) \sim N(a_t; R_t) \quad (3.6)$$

$$a_t = G_t m_{t-1};$$

$$R_t = G_t C_{t-1} G_t^0 + W_t;$$

(b) Previsão 1 passo a frente:

$$(Y_t | D_{t-1}) \sim N(f_t; q_t) \quad (3.7)$$

$$f_t = F_t^0 a_t;$$

$$q_t = F_t^0 R_t F_t + V;$$

(c) *posteriori*:

$$(Y_t | D_t) \sim N(m_t, C_t) \quad (3.8)$$

$$m_t = a_t + R_t F_t (e_t) = (q_t + 1); \quad (3.9)$$

$$C_t = R_t^{-1} F_t (F_t R_t)^0 = (q_t + 1) \quad (3.10)$$

$$e_t = Y_t - f_t$$

Harrison and Stevens (1976) indicam que a especificação de  $W_t$  pode ser difícil de ser feita, então, uma alternativa que foi proposta por Ameen and Harrison (1984) é utilizada. Eles propuseram a utilização de fatores de desconto, sendo estes a porcentagem de informação que passa de um período a outro. Desta forma, a matriz  $W_t$  passa a ser especificada a partir de uma matriz de fatores de desconto  $B_t$  e  $R_t$  passa a ser dada por:

$$R_t = B_t G_t C_t^{-1} G_t^0 B_t;$$

sendo  $B$  a matriz diagonal  $r = r$  de descontos  $\rho_i^{-1}$  com  $0 < \rho_i < 1$  para  $i = 1, \dots, p$ . A idéia de fator de desconto é a seguinte: à medida que a informação "envelhece", torna-se menos útil e, portanto, seu peso deve diminuir. Ou seja, o fator de desconto pode ser pensado como a porcentagem de informação passada, a cada tempo  $t$ , aos tempos posteriores. Quanto mais próximo de 1 for o desconto, mais suave será a evolução dos estados, enquanto quanto menor, mais brusca ela será, pois teremos mais incerteza para a predição. Se o desconto for 1, desde que  $G$  seja a matriz identidade e os erros de evolução tenham média nula, o estado será fixo no tempo, ou seja, o modelo é estático no tempo. Caso haja mais de uma componente no modelo, seja de tendência ou sazonalidade, é importante ressaltar a aplicação prática sugerida por (West and Harrison, 1996, p. 193:202) de aplicar não aplicar desconto fora da diagonal de  $G_t C_t^{-1} G_t^0$ .

## 3.2 Modelos Lineares Dinâmicos Generalizados

Esta seção tem por objetivo discutir métodos de inferência para a classe dos modelos lineares dinâmicos generalizados (MLDG). Com o intuito de tornar mais claro o objetivo da discussão de tais métodos de inferência, segue como exemplo o caso Poisson.

Seja  $y_t^j \mid \theta_t \sim \text{Poisson}(\theta_t)$ . Observe que a densidade desta distribuição pode ser escrita na família exponencial da seguinte forma:

$$\begin{aligned} p(y_t^j \mid \theta_t) &= \frac{\theta_t^{y_t^j} e^{-\theta_t}}{y_t^j!} \\ &= \frac{1}{y_t^j!} \exp \left[ y_t^j \log(\theta_t) - \theta_t \right] \end{aligned}$$

Vimos na Seção 2.1.1 que, uma vez que a distribuição é pertencente à família exponencial, torna-se simples encontrar a conjugada para seu parâmetro. Nessa mesma seção foi mostrado que a distribuição conjugada para o parâmetro  $\theta_t$  é uma Gama.

Alguns métodos de inferência para a classe dos MLDG como os propostos por [Smith \(1979\)](#), [Harvey and Fernandes \(1989\)](#) e [Gamerman et al. \(2013\)](#), atribuem uma evolução na média a cada tempo  $t$ , utilizando propriedades de conjugação. Desta forma, não é necessário que se faça uso de nenhuma forma de aproximação, uma vez que tanto a distribuição a *posteriori* para o parâmetro  $\theta_t$  como a distribuição preditiva, terão formas analíticas fechadas e conhecidas. Contudo, os modelos propostos por tais autores, embora permitam a presença de componentes sazonais e de tendência, não permitem que estas sejam estocásticas.

Por outro lado, [West et al. \(1985\)](#) propõe que a média não seja modelada diretamente e sim através de uma função de ligação que relaciona a tais componentes de tendência e sazonalidade. No caso da Poisson especificada anteriormente, é sabido que a média é igual a  $\theta_t$ . Suponha que a função de ligação escolhida, seja a canônica, ou seja,  $\log(\theta_t)$ .

$$g(\theta_t) = \log(\theta_t) = F_t^\theta \theta_t \quad (3.11)$$

onde  $F_t$  é o vetor de planejamento, que pode conter ou não covariáveis, e  $\theta_t$  é o vetor

de estados que, conforme visto na equação 3.5, estão relacionados às componentes de tendência e sazonalidade. Observe que uma vez atribuída uma distribuição a *priori* para  $\theta_t$ , teremos uma distribuição a *priori* induzida para  $\theta_t$ . Contudo, a adoção de distribuição a *priori* conjugada é conveniente, entre outros motivos, pela decorrente existência de forma analítica fechada para a distribuição preditiva de futuras observações.. Dessa forma, a solução encontrada por West et al. (1985) foi igualar os momentos dessa distribuição a *priori* induzida pelos estados aos momentos da *priori* conjugada. Observe inclusive que os autores não atribuem uma distribuição específica para  $\theta_t$ , sendo esta especificada somente em termos de primeiro e segundo momentos. Uma vez definidos os momentos da *priori* conjugada para  $\theta_t$ , é possível incorporar a nova informação obtida para que a *posteriori* para tal parâmetro seja definida. Veja que  $\theta_t$  é um escalar, enquanto  $\theta_t$  é um vetor. Assim, a atualização dos estados a partir da *posteriori* de  $\theta_t$  não é trivial, sendo necessário o uso de um método de aproximação chamado Linear Bayes. Dessa forma, a distribuição a *posteriori* para  $\theta_t$  é obtida. É importante lembrar que, embora sejam utilizados métodos de aproximação, o modelo permite que sejam atribuídas componentes de tendência e sazonalidade estocasticamente e que a média seja decomposta em termos de cada uma delas.

No método que proporemos no capítulo seguinte, sugerimos que seja atribuída uma distribuição Normal para o vetor de estados e que a *priori* induzida, por essa especificação, para o vetor canônico e sua *priori* conjugada sejam compatibilizadas a partir do uso de conceitos de Geometria da Informação.

Ao longo desta seção serão apresentados os modelos citados nesta breve introdução, de forma mais aprofundada.

### 3.2.1 Conjugate Updating

Modelos lineares dinâmicos generalizados (MLDG) são uma extensão para os modelos lineares dinâmicos e para os modelos lineares generalizados. Os MLDG:

- (a) Acomodam respostas pertencentes à família exponencial;
- (b) Não pressupõem, necessariamente, efeitos fixos associados a covariáveis no tempo.

West et al. (1985) propuseram uma extensão do tradicional modelo linear generalizado relacionando  $y_t$  a um preditor dinâmico  $F_t^0 x_t$ , através de uma função de ligação monótona e diferenciável  $g(\cdot)$ :

$$y_t = g(x_t) = F_t^0 x_t \quad (3.12)$$

Admite-se, nesta classe de modelos, que  $Y_t$  pertence à família exponencial e que o parâmetro natural  $x_t$  tem *priori* conjugada da forma  $PC(x_t; \theta_t)$  para algum  $\theta_t$  e  $x_t$ , podendo ser escrita da seguinte forma

$$p(x_t | D_{t-1}) = c(x_t; \theta_t) \exp(x_t' \theta_t - t a(x_t)):$$

O problema é que impor (3.12) traria muitas restrições para a *priori*  $x_t$ , então, para contornar o problema, foi sugerido que a relação entre  $g(x_t)$  e  $x_t$  seria usada somente como guia para formar a *priori*  $x_t$ , sendo notada por  $g(x_t) = y_t$

As distribuições à *priori* e *posteriori* não serão normalmente distribuídas, contudo, por analogia ao MLD, os primeiro e segundo momentos do vetor de estados  $x_t$  serão dados por

$$E(x_t | D_{t-1}) = (m_{t-1}; C_{t-1}); \quad (3.13)$$

$$Var(x_t | D_{t-1}) = (a_t; R_t); \quad (3.14)$$

onde,

$$a_t = G_t m_{t-1} \text{ e } R_t = B_t G_t C_{t-1} G_t^0 B_t$$

com matriz de evolução  $G_t$  e matriz de desconto  $B_t$  conhecidas.

De (3.12),

$$f_t = E(y_t | D_{t-1}) = F_t^0 a_t;$$

$$q_t = Var(y_t | D_{t-1}) = F_t^0 R_t F_t;$$

Até aqui, a *priori* natural para o parâmetro  $x_t$  é somente parcialmente especificada,

tendo a forma

$$p(\theta_t | D_{t-1}) = PC(\theta_t; \theta_{t-1})$$

sem nenhuma restrição a  $\theta_t$  e  $\theta_{t-1}$ . Esses valores serão escolhidos de acordo com a relação  $g(\theta_t) = \theta_t$ , de modo que fixando os dois primeiros momentos de  $g(\theta_t)$ , será possível determinar  $\theta_t$  e  $\theta_{t-1}$  através de  $f_t$  e  $q_t$ .

Dessa forma, fica claro que a preditiva tem forma

$$p(Y_t | D_{t-1}) = \frac{c(\theta_t; \theta_{t-1})}{c(\theta_{t+1}; Y_t; \theta_{t+1})} \quad (3.15)$$

Enquanto a *posteriori* para  $(\theta_t | D_t)$  tem forma conjugada

$$PC(\theta_{t+1}; Y_t; \theta_{t+1})$$

Uma análise Bayesiana completa requer também a distribuição a *posteriori* para  $\theta_t$ , mas esta não está disponível porque a *priori*  $(\theta_t | D_{t-1})$  é apenas parcialmente especificada e o modelo não fornece uma verossimilhança para  $\theta_t$ . Entretanto, conforme visto anteriormente o modelo não precisa da especificação completa da distribuição para passar para o tempo  $t+1$  e sim dos primeiros e segundo momentos. Estes podem ser encontrados da seguinte forma:

$$m_t = E(E(\theta_t | D_t)); \quad (3.16)$$

$$C_t = Var(E(\theta_t | D_t)) + E(Var(\theta_t | D_t)); \quad (3.17)$$

Como no caso normal,  $(\theta_t | D_t)$  é condicionalmente independente de  $I_t = (Y_t; F_t)$ , sabe-se que os momentos (3.16) e (3.17) são de  $(\theta_t | D_{t-1})$ . Na maioria das vezes, esses momentos não serão conhecidos e as únicas informações disponíveis serão os momentos da conjunta  $(g(\theta_t); \theta_t | D_{t-1})$ ,



$$g_t = \frac{R_t F_t (R_t F_t)^0}{R_t + (R_t F_t)^0} \quad (3.17)$$

onde a matriz de covariância completa é singular. Assim, uma abordagem alternativa é necessária para que a informação em  $I_t$  seja incorporada de volta a  $t$ .

O método Linear Bayes pode ser usado para fornecer o retorno da informação em  $I_t$  para  $t$ . A densidade  $p(y_t | D_{t-1})$  é a distribuição preditiva desconhecida de  $(y_t | t)$ . A média é o preditor no sentido de minimizar traço do risco quadrático  $[A_t(\mathbf{d})]$  com respeito a  $\mathbf{d}$  onde

$$A_t(\mathbf{d}) = E[(y_t - \mathbf{d})(y_t - \mathbf{d})^j | D_{t-1}]:$$

A matriz de covariância da distribuição é o valor de  $A_t(\mathbf{d})$  na média. Dado que a média é desconhecida, adota-se uma função linear  $g_t$  como preditor de  $y_t$ . Especificamente,  $\mathbf{d}$  deve ser escolhido de forma que  $\mathbf{d} = \mathbf{d}_0 + \mathbf{d}_1 g_t$  para algum  $\mathbf{d}_0, \mathbf{d}_1$  e  $\mathbf{d}$  minimize a soma de variâncias dada por

$$r_t(\mathbf{d}) = \text{traço } E(A_t(\mathbf{d}) | D_{t-1}); \quad (3.18)$$

com esperança em relação a  $p(y_t | D_{t-1})$ .

Nesse modelo, os momentos da conjunta são suficientes para determinar o preditor requerido. Minimizando diretamente  $r_t(\mathbf{d})$  com respeito a  $\mathbf{d}_0$  e  $\mathbf{d}_1$ , mostra-se que o ponto de mínimo é obtido em  $\mathbf{d} = \hat{\mathbf{a}}_t$ , onde

$$\hat{\mathbf{a}}_t = \mathbf{a}_t + \frac{R_t F_t (g_t - f_t)}{q_t}; \quad (3.19)$$

Assim, o resultado de  $E(A_t(\mathbf{d}) | D_{t-1})$  no ponto de mínimo é

$$\hat{R}_t = R_t - \frac{R_t F_t (R_t F_t)^0}{q_t}; \quad (3.20)$$

Os valores  $\hat{\mathbf{a}}_t$  e  $\hat{R}_t$  fornecem um preditor linear ótimo de  $p(y_t | D_{t-1})$ . O retorno da informação  $I_t$  pode ser completado agora, substituindo a média e covariância condicionais

em (3.16) (3.17) por  $\hat{\mathbf{a}}_t$  e  $\hat{R}_t$  para encontrar o preditor esperado e risco relativo dados por

$$m_t = a_t + R_t(g_t - f_t) = q_t; \quad (3.21)$$

$$C_t = R_t \frac{R_t F_t (R_t F_t)^0 (1 - \rho_t = q_t)}{q_t}; \quad (3.22)$$

onde  $g_t = E(g(\cdot | D_t))$  e  $p_t = Var(g(\cdot | D_t))$  podem ser calculados a partir da *posteriori*  $(\cdot | D_t)$ .

### 3.2.2 Modelos Dinâmicos de Nível Local

Smith (1979) estende a metodologia de Harrison and Stevens (1976) no sentido de considerar a modelagem de dados com distribuição diferente da gaussiana. A ideia proposta é pensar na evolução dos estados como um processo de decisão bayesiano de forma que as decisões com respeito a uma certa função perda permaneçam constantes enquanto não houver novas observações e a perda esperada associada a essas decisões aumente durante esse intervalo. Verificou-se que a função (de calibre  $b$ ) que satisfaz esses critérios é tal que só assume dois valores, ou seja:

$$L_b(d) = \begin{cases} 0; & j < dj - b; & b > 0 \\ 1; & \text{caso contrário;} \end{cases} \quad (3.23)$$

Smith (1979) demonstrou que a distribuição a *priori* de hoje (tempo  $t$ ) para os estados é proporcional a *posteriori* do tempo anterior (tempo  $t - 1$ ) elevada a uma determinada potência  $!$ , tais critérios são satisfeitos. Assim,

$$f(\cdot | D_t) \propto f(\cdot | D_{t-1}) g_t^!; \text{ onde } 0 < ! < 1. \quad (3.24)$$

Dessa forma, garante-se que a média da distribuição seja mantida e que a incerteza aumente a cada tempo  $t$ .

$$\text{Seja } f_{t-1}(\cdot) = \frac{N(m_{t-1}, C_{t-1})}{f_{t-1}(\cdot)};$$

Desse modo,

$$f_{t-1}(\cdot) \propto \exp \left( -\frac{(y_t - m_{t-1})^2}{2C_{t-1}} \right) ; \quad (3.25)$$

De (3.24),

$$\begin{aligned} y_t | D_{t-1} &\sim c_t f_{t-1}(\cdot) ; \\ &\propto f_{t-1}(\cdot) ; \end{aligned} \quad (3.26)$$

De (3.25) e (3.26),

$$\begin{aligned} f(y_t | D_{t-1}) &\propto \exp \left( -\frac{(y_t - m_{t-1})^2}{2C_{t-1}} \right) ; \\ &\propto \exp \left( -\frac{(y_t^2 - 2m_{t-1}y_t)}{2C_{t-1}} \right) ; \\ &\propto \exp \left( -\frac{1}{2} \frac{y_t^2}{C_{t-1}} + \frac{m_{t-1}}{C_{t-1}} y_t \right) ; \end{aligned}$$

Assim,

$$\begin{aligned} a_t &= E(y_t | D_{t-1}) = m_{t-1} = E(y_t | D_{t-1}); \\ R_t &= V(y_t | D_{t-1}) = \frac{1}{I} C_{t-1} = \frac{1}{I} V(y_t | D_{t-1}); \end{aligned}$$

Verifica-se, portanto, que a média da *posteriori* foi equivalente à da *priori* enquanto a variância foi inflacionada por um fator  $I=I$ !

Uma vantagem desta forma de atualização proposta por [Smith \(1979\)](#) é a capacidade de se poder compatibilizar e depois atualizar a *priori* via Teorema de Bayes, ou atualizar e então compatibilizar *posteriors*, preservando-se o mesmo resultado.. A distribuição a *priori* utilizada para formular o modelo de nível local com essas características foi a conjugada natural da família exponencial. Foi observado que seus procedimentos deram origem a previsões do tipo média móvel exponencial ponderada. [Harvey and Fernandes \(1989\)](#), também interessados em modelagem de dados qualitativos e de contagem, se baseiam neste modelo. De forma análoga, introduzem hiperparâmetro que atua da mesma forma que um fator de desconto. A diferença se dá na forma de estimação desse hiper-

parâmetro: enquanto o primeiro não menciona um método específico, o segundo sugere o uso do estimador de máxima verossimilhança. [Smith and Miller \(1986\)](#) generalizam a metodologia de [Smith \(1979\)](#); consideram a equação de evolução exata, de forma que não sejam necessárias aproximações para a obtenção da distribuição preditiva e dos estados. [Gamerman et al. \(2013\)](#) se baseiam em [Smith and Miller \(1986\)](#), introduzem a classe dos Modelos Dinâmicos da Família Gama e apresentam uma forma de suavização para a componente do nível do modelo.

É importante salientar que os modelos propostos por tais autores, embora permitam a presença de componentes sazonais e de tendência, não permitem que estas sejam estocásticas. Os autores argumentem que esta não é uma grave limitação. Outro ponto que deve ser mencionado é que a modelagem proposta permite que a distribuição dos dados sejam tanto uniparamétricas, como  $k$  paramétricas, sendo esta uma vantagem do método. Será apresentada nesta seção uma breve descrição dos modelos mencionados. Mais detalhes podem ser encontrados em [Smith \(1979\)](#), [Smith and Miller \(1986\)](#), [Harvey and Fernandes \(1989\)](#) e [Gamerman et al. \(2013\)](#).

Seja o Modelo Dinâmico da Família Gama (MDFG) como definido em [Gamerman et al. \(2013\)](#):

$$p(y_t | y_{t-1}) = a(y_{t-1}) \frac{b(y_t)}{b(y_{t-1})} \exp \{ -c(y_t) \} g \quad (3.27)$$

Onde,  $y_t \geq H(\cdot) < \infty$  e  $p(y_t | y_{t-1}) = 0$ , caso contrário.

Variáveis explicativas, podem ser inseridas em vetor de covariáveis  $x_t$ . Este vetor se relacionará com  $y_t$  através da função de ligação, de forma que  $y_t = g(x_t)$ , onde se refere aos coeficientes da regressão e  $\beta$  é um parâmetro referente ao nível do modelo dinâmico. Tal parâmetro é definido através da equação de evolução  $y_t = \beta^{-1} y_{t-1} + \epsilon_t$ , onde  $\epsilon_t \sim \text{Beta}(\beta a_{t-1}; (1 - \beta) a_{t-1})$ , isto é,

$$\beta \frac{y_t}{(1 - \beta) y_t} \sim \text{Beta}(\beta a_{t-1}; (1 - \beta) a_{t-1}) \quad (3.28)$$

As informações iniciais acerca do nível são definidas através da *priori*  $y_0 | D_0 \sim \text{Gama}(a_0; b_0)$ .

Resultados referentes ao modelo apresentado em 3.27, são apresentados a seguir.

A *priori* para  ${}_t jD_{t-1}$  é uma distribuição  $Gama(a_{tjt-1}; b_{tjt-1})$ , onde :

$$a_{tjt-1} = \beta a_{t-1}; \quad (3.29)$$

$$b_{tjt-1} = \beta b_{t-1}; \quad (3.30)$$

$$0 < \beta < 1;$$

Dessa forma,

$$E({}_t jD_{t-1}) = \beta a_{t-1} / \beta b_{t-1} = a_{t-1} / b_{t-1}; \quad (3.31)$$

$$Var({}_t jD_{t-1}) = \beta a_{t-1} / (\beta b_{t-1})^2 = \beta^{-1} Var({}_{t-1} jD_{t-1}); \quad (3.32)$$

Nota-se, portanto, que a média se mantém e a variância é inflacionada por um fator  $1/\beta$  a cada tempo  $t$ . Pode-se observar que  $\beta$  atua conforme um fator de desconto.

Sendo  $\pi_t = \pi_t g(x_t^d)$ , a distribuição  ${}_t jD_{t-1}$  é uma Gama com parâmetros  $a_{tjt-1}$  e  $b_{tjt-1}$ , onde tais parâmetros podem ser definidos como:

$$a_{tjt-1} = \beta a_{t-1}; \quad (3.33)$$

$$b_{tjt-1} = \beta b_{t-1} g(x_t^d); \quad (3.34)$$

Dada uma nova observação a distribuição a *posteriori*,  ${}_t jD_t$ , é uma  $Gama(a_t; b_t)$  com parâmetros especificados por:

$$a_t = a_{tjt-1} + b(y_{t-1}); \quad (3.35)$$

$$b_t = b_{tjt-1} + c(y_{t-1}); \quad (3.36)$$

Desta maneira, a distribuição a *posteriori* para  ${}_t jD_t$  também será uma Gama com parâmetros  $a_t$  e  $b_t$ , com

$$a_t = a_{tjt-1} + b(y_{t-1}); \quad (3.37)$$

$$b_t = b_{tjt-1} + c(y_{t-1}) g(x_t^d); \quad (3.38)$$

A distribuição preditiva um passo à frente, de forma que:

$$\begin{aligned}
 p(y_{t+1}|D_t) &= \int_0^Z p(y_{t+1}|y_t) p(y_t|D_t) p(d_{t+1}|y_t) \\
 &= \frac{a_{t+1} + b(y_t; \cdot) c(y_t; \cdot) (b_{t+1})^{a_{t+1}}}{a_{t+1} c(y_t; \cdot) + b_{t+1}} \quad (3.39)
 \end{aligned}$$

# Capítulo 4

## MLDG via Geometria da Informação

### 4.1 Introdução

Até o momento foram apresentados dois métodos de estimação para a classe de modelos lineares dinâmicos generalizados: *Conjugate Updating* e Modelos de Nível Local. O primeiro utiliza Linear Bayes para obter as estimativas do modelo, uma vez que somente as distribuições do parâmetro canônico e da preditiva têm forma analítica fechada. A evolução dos estados é descrita somente pelos primeiro e segundo momentos. Uma das dificuldades encontradas é que há duas distribuições a *priori* para o parâmetro natural da família exponencial: uma induzida pelo vetor de estados e outra vinda por conjugação da família exponencial. A solução encontrada por West, Harrison e Migon, foi igualar os momentos de primeira e segunda ordem de tais *prioris*. Abre-se mão de ter forma analítica fechada para as distribuições associadas ao vetor de estados, mas preserva-se a conveniência das formas analíticas fechadas para a distribuição a posteriori do parâmetro canônico e para a distribuição preditiva. Um dos objetivos do presente trabalho é propor uma forma alternativa de compatibilização de *prioris*. Para isso, são utilizados conceitos de Geometria da Informação tais como teorema da Projeção e divergência de Kullback-Leibler. No método proposto, assume-se evolução gaussiana para os estados. É importante ressaltar que, diferentemente dos modelos apresentados em 3.2.2, esta metodologia permite que componentes de sazonalidade e de tendência possam ser incorporadas estocasticamente e que as regressores tenham parâmetros associados variando no tempo.

## 4.2 Método proposto

O método a ser apresentado é muito similar ao proposto por West *et. al* (1985) e utiliza grande parte de sua estrutura. De forma análoga, é um método de inferência proposto para a família exponencial uniparamétrica. A diferença aqui se dá na forma com que duas *prioris* para o parâmetro canônico são compatibilizadas: a *priori* conjugada para tal parâmetro e a priori implicada pela suposição de normalidade dos parâmetros de estado que controlam o preditor linear do modelo.

Sabe-se que,

$$t = g(\eta) = F_t^\theta \eta; \quad (4.1)$$

Onde,  $F_t$  é uma matriz de forma de tamanho  $p \times n$ ,  $\eta$  é o vetor de estados de tamanho  $p - 1$ ,  $\eta$  é o parâmetro da distribuição uniparamétrica de interesse e  $g(\eta)$  é a função de ligação escolhida.

Observe que uma vez atribuída uma priori para  $\eta$ , haverá duas *prioris* para a mesma quantidade  $\eta$ : de um lado a *priori* implicada pela conjugação na família exponencial e de outro, a *priori* induzida pelo vetor de estados.

A metodologia proposta no presente trabalho funciona da seguinte forma: uma vez atribuída uma distribuição a *priori* para o vetor de estados, obtenha a *priori* para o parâmetro natural induzida por tal vetor. Projete-a no espaço da distribuição conjugada da família exponencial. Esta será a *priori* utilizada no prosseguimento do método.

Como esta é conjugada da família exponencial, sua *posteriori* pode ser facilmente encontrada. Contudo, observe que agora temos duas *posteriors* para o preditor linear: uma induzida pela conjugação do parâmetro natural e outra Normal (já que o preditor é uma função linear dos estados que, sob normalidade, têm distribuição a *posteriori* normal). A solução adotada, para lidar com essa duplicidade, foi projetar a distribuição a *posteriori* induzida pela conjugação do parâmetro natural no espaço da distribuição Normal. Dessa forma, gostaríamos de encontrar a Normal que melhor aproxima tal distribuição. Uma vez que tal compatibilização seja feita, utilizamos o método Linear Bayes para obter os momentos do vetor de estados. Observe que a normalidade é mantida.



O uso do método Linear Bayes se faz necessário porque, enquanto o preditor linear é um escalar, o vetor de estados é p-dimensional. Desta forma, escolhemos o esse método para que tal expansão pudesse ser feita.

Um resumo da metodologia proposta é apresentado a seguir.

### Modelo

$$\begin{aligned} & \mathcal{Y} \\ & \sim y_{t|t-1} \sim \text{FamExp}(y_t) \\ & \mathcal{D} \\ & \sim y_{t|D_{t-1}} \sim C \text{FamExp}(y_t) \end{aligned}$$

### Função de Ligação

$$g(y_t) = F_t^0 y_t = y_t$$

onde  $g$  é uma função contínua, monótona e duas vezes diferenciável.

### Evolução do Vetor de Estados

$$x_t = G_t x_{t-1} + \epsilon_t; \quad \epsilon_t \sim N(0; W);$$

o que implica em

$$p(y_t | D_{t-1}) = p_{y_t | D_{t-1}}(g^{-1}(y_t)) \frac{\partial y_t}{\partial g^{-1}(y_t)}$$

### Distribuições a *priori* resultantes

$$\begin{aligned} & y_{t|D_{t-1}} \sim C \text{FamExp}(y_t) \\ p(y_t | D_{t-1}) & = p_{y_t | D_{t-1}}(g^{-1}(y_t)) \frac{\partial y_t}{\partial g^{-1}(y_t)} \end{aligned}$$

### Metodologia Proposta

Seja  $q(x)$  a distribuição a *priori* conjugada e seja  $p(x)$  a *priori* induzida pelo vetor de estados. Obtém-se parâmetros tais que

$$= \text{argmin} D_{KL}[p : q] \tag{4.2}$$

Uma vez que esse vetor paramétrico seja definido,

$${}_{tj}D_t = C \text{ FamExp}(\theta_{t-1}): \quad (4.3)$$

Uma vez que obtemos a *priori* compatibilizada para  $\theta_t$ , podemos incorporar facilmente novas informações através da conjugação. Observe que agora, duas distribuições a *posteriori* para o preditor linear devem ser combinadas:

$$\begin{aligned} {}_{tj}D_t &= p_{{}_{tj}D_t}(g(\theta_t) | \frac{\partial}{\partial \theta_t} g(\theta_t)) \\ {}_{tj}D_t &= N(h_t; p_t) \end{aligned}$$

Em seguida, projetamos a distribuição a *posteriori* obtida através da conjugação no espaço da distribuição Normal, encontrando seus momentos. Observe que  $\theta_t$  é um escalar, para cada tempo  $t$ , enquanto  $\theta_t$  é um vetor paramétrico  $p$  dimensional. Uma vez que a distribuição a *posteriori* para  $\theta_t$  é obtida, gostaríamos de obter a *posteriori* para  $\theta_t$ , levando em consideração a diferença de dimensões para essas quantidades. Usamos o método Linear Bayes para encontrar os momentos de  $({}_{tj}D_t)$ .

Nas subseções 4.2.1 e 4.2.2, detalha-se a metodologia descrita, considerando-se MLDGs para respostas Poisson e Binomial, respectivamente. Na subseção 4.2.3, aplica-se o método a dados artificialmente gerados, seguindo um MLDG para resposta Poisson.

### 4.2.1 Modelo Linear Dinâmico Generalizado Poisson

Seja  $(y_{tj} | \theta_t) \sim \text{Poisson}(\theta_t)$ :

$$p(y_{tj} | \theta_t) = \frac{\theta_t^{y_{tj}} e^{-\theta_t}}{y_{tj}!}: \quad (4.4)$$

Esta densidade pode ser reescrita na forma da família exponencial da seguinte forma:

$$p(y|t) = \frac{1}{y!} \exp(-y \log(t)) \quad t \in \mathcal{G} \quad (4.5)$$

Assim temos que a precisão é igual a 1, o preditor linear  $\eta_t$  é igual ao  $\log(t)$ , que é o parâmetro canônico, neste caso. Tem-se, ainda,  $b(\eta_t) = \eta_t = e^{-\eta_t}$ . Assim, a média e variância da distribuição são dadas por  $\dot{b}(\eta_t) = \eta_t = e^{-\eta_t}$  e  $\ddot{b}(\eta_t) = -\eta_t = -e^{-\eta_t}$ , respectivamente.

No caso da distribuição Poisson, sabemos que a *priori* conjugada para o parâmetro  $t$  é a Gama( $\alpha; \lambda$ ) e portanto, a *priori* implicada para o preditor linear  $\eta_t$ , adotando-se a especificação conjugada, é a Log-Gama.

$$(\eta_t | D_{t-1}) \sim PC(\eta_t; \lambda) \sim Gama(\alpha; \lambda) \quad (4.6)$$

Onde,

$$\eta_t = \log(t) = F_t^\eta \lambda \quad (4.7)$$

Por outro lado, a informação inicial  $(\eta_0 | D_0)$  tem distribuição Normal com parâmetros  $(m_0; C_0)$  e usando  $\eta_t = G_t \eta_0 + \lambda$ , tem-se que

$$(\eta_t | D_{t-1}) \sim N(a_t; R_t) \text{ e } (\eta_t | D_{t-1}) \sim N(f_t; q_t) \quad (4.8)$$

em que  $f_t = F_t^\eta a_t$  e  $q_t = F_t^\eta R_t F_t^\eta$ .

Dessa forma, desejamos projetar a *priori* induzida pelo vetor de estados 4.8 no espaço da distribuição conjugada. Sabendo que a *priori* induzida tem distribuição Log-Normal, gostaríamos de projetá-la no espaço da distribuição Gama com o intuito de preservar a conveniência de se trabalhar com a conjugada.

Temos portanto que

$$(\eta_t | D_{t-1}) \sim \text{Log-Normal}(f_t; q_t) \text{ e desejamos obter } (\eta_t | D_{t-1}) \sim Gama(\alpha; \lambda)$$

Como visto na Seção (2.2), os parâmetros da Gama deduzida da Log-Normal são:

$$\tau_t = \frac{1}{q_t} \quad (4.9)$$

$$\tau_t = q_t^{-1} \exp \int f_t - 0.5 q_t g \quad (4.10)$$

Uma vez encontrada a *priori* compatibilizada para  $\tau_t$ , pertencente à família de *prioris* conjugadas, sua *posteriori* é facilmente obtida por conjugação. Agora temos que

$$(\tau_t | D_t) \sim \text{Log-Gama}(\tau_t; \tau_t) \text{ e desejamos obter } (\tau_t | D_t) \sim \text{Normal}(\tau_t; \tau_t^2):$$

Projeta-se então a Log-Gama no espaço da distribuição Normal e aplica-se o método Linear Bayes para encontrar os momentos de  $(\tau_t | D_t)$ . Tal distribuição terá como parâmetros  $(m_t; C_t)$ . Sendo estes:

$$\begin{aligned} m_t &= m_{t-1} + R_t F_t (g_t - f_t) = q_t; \\ C_t &= R_t - R_t F_t (R_t F_t)^0 (1 - p_t = q_t) = q_t; \end{aligned}$$

onde  $g_t = E(f(\tau_t) | D_t)$  e  $p_t = \text{Var}(f(\tau_t) | D_t)$  podem ser calculados a partir da *posteriori*  $(\tau_t | D_t)$  já compatibilizada, onde  $f(\tau_t) = \log(\tau_t)$ .

Observe que no modelo Poisson via *Conjugate Updating* os parâmetros  $\tau_t$  e  $q_t$  são dados por:

$$\begin{aligned} \tau_t &= \frac{1}{q_t} \\ \tau_t &= q_t^{-1} \exp \int f_t - f_t g \end{aligned} \quad (4.11)$$

Dessa forma, nota-se que os parâmetros são bem similares nos dois métodos, havendo diferença somente no parâmetro  $\tau_t$ .

## 4.2.2 Modelo Linear Dinâmico Generalizado Bernoulli/Binomial

Seja  $(y_t | p_t)$  Bernoulli  $(p_t)$ :

$$p(y_t | p_t) = p_t^{y_t} (1 - p_t)^{(1 - y_t)}; \quad (4.12)$$

Pode-se escrever a função de probabilidade acima na família exponencial da seguinte forma:

$$p(Y_t | p_t) = \exp \left[ y_t \frac{\log(p_t)}{\log(1 - p_t)} + \log(1 - p_t) \right]; \quad (4.13)$$

Assim temos que a precisão  $\eta$  é igual a 1, o preditor linear  $\mu_t$  é igual ao  $\log \frac{p_t}{1 - p_t}$ , sendo este o parâmetro natural ou canônico, e  $b(\eta) = \log(1 - p_t)$ . Assim, a média e variância da distribuição são dadas por  $\hat{\mu}(\eta) = p_t = \frac{1}{1 + e^{-\eta}}$  e  $\hat{\sigma}(\eta) = p_t(1 - p_t) = \frac{e^{-\eta}}{(1 + e^{-\eta})^2}$ , respectivamente.

No caso da distribuição Bernoulli, sabemos que a *a priori* conjugada para o parâmetro  $p_t$  é a Beta  $(\alpha; \beta)$  e portanto, a *priori* implicada para o preditor linear  $\mu_t$  é a Logito-Beta.

$$(p_t | D_{t-1}) \sim PC(\mu_t; \eta) \sim \text{Beta}(\alpha; \beta); \quad (4.14)$$

de forma que,

$$\mu_t = \text{logito}(p_t) = F_t^\eta \eta \sim \text{Logito} \sim \text{Beta}(\alpha; \beta); \quad (4.15)$$

Por outro lado a informação inicial  $(\mu_0 | D_0)$  tem distribuição Normal com parâmetros  $(m_0; C_0)$ , tem-se que

$$(\mu_t | D_{t-1}) \sim N(a_t; R_t) \text{ e } (\mu_t | D_{t-1}) \sim N(f_t; q_t) \quad (4.16)$$

em que  $f_t = F_t^\eta a_t$  e  $q_t = F_t^\eta R_t F_t$ .

Conforme mencionado anteriormente, a ideia é projetar a *priori* induzida pelo vetor de estados no espaço da distribuição conjugada. Sabendo que a *priori* induzida é Logito-

Normal, deseja-se projetá-la no espaço da distribuição Beta.

$$(p_t | D_{t-1}) \text{ Logito } \text{Normal}(f_t; q_t) \text{ e buscamos } (p_t | D_{t-1}) \text{ Beta}(a_t; b_t):$$

Como visto na Seção (2.2), os parâmetros da Beta compatibilizada não têm solução analítica, portanto foram utilizados métodos computacionais como Newton-Raphson e Monte Carlo para que esses pudessem ser obtidos. A cada tempo  $t$ , esses métodos devem ser utilizados para calculá-los.

Uma vez encontrada a *priori* compatibilizada para  $p_t$  na família de *prioris* conjugadas, sua *posteriori* é facilmente obtida via Teorema de Bayes.

Agora temos que

$$(p_t | D_t) \text{ Logito } \text{Beta}(a_t; b_t) \text{ e desejamos obter } (p_t | D_t) \text{ Normal}(m_t; C_t):$$

Projeta-se então a distribuição Logito-Beta no espaço da distribuição Normal e aplica-se o método Linear Bayes para encontrar os momentos de  $(p_t | D_t)$ . Tal distribuição terá como parâmetros  $(m_t; C_t)$ . Sendo estes:

$$m_t = m_{t-1} + R_t F_t (g_t - f_t) = q_t;$$

$$C_t = R_t - R_t F_t (R_t F_t)^{-1} (1 - p_t = q_t) = q_t;$$

onde  $g_t = E(f(p_t) | D_t)$  e  $p_t = Var(f(p_t) | D_t)$  podem ser calculados a partir da *posteriori*  $(p_t | D_t)$  já compatibilizada, onde  $f(p_t) = \text{logito}(p_t)$ .

### 4.2.3 Estudo de Simulação

Um estudo simulado é apresentado nesta seção com o intuito de verificar o comportamento da metodologia proposta no que concerne a estimativas pontuais e predição a um passo. Visando avaliar a eficácia do método quando comparado a outros que fazem inferência para esta mesma classe de modelos, também será apresentado o fator de Bayes comparando-os.

### 4.2.3.1 Modelo Poisson

Seja  $Y_{tj} \mid t \sim \text{Poisson}(\lambda_t)$  em que:

$$\lambda_t = \log(\lambda_t) = F_t^\theta \beta; \\ \beta \sim G_t \mid t \sim N(0; 0.001^2);$$

$F_t$  e  $G_t$  foram definidas da seguinte forma:

$$F_t = \begin{bmatrix} 1 & 0 & 1 & 0 \\ \delta t; t = 1; \dots; 80; \end{bmatrix}; \\ G_t = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos(w) & \text{sen}(w) \\ 0 & 0 & \text{sen}(w) & \cos(w) \end{bmatrix}; \text{ onde } w = \pi/2.$$

Foram geradas 50 réplicas de um modelo Poisson com essa estrutura. No processo inferencial foram utilizados fatores de desconto tanto para as componentes de tendência quanto de sazonalidade. Esses foram iguais a 0.95 e 0.99, respectivamente. A Figura 4.1 exibe os boxplots obtidos através de tais réplicas para a média a *posteriori* do parâmetro natural  $\lambda_t$  e seu valor verdadeiro (linha vermelha). Nota-se que o modelo é capaz de captar bem a estrutura do dado, de forma que o  $E(\lambda_j | D_t)$  ficou bem próximo ao valor verdadeiro em todo tempo  $t$ . É importante ressaltar que o custo computacional da metodologia proposta é muito baixo, com cada modelo levando em média 0.2 segundos para ser estimado em um computador com processador Intel Core i5-6500, com 4GB de RAM e sistema operacional de 64 Bits.

Verifica-se, através da Figura 4.2, que o método proposto é capaz de fazer previsões a um passo que acompanham a estrutura dos dados observados independentemente do tamanho da amostra. Nota-se que a grande maioria dos valores observados encontram-se dentro do intervalo de credibilidade obtido. Nota-se, inclusive, que tanto a sazonalidade quanto a tendência crescente, são captadas pelo modelo, resultado indicativo de um bom ajuste.

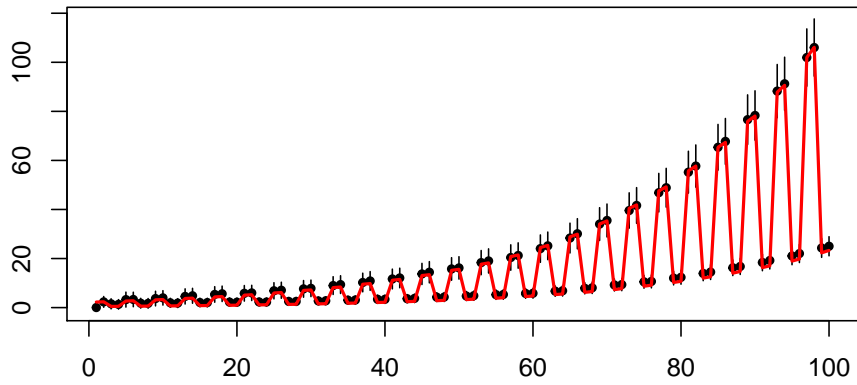


Figura 4.1: Boxplots resultantes das réplicas para a média *a posteriori* do parâmetro natural e seus valores verdadeiros representados pela linha vermelha

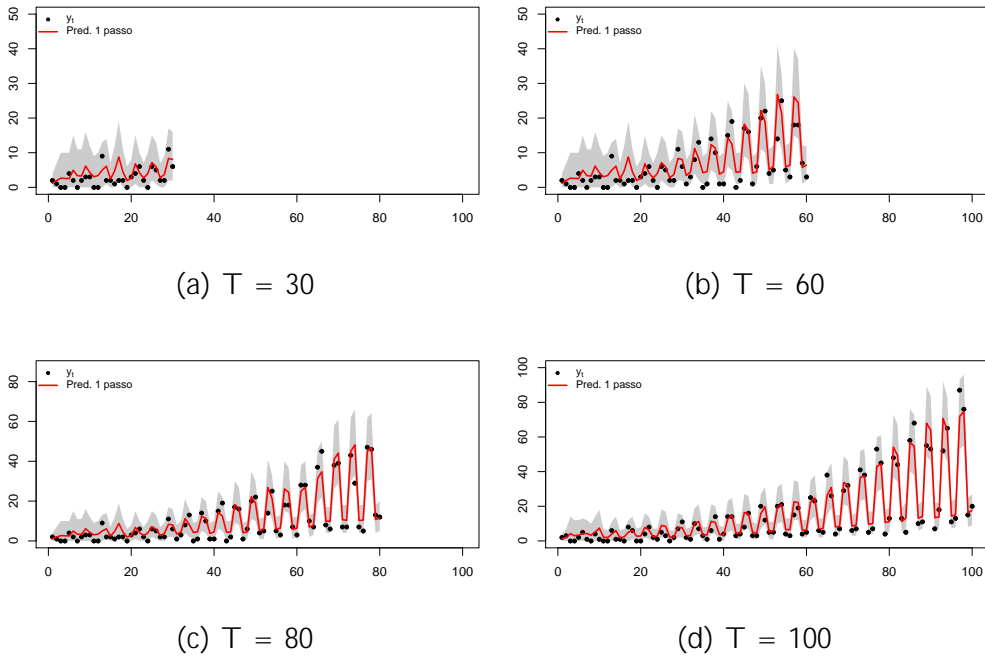


Figura 4.2: Média *a posteriori* da distribuição preditiva a um passo para diferentes tamanhos de amostra junto ao seu intervalo de credibilidade. As escalas dos gráficos diferem para facilitar a visualização.

### 4.3 Compatibilizando diferentes modelos

Observe que o modelo proposto por [Smith \(1979\)](#) não permite que sejam introduzidas componentes estruturais estocasticamente. Suponha que haja interesse em não só



utilizar a estrutura de inferência proposta em artigo, mas também em inserir componentes de tendência e sazonalidade estocasticamente no modelo. Uma combinação entre os métodos propostos por [Smith \(1979\)](#) e [West et al. \(1985\)](#) pode ser feita utilizando conceitos de geometria da informação. A descrição de tal metodologia será apresentada ao longo desta seção.

O funcionamento do método é similar ao explicitado anteriormente com o acréscimo de um novo passo referente à combinação das distribuições de probabilidade *a priori* e a *posteriori*, baseadas em ambas as especificações. De um lado, tem-se um modelo como o apresentado por [West et al. \(1985\)](#), completamente especificado. De outro, um modelo como o retratado por [Smith \(1979\)](#) e outros, fazendo uso da *priori* conjugada para o parâmetro canônico da família exponencial. Analogamente o procedimento apresentado na seção 4.2, projeta-se a *priori* para o parâmetro usual da família exponencial, induzida pelo vetor de estados, no espaço da distribuição conjugada da família exponencial. Uma vez que tal projeção tenha sido feita, torna-se simples encontrar a *priori* compatibilizada que leva em consideração as informações do modelo de nível local (*priori* conjugada) e do modelo estruturado (*priori* induzida pelo vetor de estados), utilizando-se, para tanto, o teorema 2.2.2.

Observe-se que os parâmetros da *priori* conjugada não serão obtidos através da compatibilização dos momentos, como é feito em [West et al. \(1985\)](#), o que faz com que seja necessário atribuir uma estrutura evolutiva para esses parâmetros que independe do vetor de estados. O método escolhido foi evoluir da *posteriori* no tempo  $t-1$  para *priori* no tempo  $t$ , preservando-se a média a *posteriori* e inflacionando-se a variância da *posteriori* por um fator multiplicativo  $1/\lambda$ ,  $0 < \lambda < 1$ .

Assim, os parâmetros da *priori* compatibilizada para o parâmetro canônico dependerão tanto dos parâmetros da *priori* conjugada quanto da *priori* induzida.

Como as duas distribuições a *priori* compatibilizadas pertencem ao espaço das *prioris* conjugadas para o parâmetro canônico (a primeira, projetada neste espaço utilizando-se o teorema da projeção e a segunda, já naturalmente pertencente a este espaço), também a *priori* resultante da aplicação do teorema 2.2.2, para o parâmetro canônico, será conjugada à família exponencial. Tal distribuição a *priori* terá seus parâmetros influenciados

tanto pelas componentes estruturais do modelo baseado em um preditor dinâmico com componentes estruturais, quanto pelo modelo dinâmico, mas de nível local, especificado somente com base na informação vinda da variável resposta. São fatores relevantes, na ponderação entre esses dois modelos, tanto o parâmetro  $S$ , que controla a combinação convexa apresentada no teorema 2.2.2, quanto as precisões associadas a cada uma das *prioris* a serem combinadas. Uma vez obtida a priori compatibilizada, pertencente à família de *prioris* conjugadas à família exponencial, sua *posteriori* pode ser facilmente encontrada. Contudo, observe que agora temos duas *posteriores* para o preditor linear: uma induzida pela conjugação do parâmetro natural da família exponencial e outra proveniente da normalidade dos estados. A solução encontrada foi projetar a distribuição induzida pela conjugação do parâmetro natural no espaço da distribuição Normal. Dessa forma, gostaríamos de encontrar a Normal que melhor aproxima tal distribuição. Uma vez que tal aproximação seja feita, utilizamos o método Linear Bayes para obter os momentos do vetor de estados, dada a diferença de dimensão entre o preditor linear (escalar) e tal vetor ( $\rho$ -dimensional).

Nas subseções 4.3.1 e 4.3.2, detalha-se a metodologia descrita, considerando-se MLDGs para respostas Poisson e Binomial, respectivamente. Na subseção 4.3.3, aplica-se o método a dados de vendas, seguindo um MLDG para resposta Poisson.

### 4.3.1 Modelo Linear Dinâmico Generalizado Poisson

Seja  $(y_t^j | t) \sim \text{Poisson}(\lambda_t)$ :

$$p(y_t^j | t) = \frac{\lambda_t^{y_t^j} e^{-\lambda_t}}{y_t^j!} \quad (4.17)$$

A distribuição a *priori* conjugada para o parâmetro  $\lambda_t$  é a Gama( $\alpha_t, \beta_t$ ) e, portanto, a *priori* implicada para o preditor linear é a Log-Gama. Isto é,

$$(\lambda_t | D_{t-1}) \sim \text{PC}(\lambda_t | t) \sim \text{Gama}(\lambda_t | t) \quad (4.18)$$

$$\lambda_t = \exp(\eta_t) = \exp(F_t^0 | t) \quad (4.19)$$

Observe que uma vez atribuída uma priori para  $\theta_t$ , haverá duas *prioris* para a mesma quantidade  $\theta_t$ : de um lado a distribuição a priori conjugada da família exponencial e de outro, a priori induzida pelo vetor de estados.

Assumindo que a informação inicial  $(\theta_0 | D_0)$  tem distribuição Normal com parâmetros  $(m_0; C_0)$ , tem-se que

$$(\theta_t | D_{t-1}) \sim N(a_t; R_t) \text{ e } (\theta_t | D_{t-1}) \sim N(f_t; q_t);$$

onde  $f_t = F_t^0 a_t$  e  $q_t = F_t^0 R_t F_t$ .

De forma que as duas prioris a serem compatibilizadas serão:

$$(\theta_t | D_{t-1}) \sim \text{Log Normal}(f_t; q_t) \text{ e } (\theta_t | D_{t-1}) \sim \text{Gama}(\alpha_t; \beta_t);$$

Como visto na Seção 2.2, os parâmetros da Gama já compatibilizada são:

$$a_t = s \frac{1}{q_t} + (1-s) \alpha_t \tag{4.20}$$

$$b_t = s(q_t^{-1} \exp(f_t - f_t - 0.5q_t g)) + (1-s) \beta_t \tag{4.21}$$

Observe aqui, que  $\alpha_t$  e  $\beta_t$  não são atualizados de acordo com a estrutura do vetor de estados, portanto é necessário que se crie uma estrutura de atualização para os mesmos. A forma escolhida foi a proposta por [Smith \(1979\)](#):

$$\alpha_t = \alpha_{t-1} + y_{t-1}$$

$$\beta_t = \beta_{t-1} + 1$$

Com essa estrutura de atualização, nota-se que, ao se passar da posteriori no tempo  $t-1$  à priori no tempo  $t$ , a média da distribuição Gama é mantida e que sua variância é inflacionada por um fator multiplicativo  $\frac{1}{\gamma}$  onde  $0 < \gamma < 1$ .

Nesta etapa do procedimento, tem-se duas distribuições Gama, a priori, para o parâmetro  $\theta_t$ , que devem ser compatibilizadas via teorema 2.2.2. Uma vez encontrada a priori compatibilizada para  $\theta_t$  pertencente à família de *prioris* conjugadas, sua posteriori

é facilmente obtida via Teorema de Bayes. Agora temos que

$$(y_t | D_t) \sim \text{Log-Gama}(f_t; \tau_t) \text{ e desejamos obter } (y_t | D_t) \sim \text{Normal}(f_t; \tau_t^2):$$

Projeta-se, então, a distribuição Log-Gama no espaço da distribuição Normal e aplica-se o método Linear Bayes para encontrar os momentos de  $(y_t | D_t)$ . Tal distribuição terá como parâmetros  $(m_t; C_t)$ . Sendo estes:

$$m_t = m_{t-1} + R_t F_t (g_t - f_t) = q_t;$$

$$C_t = R_t - R_t F_t (R_t F_t)^0 (1 - \rho_{t=q_t}) = q_t;$$

onde  $g_t = E(f_t | D_t)$  e  $p_t = \text{Var}(f_t | D_t)$  podem ser calculados a partir da *posteriori*  $(y_t | D_t)$  já compatibilizada, onde  $f_t = \log(y_t)$ .

### 4.3.2 Modelo Linear Dinâmico Generalizado Bernoulli/Binomial

Seja  $(y_t | p_t) \sim \text{Bernoulli}(p_t)$ :

$$p(y_t | p_t) = p_t^{y_t} (1 - p_t)^{(1 - y_t)}. \quad (4.22)$$

No caso da distribuição Bernoulli, sabemos que a distribuição a *priori* conjugada para o parâmetro  $p_t$  é a  $\text{Beta}(\alpha; \beta)$  e portanto, a *priori* implicada para o preditor linear  $\mu_t$  é a Logito-Beta.

$$(p_t | D_{t-1}) \sim \text{PC}(\alpha; \beta) \sim \text{Beta}(\alpha; \beta); \quad (4.23)$$

$$\mu_t = \text{logito}(p_t) = F_t^0 \mu; \quad (4.24)$$

Observe novamente que, uma vez atribuída uma *priori* para  $\mu_t$ , haverá duas *prioris* para a mesma quantidade  $\mu_t$ : de um lado a *priori* conjugada da família exponencial e de outro, a *priori* induzida pelo vetor de estados.

Assuma-se que a informação inicial  $(\mu_0 | D_0)$  tem distribuição Normal com parâmetros  $(m_0; C_0)$ . Assim,

$$(\mu_t | D_{t-1}) \sim N(a_t; R_t) \text{ e } (\mu_t | D_{t-1}) \sim N(f_t; q_t)$$

onde  $f_t = F_t^0 a_t$  e  $q_t = F_t^0 R_t F_t$ .

As duas *prioris* a serem compatibilizadas serão:

$$(p_t | D_{t-1}) \sim \text{Logito-Normal}(f_t; q_t) \text{ e } (p_t | D_{t-1}) \sim \text{Beta}(\alpha; \beta);$$

Como visto na Seção (2.2), os parâmetros da Beta compatibilizada não têm solução analítica, portanto métodos como Newton-Raphson e Monte Carlo são necessários para que tais parâmetros possam ser obtidos.

Novamente os parâmetros da *priori* naturalmente conjugada,  $\alpha$  e  $\beta$  não são atualizados de acordo com a estrutura do vetor de estados, portanto é necessário que se crie uma estrutura de atualização para os mesmos. Como já descrito, adota-se a proposta de

Smith (1979):

$$a_t = \alpha(a_{t-1} + y_{t-1})$$

$$b_t = \beta(b_{t-1} + n - y_{t-1}); \text{ sendo } n \text{ o número de ensaios de Bernoulli associados à distribuição Binomial}$$

Com essa estrutura de atualização, nota-se que a média da distribuição Gama é mantida e que sua variância é inflacionada por um fator multiplicativo  $\frac{1}{\alpha}$ ,  $0 < \alpha < 1$ .

Tem-se duas *prioris* Beta para o parâmetro  $\rho_t$ , que são compatibilizadas por meio do teorema 2.2.2. Uma vez encontrada a distribuição a *priori* compatibilizada para  $\rho_t$ , sua *posteriori* é facilmente obtida por Teorema de Bayes. Agora temos que

$$(y_t | D_t) \sim \text{Logito-Beta}(\rho_t; \alpha, \beta) \text{ e desejamos obter } (y_t | D_t) \sim \text{Normal}(\mu_t; \sigma_t^2):$$

Projeta-se, então, a distribuição Logito-Beta no espaço da distribuição Normal e aplica-se o método Linear Bayes para encontrar os momentos de  $(y_t | D_t)$ . Tal distribuição terá como parâmetros  $(m_t; C_t)$ . Sendo estes:

$$m_t = m_{t-1} + R_t F_t (g_t - f_t) = q_t;$$

$$C_t = R_t - R_t F_t (R_t F_t)^0 (1 - \rho_t = q_t) = q_t;$$

onde  $g_t = E(f(\rho_t) | D_t)$  e  $p_t = Var(f(\rho_t) | D_t)$  podem ser calculados a partir da *posteriori*  $(\rho_t | D_t)$  já compatibilizada, onde  $f(\rho_t) = \text{logito}(\rho_t)$ .

### 4.3.3 Aplicação

A série a ser analisada é uma série trimestral de vendas com padrão sazonal conhecido na literatura. Adota-se um modelo Poisson para a contagem de vendas.

É observado que essa possui uma mudança de padrão sazonal a partir, aproximadamente, do tempo 20, além de possuir uma tendência crescente. O intuito do uso dessa série é verificar se o modelo consegue captar essa mudança de padrão sazonal. Além da estimação via metodologia que compatibiliza modelos (GI - com comp), apresentada

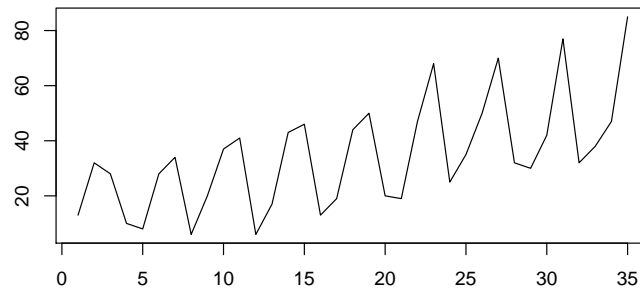


Figura 4.3: Série Observada

nesta subseção, foram feitas estimações via geometria da Informação sem compatibilizar modelos (GI - sem comp), *Conjugate Updating* e Modelos de Nível Local (MNL) conforme descrito na seção 3.2.2.

No que concerne à estimação via GI - sem comp. e com comp., componentes de sazonalidade foram introduzidas no modelo através da matriz  $G$  de forma que o primeiro e segundo harmônicos foram considerados. Como fatores de desconto foram utilizados 0.9 para componentes de tendência e 0.7 para componentes de sazonalidade. Não é usual que o desconto atribuído para a sazonalidade seja maior que o da tendência, entretanto os resultados encontrados por Pole et al. (1994) foram similares a esses apresentados, fornecendo o melhor resultado no que concerne a função de log-verossimilhança.

Por conseguinte, considera-se o modelo

$$y_{tj} \sim \text{Poisson}(\lambda_{tj})$$

$$\lambda_{tj} = \log(\lambda_{tj}) = F_t^j \theta_j$$

$$\theta_j = G_{tj} \theta_{j-1}$$

De modo que,

$$\theta_j = (\theta_{j1} \ \theta_{j2} \ \theta_{j3} \ \theta_{j4} \ \theta_{j5} \ \theta_{j6} \ \theta_{j7});$$

$$F_t = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 2 & & & & & & 3 \end{bmatrix};$$

$$G_t = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos(w) & \sin(w) & 0 & 0 & 0 \\ 0 & 0 & \sin(w) & \cos(w) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos(2w) & \sin(2w) & 0 \\ 0 & 0 & 0 & 0 & \sin(2w) & \cos(2w) & 0 \end{bmatrix}; \text{sendo } w = 2\pi/4;$$

Essas mesmas matrizes de desenho e de evolução foram utilizadas no método de geometria da informação que compatibiliza modelos estruturado e não estruturado e no método *Conjugate Updating*. De forma análoga, os fatores de desconto foram os mesmos.

Conforme mencionado anteriormente, os parâmetros da *priori* conjugada não serão mais obtidos através da compatibilização dos momentos, fazendo com que seja necessário atribuir uma evolução para eles que independe do vetor de estados. O método escolhido foi evoluir da posteriori no tempo  $t-1$  para *priori* no tempo  $t$ , preservando-se a média a *posteriori* e inflacionando-se a variância da *posteriori* por um fator multiplicativo  $1/\beta$ , como em [Smith \(1979\)](#); [Harrison and Stevens \(1976\)](#); [Gamerman et al. \(2013\)](#). Para que tal parâmetro fosse definido, foi especificada uma grade de valores possíveis e o modelo estimado com cada um desses valores. Como este atua como um fator de desconto, os possíveis valores foram definidos como (0.5;0.7;0.75;0.8;0.85;0.9;0.95;0.975;0.99;1). Na abordagem baseada em compatibilização de modelos estruturado e de nível local,  $S$ ,



necessária aplicação do teorema 2.2.2, também deve ser escolhido e portanto uma grade de valores para tal foi devida. Essa foi igual a (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0), com menores valores atribuindo menor peso ao modelo estruturado que ao modelo de nível local. Em seguida foi criada uma grade única com as todas as combinações possíveis para esses parâmetros. Exemplo:  $\beta = 0.5$ ;  $s = 0.1$ ;  $\lambda = 0.975$ ;  $s = 0.8$  e assim por diante. O melhor modelo foi devida como aquele que forneceu maior valor para a log-verossimilhança preditiva. A tabela resultante pode ser encontrada no apêndice 1 deste trabalho. Segundo esse critério,  $\beta = 0.9$  e  $\lambda = 0.5$ . Este modelo também foi o que forneceu o menor EQM preditivo.

Para aplicação do modelo de nível local baseado em covariáveis sem efeito dinâmico e que, portanto, apresentando apenas um nível dinâmico, foram devidas 3 covariáveis:  $x_1 = \cos(\omega t)$ ,  $x_2 = \sin(\omega t)$ ,  $x_3 = \cos(2\omega t)$ , onde  $t=1; \dots; 35$  e  $\omega = 2\pi = 4$ . Observe-se que, para este valor de  $\omega$ ,  $\sin(2\omega t)$  é sempre nulo. Dessa forma o modelo foi devida como:

$$Y_{jt} \mid \theta_t \sim \text{Poisson}(\theta_t) \\ \theta_t = \theta_t \exp(\beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t})$$

Além dos coeficientes atrelados às covariáveis, o fator de incidência do modelo também deve ser determinado. Tal estimação foi feita a partir do método de Monte Carlo via Cadeia de Markov com algoritmo Metropolis-Hastings. A priori para o parâmetro natural,  $\theta_t$ , foi devida como  $\theta_t \sim G(0.01; 0.01)$ . Para o parâmetro  $\beta$ , uma Uniforme(0,1) foi atribuída e para os coeficientes atrelados às covariáveis ( $\beta_k$ ;  $k = 1; \dots; 3$ ) foram devidas como priors distribuições Normal(0; 100). As amostras da distribuição a posteriori foram obtidas utilizando o método MCMC com 100.000 iterações, descartando-se burn-in de 1000 iterações e aplicando um espaçamento de tamanho 20, resultando em uma amostra de tamanho 4951. A convergência pôde ser verificada com auxílio de métodos de diagnóstico presentes no pacote coda no R (Plummer et al. (2006)). O modelo também foi estimado seguindo abordagem clássica.

A tabela 4.1 indica que as estimativas clássica e bayesiana ficaram bem próximas. Os

|   | EMV   | EB-Mediana | 2.5% via MCMC | 97.5% via MCMC |
|---|-------|------------|---------------|----------------|
| ! | 0.65  | 0.61       | 0.36          | 0.74           |
| 1 | -0.41 | -0.41      | -0.52         | -0.30          |
| 2 | -0.43 | -0.44      | -0.53         | -0.35          |
| 3 | -0.13 | -0.12      | -0.19         | -0.05          |

Tabela 4.1: Resultados da estimaco dos parmetros do modelo de nvel local

estimadores bayesianos foram considerados para futuras anlises.

| Método             | segundos |
|--------------------|----------|
| MCMC               | 456.225  |
| GI - sem comp.     | 0.178    |
| GI - com comp.     | 0.235    |
| Conjugate Updating | 0.199    |

Tabela 4.2: Tempo computacional gasto em cada um dos métodos

Analisando a tabela 4.2, nota-se que o método MCMC foi o mais custoso computacionalmente. Por outro lado, todos os métodos sequenciais levaram menos de 1 segundo para obter as estimativas do modelo.

Nota-se que a predico a 1 passo via GI - sem comp. e bem próxima a estimativa obtida através do método Conjugate Updating(Figura 4.4). Assim como no modelo de geometria via compatibilizaco de modelos,esses modelos conseguem captar a mudana no padro sazonal, sem necessidade de uma interveno manual. As estimativas do modelo GI - com comp. diferem ligeiramente das obtidas pelos modelos citados anteriormente, mas ainda assim esto próximas aos valores observados. O modelo de nvel local no capturou a mudana no padro sazonal. **Harvey and Fernandes (1989)** afirma em seu artigo que no permitir que componentes de sazonalidade sejam estocsticas no e uma grave limitaco do método, contudo o que pode ser visto que melhores resultados so obtidos quando se permite que essas sejam estocsticas. Isso se dá porque modelos mais estruturados têm vantagem, pois permitem que tais componentes variem no tempo, ou seja, ha uma exibibilidade maior para os parmetros que capturar mudanas estruturais no dado.

(a)

(b)

(c)

(d)

Figura 4.4: Predição 1 passo a frente via geometria da informação compatibilizando modelos e sem compatibilizar Conjugate Updatinge modelo de nível local junto a série observada. Considerou-se que após 10 passos todos os modelos já tinham aprendido o comportamento da série.

Analisando a Figura 4.5, fica claro que as estimativas do nível obtidas através dos modelos estruturados são bem próximas durante todo o período de tempo. Contudo, as estimativas provenientes do modelo de nível local apresentam um nível menor que o obtido através dos outros modelos.

Figura 4.5:  $E(\hat{y}_T | D_T)$  estimado por diferentes métodos

Pode-se observar que, também no que se refere as estimativas suavizadas para o par de harmônicos, verifica-se que as obtidas através de todos os modelos encontram-se muito próximas uma das outras.

(a) Harmônico 1

(b) Harmônico 2

Figura 4.6: Média a posteriori suavizada para o par de harmônicos.

Como forma de compararmos o desempenho de tais modelos, utilizou-se como critério o Fator de Bayes a ser descrito a seguir.

Seja o Fator de Bayes do modelo 1 com respeito ao modelo 2 de nido como em **Kass and Raftery (1995)**:

$$FB_{12} = \frac{p(x|m_1)}{p(x|m_2)}$$

Suponha que desejemos comparar 2 modelos diferentes, problema este que pode ser tratado como um teste de hipóteses. Um problema de decisão é completamente especificado pela trípla  $f A; H; X g$ , onde  $A$  é o espaço de decisão,  $H$  o espaço de parâmetros e  $X$  é o espaço amostral.

Seja o modelo  $m_1$  de nido por  $H_1$  e o modelo  $m_2$  de nido por  $H_2$ , onde esses são designados por  $R_i$ ,  $i = 1; 2$ . O espaço de parâmetros é particionado em dois componentes disjuntos  $H_1$  e  $H_2$ . O espaço de ação é de nido por dois componentes,  $A = f a_1; a_2 g$ , onde  $a_i$  indica que a hipótese de  $R_i$  é a verdadeira e, portanto, deve ser aceita. Sejam  $k_1$  e  $k_2$  as perdas associadas a escolha dos modelos com respeito a uma determinada função perda.

Sabe-se pela literatura de teoria da decisão que deve-se tomar a ação  $a_1$  e somente se a perda esperada com relação a  $a_1$  for menor que em  $a_2$ , sendo este valor esperado

relacionado as distribuições a posteriori. Isso equivale a aceitar  $H_1$  se e somente se

$$\frac{P(R_1|x)}{P(R_2|x)} > \frac{k_2}{k_1}, \text{ que e equivalente a}$$

$$FB_{12} > \frac{k_2 P(R_2)}{k_1 P(R_1)}:$$

Se considerarmos  $k_1 = k_2$ , teremos que

$$\frac{FB_{12}}{(1 + FB_{12})} > \frac{p(R_2)}{p(R_1) + p(R_2)}:$$

Esse resultado será usado neste trabalho para escolha de um modelo dentre os modelos de interesse.

Seja  $K_{ij} = \frac{FB_{ij}}{(1 + FB_{ij})}$ ,  $i \in j; i = 1; 2; 3; 4$  e  $j = 1; 2; 3; 4$ , onde 1 representa o método via Geometria da Informação - Sem compatibilizar modelos, 2 - o método via Geometria da Informação - Compatibilizando Modelos, 3 - O método Conjugate Updating e 4 - o modelo de nível local.

Temos que

| Comparação de Modelos |
|-----------------------|
| $K_{12} = 0.4768$     |
| $K_{13} = 0.4959$     |
| $K_{14} = 0.7222$     |
| $K_{23} = 0.5191$     |
| $K_{24} = 0.7404$     |
| $K_{34} = 0.7254$     |

Tabela 4.3: Comparação dos modelos via geometria da informação compatibilizando modelos e sem compatibilizar Conjugate Updating e modelo de nível local

Assim fica claro que os modelos estruturados, com estimação via via geometria da informação e Conjugate Updating são muito próximos, pois as probabilidades variaram em torno de 0.5. Tais modelos apresentaram desempenho superior ao modelo de nível local de forma que as probabilidades variaram em torno de 0.70 indicando que o modelo de nível local seria escolhido em cerca de apenas 30% dos casos.

# Captulo 5

## Conjugate Updating estendido

O metodo Conjugate Updating estendido sugerido por Souza et al. (2016) e uma extensao ao modelo proposto por West et al. (1985) para a familia exponencial biparametrica. De modo igual, destaca-se por sua rapidez computacional quando comparado a metodos como MCMC. A estimacao para os parâmetros da distribuicao do vetor de estados e preditiva e feita de forma sequencial, assim cada nova observacao que chega e imediatamente incorporada. A seguir, uma breve descricao do metodo e feita.

Seja a familia exponencial biparametrica definida como em Bernardo and Smith (2001):

$$p(y_t | \theta; \tau) = f(y_t) \exp\{\tau [d_1(y_t) + d_2(y_t)] - g(\tau; \tau)\} \quad (5.1)$$

sendo  $f(y_t)$  uma funcao não negativa,  $\tau$  e  $\tau$  parâmetros desconhecidos,  $d_1(\cdot)$ ,  $d_2(\cdot)$  e  $g(\tau; \tau)$  funcoes conhecidas.

A partir de 5.1, pode-se encontrar a media do processo, denotada por  $\mu = E(y_t | \tau; \tau)$ . Suponha que tanto as medias como os parâmetros de precisao podem ser descritos por um conjunto de variaveis explicativas a cada tempo atraves de funcoes de ligacao diferentes.

Assim,

$$\begin{aligned} \mu &= \frac{\int y_t p(y_t | \tau; \tau) d y_t}{\int p(y_t | \tau; \tau) d y_t} \\ &= \frac{\int y_t f(y_t) \exp\{\tau [d_1(y_t) + d_2(y_t)] - g(\tau; \tau)\} d y_t}{\int f(y_t) \exp\{\tau [d_1(y_t) + d_2(y_t)] - g(\tau; \tau)\} d y_t} \end{aligned} \quad (5.2)$$

A priori para o vetor  $(\tau; \tau)$ , também seguindo a forma proposta por Bernardo and

Smith (2001), pode ser definida como:

$$p(\theta_t; y_{1:t} | D_{t-1}) / \exp\{\theta_1' y_{1:t} + \theta_2' y_{1:t} - \eta(\theta_t)\} \quad (5.3)$$

Desta forma, fica claro que a posteriori será tal que

$$p(\theta_t; y_{1:t} | D_t) / \exp\{\theta_1' y_{1:t} + \theta_2' y_{1:t} - \eta(\theta_t)\} \quad (5.4)$$

de forma que  $\theta_0 = \theta_0 + 1$ ,  $\theta_1 = \theta_1 + d_1(y_t)$  e  $\theta_2 = \theta_2 + d_2(y_t)$ .

O primeiro passo do método consiste na evolução dos estados e da preditiva dada  $\theta_{t-1}$  e  $C_{t-1}$ . Tal evolução é feita do mesmo modo como apresentado em West et al. (1985).

Sejam:  $m_t = E(\theta_t | D_t)$ ;  $C_t = \text{Var}(\theta_t | D_t)$ ;  $a_t = E(\theta_t | D_{t-1})$ ;  $R_t = \text{Var}(\theta_t | D_{t-1})$ ;  $f_t = E(\theta_t | D_{t-1})$ ;  $Q_t = \text{Var}(\theta_t | D_{t-1})$ . Tem-se então:

$$\begin{aligned} a_t &= G_t m_{t-1} \\ R_t &= G_t C_{t-1} G_t' + W_t \\ f_t &= F_t' a_t \\ Q_t &= F_t' R_t F_t \end{aligned} \quad (5.5)$$

Seguidamente, deseja-se encontrar os parâmetros  $\theta_{1t}$  e  $\theta_{2t}$ . Para tal utiliza-se os momentos a priori de  $(\theta_t | D_{t-1})$  e  $(\theta_t | D_t)$ . A partir destes, é possível encontrar os momentos a priori de  $(\theta_{1t} | D_{t-1})$  e  $(\theta_{2t} | D_{t-1})$  em função de  $\theta_{0t}$ ,  $\theta_{1t}$  e  $\theta_{2t}$ .

Seguindo a notação de Souza et al. (2016), tem-se que:

$$\begin{aligned} E(\theta_{1t} | D_{t-1}) &= h_1(\theta_{0t}; \theta_{1t}; \theta_{2t}); \\ E(\theta_{2t} | D_{t-1}) &= h_2(\theta_{0t}; \theta_{1t}; \theta_{2t}); \\ V(\theta_{1t} | D_{t-1}) &= h_3(\theta_{0t}; \theta_{1t}; \theta_{2t}); \\ V(\theta_{2t} | D_{t-1}) &= h_4(\theta_{0t}; \theta_{1t}; \theta_{2t}); \\ \text{Cov}(\theta_{1t}; \theta_{2t} | D_{t-1}) &= h_5(\theta_{0t}; \theta_{1t}; \theta_{2t}); \end{aligned} \quad (5.6)$$

De modo que ao compatibilizar com as equações em 5.5, o seguinte sistema é obtido:

$$\begin{aligned}
 f_{1t} &= h_1(\theta_t; \theta_{1t}; \theta_{2t}); \\
 f_{2t} &= h_2(\theta_t; \theta_{1t}; \theta_{2t}); \\
 q_{11t} &= h_3(\theta_t; \theta_{1t}; \theta_{2t}); \\
 q_{22t} &= h_4(\theta_t; \theta_{1t}; \theta_{2t}); \\
 q_{12t} &= h_5(\theta_t; \theta_{1t}; \theta_{2t});
 \end{aligned} \tag{5.7}$$

Pode-se observar que este sistema é superespecificado, ou seja, há mais equações do que parâmetros, portanto não há uma solução única para o sistema. A abordagem adotada para lidar com esse problema, em Souza et al. (2016) e no presente trabalho foi utilizar o Método de Momentos Generalizado. Este método consiste em encontrar a solução que minimiza a função

$$g_n(\theta; f_t; Q_t) = \sum_{n=1}^n g_n(\theta; f_t; Q_t); \tag{5.8}$$

onde,  $\theta = (\theta_t; \theta_{1t}; \theta_{2t})$  e  $g_n(\theta; f_t; Q_t)$  é a seguinte função vetorial:

$$g_n(\theta; f_t; Q_t) = \begin{pmatrix} f_{1t} - h_1(\theta_t; \theta_{1t}; \theta_{2t}) \\ f_{2t} - h_2(\theta_t; \theta_{1t}; \theta_{2t}) \\ q_{11t} - h_3(\theta_t; \theta_{1t}; \theta_{2t}) \\ q_{22t} - h_4(\theta_t; \theta_{1t}; \theta_{2t}) \\ q_{12t} - h_5(\theta_t; \theta_{1t}; \theta_{2t}) \end{pmatrix}; \tag{5.9}$$

Neste trabalho considerou-se a matriz  $I_n$  como sendo uma matriz identidade, isto é, considerou-se que todas as equações possuem o mesmo peso. Uma vez que o argumento mínimo da função é encontrado, é feita a atualização dos parâmetros com base na nova observação. Obtêm-se  $\theta_0; \theta_1; \theta_2$ , tornando possível, portanto, encontrar os momentos atualizados da posteriori do preditor linear e por fim, os estados. A seguir, alguns exemplos serão apresentados.



## 5.1 Distribuição Gama

Modelo Observacional:

Seja  $y_t = \frac{t}{\theta} e^{-\frac{t}{\theta}}$ , de forma que:

$$y_t = \frac{t}{\theta} e^{-\frac{t}{\theta}}$$

$$\begin{aligned} p(y_t | \theta) &= \frac{1}{\theta} y_t^{\theta-1} \exp(-y_t) \\ &= \frac{1}{\theta} \left(\frac{t}{\theta}\right)^{\theta-1} \exp\left(-\frac{t}{\theta}\right) \\ &= \frac{1}{\theta^\theta} t^{\theta-1} \exp\left(-\frac{t}{\theta}\right) \end{aligned}$$

Distribuição a priori conjugada:

$$p(\theta | \tau_1, \tau_2) \propto \exp\left(-\frac{\tau_1}{\theta}\right) \exp\left(-\frac{\tau_2}{\theta}\right) \exp\left(-\frac{\tau_1 + \tau_2}{\theta}\right) \exp\left(-\frac{\tau_1 + \tau_2}{\theta}\right)$$

Onde,  $\tau_1 > 0$  e  $\tau_2 > 0$ .

Aproximando os momentos da priori conjugada:

Como a priori não tem fórmula analítica fechada, obtou-se por encontrar sua moda e curvatura. Sejam  $\tau_1$  e  $\tau_2$  dois preditores dinâmicos para  $\mu_t$  e  $\sigma_t$ , respectivamente e considere funções de ligação logartímicamente relacionando tais preditores aos parâmetros de interesse:

$$\tau_1 = \log(\theta);$$

$$\tau_2 = \log(\theta);$$

Equações Resultantes no Modelo Dinâmico:

Nota-se que a distribuição a priori conjugada não tem forma conhecida. Desse modo, obtou-se por encontrar sua moda e matriz de curvatura, que resultou em:

$$f_{1t} = E(\log(\theta_t) | D_{t-1}) = \log \frac{1t}{0t} ;$$

$$f_{2t} = E(\log(\theta_t) | D_{t-1}) = \log \left[ \frac{0t}{2 \log \frac{1t}{0t}} \right] A ;$$

$$Q_{1t} = \text{Var}(\log(\theta_t) | D_{t-1}) = \frac{2}{0t} \log \frac{1t}{0t} ;$$

$$Q_{2t} = E(\log(\theta_t) | D_{t-1}) = \frac{2}{0t} ;$$

Restrições:

Como as quantidades  $Q_{1t}$  e  $Q_{2t}$  representam variâncias e estas são positivas por definição, o uso de restrições se faz necessário. Assim,

$$0t > 0; \quad 1t > 0; \quad \frac{2t}{0t} > f_{1t} ;$$

Atualização:

$$0t = 0t + 1$$

$$1t = 1t + y_t$$

$$2t = 2t + \log(y_t)$$



Equações Resultantes nos Modelo Dinâmico:

$$f_{1t} = E(D_t | D_{t-1}) = \frac{1_t}{0_t};$$

$$f_{2t} = E(\log(D_t) | D_{t-1}) = \log \left( \frac{(0_t + 1)^{1/2}}{2} \right) + \frac{1_t}{0_t};$$

$$Q_{1t} = \text{Var}(D_t | D_{t-1}) = \frac{1_t}{0_t} \left( \frac{0_t + 1}{2} \right);$$

$$Q_{2t} = \text{Var}(\log(D_t) | D_{t-1}) = \frac{2}{(0_t + 1)};$$

Restrições:

Como as quantidades  $Q_{1t}$  e  $Q_{2t}$  representam variâncias e estas são positivas por definição, o uso de restrições se faz necessário. Assim,

$$0_t > 1; \quad 2_t > \frac{2}{0_t} = 0_t$$

Atualização:

$$0_t = 0_{t-1} + 1$$

$$1_t = 1_{t-1} + y_t$$

$$2_t = 2_{t-1} + (y_t)^2$$

### 5.3 Distribuição Multinomial

A seguir uma extensão da metodologia proposta por Souza et al. (2016) e apresentada para o caso Multinomial.

Modelo Observacional:

$$\begin{aligned}
 p(y_t | \theta_{1t}; \theta_{2t}) &= \frac{N!}{y_{1t}! y_{2t}! y_{3t}!} \theta_{1t}^{y_{1t}} \theta_{2t}^{y_{2t}} (1 - \theta_{1t} - \theta_{2t})^{y_{3t}}; \\
 &= \frac{N!}{y_{1t}! y_{2t}! y_{3t}!} \frac{\theta_{1t}^{y_{1t}}}{(1 - \theta_{1t} - \theta_{2t})^{y_{1t}}} \frac{\theta_{2t}^{y_{2t}}}{(1 - \theta_{1t} - \theta_{2t})^{y_{2t}}} \\
 &\quad + (1 - \theta_{1t} - \theta_{2t})^N;
 \end{aligned}$$

onde  $0 < \theta_{1t}; \theta_{2t} < 1$  e  $y_t \in \mathbb{R}^+$ .

Distribuição a priori conjugada:

$$\begin{aligned}
 p(\theta_{1t}; \theta_{2t}) &/ \frac{\theta_{1t}^{\alpha_1 - 1} \theta_{2t}^{\alpha_2 - 1}}{(1 - \theta_{1t} - \theta_{2t})^{\alpha_1 + \alpha_2 - 1}}; \\
 &\quad + (1 - \theta_{1t} - \theta_{2t})^{\alpha_3 - 1}; \\
 &/ \int_0^1 \int_0^{1 - \theta_{1t}} \theta_{1t}^{\alpha_1 - 1} \theta_{2t}^{\alpha_2 - 1} (1 - \theta_{1t} - \theta_{2t})^{\alpha_3 - 1} d\theta_{1t} d\theta_{2t};
 \end{aligned}$$

que representa o núcleo da função densidade de probabilidade Dirichlet com parâmetros

$$\alpha_1 = \alpha_1 + 1; \quad \alpha_2 = \alpha_2 + 1 \quad \text{e} \quad \alpha_3 = \alpha_3 - \alpha_1 - \alpha_2 + 1.$$

Funções de ligação:

$$\begin{aligned}
 \eta_{1t} &= \log \frac{\theta_{1t}}{1 - \theta_{1t} - \theta_{2t}}; \\
 \eta_{2t} &= \log \frac{\theta_{2t}}{1 - \theta_{1t} - \theta_{2t}};
 \end{aligned}$$

Aproximando os momentos das funções de ligação:

Observe que não há forma analítica fechada para as distribuições das funções de ligação. Optou-se, portanto, em utilizar expansão de Taylor bivariada para que os momentos pudessem ser encontrados.

Seja  $z = \langle x; y \rangle$  e  $a = (a, b)$ , a série de Taylor pode ser escrita como:

$$f(z) = f(a) + [(z - a)^T \nabla f(a)] + \frac{1}{2} [(z - a)^T H(z) (z - a)]:$$

Seja  $H(x)$  a matriz de segundas derivadas (Hessiana):

$$H(x; y) = \begin{pmatrix} f_{xx}(x; y) & f_{xy}(x; y) \\ f_{yx}(x; y) & f_{yy}(x; y) \end{pmatrix}:$$

A expansão de Taylor de segunda ordem pode ser reescrita como:

$$f(x; y) = f(a; b) + f_x(a; b)(x - a) + f_y(a; b)(y - b) + \frac{1}{2} f_{xx}(x - a)^2 + 2f_{xy}(x - a)(y - b) + \frac{1}{2} f_{yy}(y - b)^2:$$

No caso da multinomial, considere

$$f_1(t_1; t_2) = \log \frac{t_1}{1 - t_1 - t_2};$$

$$f_2(t_1; t_2) = \log \frac{t_2}{1 - t_1 - t_2};$$

$$a = E(t_1);$$

$$b = E(t_2):$$

Para que a expansão de Taylor seja feita, é necessário que as primeiras e segundas derivadas, assim como os momentos de  $t_1$  e  $t_2$ , sejam encontrados. Os resultados encontram-se dispostos a seguir.

$$\frac{f(1t; 2t)}{f(1t)} = \frac{1 - 2t}{1 - (1 - 1t - 2t)};$$

$$\frac{f(1t; 2t)}{f(2t)} = \frac{1}{1 - 1t - 2t};$$

$$\frac{f(1t; 2t)}{f(1t)f(1t)} = \frac{(1 - 2t)(2 - 1t + 2t - 1)}{1 - (1 - 1t - 2t)^2};$$

$$\frac{f(1t; 2t)}{f(2t)f(2t)} = \frac{1}{(1 - 1t - 2t)^2};$$

$$\frac{f(1t; 2t)}{f(1t)f(2t)} = \frac{1}{(1 - 1t - 2t)^2};$$

$$\frac{f(1t; 2t)}{f(2t)} = \frac{1 - 1t}{1 - (1 - 1t - 2t)};$$

$$\frac{f(1t; 2t)}{f(1t)} = \frac{1}{1 - 1t - 2t};$$

$$\frac{f(1t; 2t)}{f(2t)f(2t)} = \frac{(1 - 1t)(2 - 2t + 1t - 1)}{1 - (1 - 1t - 2t)^2};$$

$$\frac{f(1t; 2t)}{f(1t)f(1t)} = \frac{1}{(1 - 1t - 2t)^2};$$

$$\frac{f(1t; 2t)}{f(1t)f(2t)} = \frac{1}{(1 - 1t - 2t)^2};$$

$$\begin{aligned}
E(1tjD_{t-1}) &= \frac{1}{1+2+3} = \frac{1+1}{0+3} = \sim 1t; \\
E(2tjD_{t-1}) &= \frac{2}{1+2+3} = \frac{2+1}{0+3} = \sim 2t; \\
\text{Var}(1tjD_{t-1}) &= \frac{1(2+3)}{(1+2+3)^2(1+2+3+1)} \\
&= \frac{(1+1)(0+1+2)}{(0+3)^2(0+4)} = \frac{2}{1}; \\
\text{Var}(2tjD_{t-1}) &= \frac{2(1+3)}{(1+2+3)^2(1+2+3+1)} \\
&= \frac{(2+1)(0+2+2)}{(0+3)^2(0+4)} = \frac{2}{2}; \\
\text{Cov}((1t; 2t)jD_{t-1}) &= \frac{1 \cdot 2}{(1+2+3)^2(1+2+3+1)} \\
&= \frac{(1+1)(2+1)}{(0+3)^2(0+4)} = 1 \cdot 2;
\end{aligned}$$



Dessa forma,

$$\begin{aligned}
 E \log \frac{1t}{1 - 1t - 2t} jD_{t-1} &= \log \frac{\tilde{1t}}{1 - \tilde{1t} - \tilde{2t}} + \frac{1(1 - \tilde{2t})(2\tilde{1t} + \tilde{2t} - 1)}{2\tilde{1t}^2(1 - \tilde{1t} - \tilde{2t})^2} \\
 &+ \frac{1}{2(1 - \tilde{1t} - \tilde{2t})^2} + \frac{1 - 2}{(1 - \tilde{1t} - \tilde{2t})^2} \\
 E \log \frac{2t}{1 - 1t - 2t} jD_{t-1} &= \log \frac{\tilde{2t}}{1 - \tilde{1t} - \tilde{2t}} + \frac{1(1 - \tilde{1t})(2\tilde{2t} + \tilde{1t} - 1)}{2\tilde{2t}^2(1 - \tilde{1t} - \tilde{2t})^2} \\
 &+ \frac{1}{2(1 - \tilde{1t} - \tilde{2t})^2} + \frac{1 - 2}{(1 - \tilde{1t} - \tilde{2t})^2} \\
 Var \log \frac{1t}{1 - 1t - 2t} jD_{t-1} &= \frac{1 - \tilde{2t}}{\tilde{1t}(1 - \tilde{1t} - \tilde{2t})^2} + \frac{1}{1 - \tilde{1t} - \tilde{2t}} \\
 &+ \frac{1}{1 - \tilde{1t} - \tilde{2t}} + \frac{1 - \tilde{2t}}{\tilde{1t}(1 - \tilde{1t} - \tilde{2t})} \\
 Var \log \frac{2t}{1 - 1t - 2t} jD_{t-1} &= \frac{1 - \tilde{1t}}{\tilde{2t}(1 - \tilde{1t} - \tilde{2t})^2} + \frac{1}{1 - \tilde{1t} - \tilde{2t}} \\
 &+ \frac{1}{1 - \tilde{1t} - \tilde{2t}} + \frac{1 - \tilde{1t}}{\tilde{2t}(1 - \tilde{1t} - \tilde{2t})}
 \end{aligned}$$

Sejaf ( 1t; 2t) = log 1t / (1 - 1t - 2t) log 2t / (1 - 1t - 2t)

$$\begin{aligned}
 f(1t; 2t) &= f(\tilde{1t}; \tilde{2t}) + \frac{\partial}{\partial 1t} \frac{\tilde{1t}}{1 - \tilde{1t} - \tilde{2t}} \text{Arf}(1t; 2t) |_{(1t; 2t)=(\tilde{1t}; \tilde{2t})} \\
 &= \log \frac{\tilde{1t}}{1 - \tilde{1t} - \tilde{2t}} + \log \frac{\tilde{2t}}{1 - \tilde{1t} - \tilde{2t}} \\
 &+ \frac{\partial}{\partial 1t} \frac{\tilde{1t}}{1 - \tilde{1t} - \tilde{2t}} \text{Arf}(1t; 2t) |_{(1t; 2t)=(\tilde{1t}; \tilde{2t})}
 \end{aligned}$$

Onde,

$$r f ( \tau_1; \tau_2 )_{j(\tau_1)=(-\tau_1)} = \frac{(1 - \tau_2) \frac{\tau_1}{1 - \tau_1 - \tau_2} + \tau_1 \log \frac{\tau_2}{1 - \tau_2 - \tau_1}}{1 - \tau_1(1 - \tau_1 - \tau_2)}$$

$$r f ( \tau_1; \tau_2 )_{j(\tau_2)=(-\tau_2)} = \frac{(1 - \tau_1) \frac{\tau_2}{1 - \tau_1 - \tau_2} + \tau_2 \log \frac{\tau_1}{1 - \tau_2 - \tau_1}}{1 - \tau_2(1 - \tau_1 - \tau_2)}$$

Mas, se aproximarmos  $E( \tau_1 j D_{t-1} ) = \tau_1$  e  $E( \tau_2 j D_{t-1} ) = \tau_2$

$$\begin{matrix} 0 & 1 & 0 & 1 \\ @ & \tau_1 & @ & 0 \\ & \tau_2 & & 0 \end{matrix} A = \begin{matrix} 0 \\ 0 \end{matrix}$$

E, portanto,

$$E( \tau_1 \tau_2 j D_{t-1} ) = \log \frac{\tau_1}{1 - \tau_1 - \tau_2} \log \frac{\tau_2}{1 - \tau_1 - \tau_2}$$

Assim, a covariância entre as funções de ligação será dada por:

$$\begin{aligned} \text{Cov}( \tau_1; \tau_2 j D_{t-1} ) &= \log \frac{\tau_1}{1 - \tau_1 - \tau_2} \log \frac{\tau_2}{1 - \tau_1 - \tau_2} \\ &\quad \log \frac{\tau_1}{1 - \tau_1 - \tau_2} \log \frac{\tau_2}{1 - \tau_1 - \tau_2} = 0 \end{aligned}$$

O sistema a ser resolvido será tal que:

$$f_{1t} = \log \frac{1+1}{0 \quad 1 \quad 2+1}$$

$$f_{2t} = \log \frac{2+1}{0 \quad 1 \quad 2+1}$$

$$q_{1t} = \frac{(0+3)(0 \quad 2+2)}{(1+1)(0 \quad 1 \quad 2+1)}^2 \frac{(1+1)(0 \quad 1+2)}{(0+3)^2(0+4)}$$

$$+ \frac{0+3}{0 \quad 1 \quad 2+1}^2 \frac{(2+1)(0 \quad 2+2)}{(0+3)^2(0+4)}$$

$$+ \frac{(0+3)(0 \quad 2+2)}{(1+1)(0 \quad 1 \quad 2+1)} \frac{0+3}{0 \quad 1 \quad 2+1} \frac{(1+1)(2+1)}{(0+3)^2(0+4)}$$

$$q_{2t} = \frac{(0+3)(0 \quad 2+2)}{(1+1)(0 \quad 1 \quad 2+1)}^2 \frac{(2+1)(0 \quad 2+2)}{(0+3)^2(0+4)}$$

$$+ \frac{0+3}{0 \quad 1 \quad 2+1}^2 \frac{(1+1)(0 \quad 1+2)}{(0+3)^2(0+4)}$$

$$+ \frac{(0+3)(0 \quad 2+2)}{(1+1)(0 \quad 1 \quad 2+1)} \frac{0+3}{0 \quad 1 \quad 2+1} \frac{(1+1)(2+1)}{(0+3)^2(0+4)}$$

### 5.3.1 Estudo Simulado

Visando avaliar o comportamento da metodologia proposta, um estudo simulado foi feito e seus resultados serão apresentados ao longo desta seção.

Considere  $(Y_{jt} | \pi_{1t}; \pi_{2t}; \pi_{3t}) \sim \text{Multinomial}(\pi_{1t}; \pi_{2t}; \pi_{3t})$  e seja a estrutura do modelo a seguinte:

$$\begin{aligned} a_t &= a_{t-1} + \epsilon_t; W \sim N(0; 0:1^2) \\ b_t &= b_{t-1} + \epsilon_t; W \sim N(0; 0:1^2) \\ \pi_{1t} &= \frac{\exp(a_t)}{1 + \exp(a_t)} \\ \pi_{2t} &= \frac{\exp(b_t)}{1 + \exp(b_t)} \end{aligned}$$

De modo que  $a_t$  e  $b_t$  são parâmetros auxiliares e

$$\begin{aligned} \pi_{1t} &= \log \frac{\pi_{1t}}{1 - \pi_{1t} - \pi_{2t}} ; \\ \pi_{2t} &= \log \frac{\pi_{2t}}{1 - \pi_{1t} - \pi_{2t}} ; \end{aligned}$$

Para análise, foram geradas 50 amostras de um modelo com essa estrutura, sendo  $a_1 = 1$  e  $b_1 = 0.8$ . A Figura 5.1 mostra que o método foi capaz de capturar a estrutura dos

(a)  $E(\pi_{1t} | D_t)$

(b)  $E(\pi_{2t} | D_t)$

Figura 5.1: Boxplots para as médias a posteriori dos parâmetros  $\pi_{1t}$  e  $\pi_{2t}$  junto aos seus valores verdadeiros (linha vermelha).

parâmetros  $\pi_{1t}$  e  $\pi_{2t}$  durante todo o período de tempo, conforme esperado. Nota-se que

a media das estimativas das amostras ficou muito próxima ao valor verdadeiro, a cada tempo  $t$ , indicando um bom ajuste.

(a)  $E(\hat{\mu}_{1t}|D_t)$

(b)  $E(\hat{\mu}_{2t}|D_t)$

Figura 5.2:  $E(\hat{\mu}_{1t}|D_t)$  e  $E(\hat{\mu}_{2t}|D_t)$  (linha pontilhada azul) junto aos seus valores verdadeiros (linha contínua preta).

Nota-se pela Figura 5.2 que as estimativas das médias posteriores para  $\mu_{1t}$  e  $\mu_{2t}$  ficaram bem próximas aos seus respectivos valores verdadeiros, novamente indicando um bom ajuste do modelo.

Como forma de ilustrar como ficam as previsões um passo a frente, o resultado de uma das réplicas é apresentado na Figura 5.3.

Verifica-se através das Figuras 5.3 que as previsões acompanharam a estrutura da série observada em todos os casos. Conclui-se, portanto, que a metodologia apresentada teve um bom desempenho no que concerne a estimação de parâmetros e previsão a um passo.

(a)  $E(Y_{1t+1} | D_t)$

(b)  $E(Y_{2t+1} | D_t)$

(c)  $E(Y_{3t+1} | D_t)$

Figura 5.3:  $E(Y_{1t+1} | D_t)$ ,  $E(Y_{2t+1} | D_t)$  e  $E(Y_{3t+1} | D_t)$  (linha pontilhada azul) junto a s rie observada (linha contnua preta).

## 5.4 Aplicações a dados reais

Como forma de ilustrar o método proposto, a série mensal de log-retorno da IBM será analisada nesta seção. Utilizou-se o modelo Normal com média e precisão desconhecidos para estimação. Os dados encontram-se disponíveis em [Tsay \(2005\)](#) e discutidos em [Triantafyllopoulos \(2008\)](#). Os resultados a serem apresentados foram obtidos com auxílio do pacote `nlptr` do software R, que permite resolver o sistema com mais equações do que parâmetros impondo restrições não lineares. O algoritmo de otimização escolhido foi o de Otimização restrita por aproximação linear (COBYLA). As matrizes  $F_t$  e  $G_t$  foram definidas como sendo iguais a 1 para todo tempo

Figura 5.4: Série mensal do log-retorno da SP500

No que se refere a previsão a 1 passo, temos que essa acompanha bem os movimentos da série observada, apresentando queda quando há decaimentos mais fortes no log-retorno. Portanto, tem-se que todas as observações encontram-se dentro do intervalo de credibilidade obtido.

Figura 5.5: S rie observada junto a predic o 1 passo a frente e seu respectivo intervalo de credibilidade.

Analisando a Figura 5.6(a), nota-se que a m dia se encontra em torno de zero, o que est  de acordo com a literatura. Suas atua es acompanham o movimento da s rie, indicando um bom ajuste. No que concerne ao par metro de precis o (Figura 5.6(b)), observa-se que este est  entre zero e 0.2, ou seja, assume valores muito pequenos. A volatilidade da s rie, portanto,   alta, resultado este condizente com o que se encontra na literatura.

(a)  $t$

(b)  $t$

Figura 5.6: M dia a posteriori ltrada e suavizada junto ao intervalo de credibilidade da m dia suavizada para o par metro da m dia e de precis o.



# Capítulo 6

## Conclusão

O presente trabalho teve por objetivo discutir métodos de inferência para a classe de modelos lineares dinâmicos para a família exponencial  $p$ -dimensional. Foram visitadas metodologias propostas por [West et al. \(1985\)](#), [Gamerman et al. \(2013\)](#), [Smith \(1979\)](#) e [Souza et al. \(2016\)](#), além de apresentar-se novas abordagens para inferência nessa classe de modelos via elementos de geometria da informação.

Foi verificado que, uma vez que se atribui uma distribuição *a priori* para o vetor de estados, uma *a priori* induzida para o parâmetro canônico da família exponencial é obtida. Por outro lado, há também a *a priori* conjugada natural nesta família. Por motivos de conveniência, uma vez que conhecemos as propriedades de tal distribuição, deseja-se trabalhar com a conjugada.

Uma abordagem de geometria da informação apresentada sugere que a distribuição induzida seja projetada no espaço da distribuição conjugada via minimização da divergência de Kullback-Leibler. Outra sugere que não apenas se faça tal projeção, mas também que se compatibilize tal informação com a proveniente de um modelo de nível local. A vantagem do segundo é que, o modelo de nível local por si não permite que as componentes de tendência e sazonalidade sejam estocásticas, enquanto o método aqui proposto admite essa propriedade.

Verificou-se, através de exercícios de simulação e aplicação a dados reais que tal metodologia tem um bom desempenho no que concerne a estimação do vetor de estados e previsão a um passo. Foi verificado, inclusive, que mudanças no padrão sazonal são

captadas sem necessidade de intervenção manual. Ficou claro que as estimativas do modelo via geometria da informação - sem compatibilizar modelos - cam muito próximas das obtidas via Conjugate Updating (West et al. (1985)). No que se refere ao modelo via geometria que compatibiliza informações de um modelo estruturado e de um modelo de média local, nota-se que suas estimativas vão depender do peso dado a cada modelo, podendo se aproximar mais de Conjugate Updating ou do modelo de nível local. Quanto ao último, foi verificado que não consegue captar mudanças no padrão sazonal, o que pode estar relacionado ao fato de não permitir que as componentes de sazonalidade sejam estocásticas, conforme mencionado anteriormente. Embora Harvey and Fernandes (1989) argumentem que essa não é uma forte limitação, acreditamos que é sim.

No que concerne a família exponencial  $k$ -paramétrica o método apresentado por Souza et al. (2016) foi visitado. Uma vez utilizada a priori definida por Bernardo and Smith (2001), nos casos aqui apresentados, 3 parâmetros devem ser estimados. Entretanto neste caso há 2 funções de ligação: uma para o parâmetro da média e outra para o parâmetro de precisão. Observe que cada uma dessas tem sua média e variância, além da covariância que pode existir entre elas, totalizando até 5 equações. Dessa forma, nota-se que o sistema é super especificado: 5 equações para 3 parâmetros. A solução encontrada pelos autores foi utilizar o método de momentos generalizado para redução do sistema. Apresentou-se nesse trabalho uma extensão do método para a distribuição Multinomial. Foi verificado, através de simulação, que a metodologia proposta consegue captar o nível dos parâmetros, assim como a previsão a um passo acompanha a estrutura dos dados observados.

Como trabalhos futuros, pretende-se desenvolver uma metodologia baseada em geometria da informação para a família exponencial  $k$ -paramétrica. Além disso, a expansão da metodologia para dados composicionais.

# Apêndice A

## Resultados secundários referentes a série de vendas

A.1 Tabela de resultados do modelo Poisson via Geometria da Informação que compatibiliza modelos

|    | s    | !    | EQM Preditivo | Log- verossimilhança |
|----|------|------|---------------|----------------------|
| 1  | 0.90 | 0.50 | 74.62         | -131.32              |
| 2  | 0.90 | 0.70 | 87.89         | -132.66              |
| 3  | 0.90 | 0.75 | 91.78         | -133.12              |
| 4  | 0.90 | 0.80 | 95.82         | -133.62              |
| 5  | 0.90 | 0.85 | 99.97         | -134.16              |
| 6  | 0.90 | 0.90 | 104.21        | -134.74              |
| 7  | 0.90 | 0.95 | 108.50        | -135.34              |
| 8  | 0.90 | 0.97 | 110.67        | -135.65              |
| 9  | 0.90 | 0.99 | 111.97        | -135.84              |
| 10 | 0.90 | 1.00 | 112.84        | -135.97              |

Tabela A.1: Tabela com os resultados obtidos de acordo com os parâmetros para os 10 melhores modelos.

## A.2 Cadeias obtidas pelo MCMC

(a)  $\theta_1$

(b)  $\theta_2$

(c)  $\theta_3$

(d)  $\theta_4$

Figura A.1: Cadeias resultantes obtidas pelo método MCMC para cada um dos parâmetros  $\theta_1, \theta_2, \theta_3$  e para o parâmetro  $\theta_4$ .

# Referências Bibliográficas

- Amari, S.-i. (2016). Information geometry and its applications. Springer.
- Ameen, J. and Harrison, P. (1984). Discount weighted estimation Journal of Forecasting, 3(3):285{296.
- Bernardo, J. M. and Smith, A. F. (2001). Bayesian theory.
- Gamerman, D., Santos, T. R., and Franco, G. C. (2013). A non-gaussian family of state-space models with exact marginal likelihood Journal of Time Series Analysis, 34(6):625{645.
- Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting Journal of the Royal Statistical Society. Series B (Methodological), pages 205{247.
- Harvey, A. C. and Fernandes, C. (1989). Time series models for count or qualitative observations. Journal of Business & Economic Statistics, 7(4):407{417.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors Journal of the American Statistical Association, 90(430):773{795.
- Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models Journal of the Royal Statistical Society, Series A, General, 135:370{384.
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for mcmc. R News, 6(1):7{11.
- Pole, A., West, M., and Harrison, J. (1994) Applied Bayesian forecasting and time series analysis. CRC press.

- Poole, D. and Raftery, A. E. (2000). Inference for deterministic simulation models: the bayesian melding approach. Journal of the American Statistical Association, 95(452):1244–1255.
- Powell, M. J. (1994). A direct search optimization method that models the objective and constraint functions by linear interpolation. In Advances in optimization and numerical analysis, pages 51–67. Springer.
- Smith, J. (1979). A generalization of the bayesian steady forecasting model. Journal of the Royal Statistical Society. Series B (Methodological), pages 375–387.
- Smith, R. and Miller, J. (1986). A non-gaussian state space model e application to prediction of records. Journal of the Royal Statistical Society. Series B (Methodological), pages 79–88.
- Souza, M. A. d. O., Migon, H. d. S., and Pereira, J. (2016). Extended dynamic generalized linear models: the two-parameter exponential family. Computational Statistics & Data Analysis.
- Triantafyllopoulos, K. (2008). Dynamic generalized linear models for non-gaussian time series forecasting. arXiv preprint arXiv:0802.0219.
- Tsay, R. S. (2005). Analysis of financial time series, volume 543. John Wiley & Sons.
- West, M. and Harrison, P. J. (1996). Bayesian forecasting. Wiley Online Library.
- West, M., Harrison, P. J., and Migon, H. S. (1985). Dynamic generalized linear models and bayesian forecasting. Journal of the American Statistical Association, 80(389):73–83.