



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

Dissertação de Mestrado

Modelos Dinâmicos  
para a Distribuição Poisson Generalizada

por

Patrícia Oliveira de Souza

Orientador: Prof<sup>a</sup>. Dr<sup>a</sup>. Cibele Queiroz da Silva

Julho de 2014

Patrícia Oliveira de Souza

**Modelos Dinâmicos  
para a Distribuição Poisson Generalizada**

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

**Universidade de Brasília**

**Brasília, Julho de 2014**

*A Deus, por tudo.*

*A meus pais, Regina e Hudson, pelo amor, força e confiança.*

*Ao Josian, meu bem.*

*À Helena, minha vida.*

# Agradecimentos

Gostaria de agradecer:

A Deus por ser imprescindível em minha vida, me guiando por caminhos de grandes realizações e de aprendizagem pra vida.

A meus pais, Regina e Hudson, por todo amor dado sem troca, pela educação e ensinamentos, pelo apoio incondicional dado em todos os momentos da minha vida, em especial, naqueles em que mais necessitei. Pelas noites mal dormidas, pelas palavras e abraços de conforto, por muitas vezes abdicarem de seus interesses e sonhos para dá espaço aos meus. Agradeço a meus pais por simplesmente serem os melhores, agradeço pelo exemplo de caráter, honestidade e bondade que hoje fazem parte de mim. Enfim, agradeço a eles por tudo.

A Meu Bem, Josian, meu amor, meu namorado e agora pai da nossa Helena. Dedico grande parte desse trabalho a ele que me proporciona os momentos mais felizes da minha vida, me apoiando, me incentivando em todas as minhas decisões. Obrigada por acreditar mais em mim do que eu mesma, por ser meu porto seguro e alicerce, por ser um grande companheiro, por me mostrar sempre o lado bom da vida, por me apoiar, principalmente, nesse momento e por ser um pai tão maravilhoso. Com você aprendi o significado da palavra generosidade, respeito e amor verdadeiro. Enfim, obrigada Meu Bem por ser um príncipe de verdade, cheio de beleza e defeitos e por fazer parte da minha vida, pois amo partilhá-la com você.

À minha Helena, a luz dos meus olhos, agradeço por fazer, em tão pouco tempo, a minha vida mudar completamente, tornando tudo mais lindo e perfeito.

À minha orientadora, professora Cibele Q. da Silva, pela paciência e confiança depositados em mim para este trabalho. Agradeço por compatilhar comigo tanta sabedoria e pela enorme compreensão e amizade nos momento difíceis.

Aos professores e funcionários do Departamento de Estatística da Universidade de Brasília e da Universidade Federal do Ceará, em especial, ao professor Antônio Eduardo e à professora Silvia Maria de Freitas, pelas oportunidades proporcionadas e toda aprendizagem adquirida.

Aos colegas da turma do mestrado 2012.1 da Universidade de Brasília e aos colegas da turma de estatística 2007.1 da Universidade Federal do Ceará, pelos muitos momentos agradáveis.

# Sumário

<b>Agradecimentos</b>	<b>ii</b>
<b>Lista de Figuras</b>	<b>5</b>
<b>Lista de Tabelas</b>	<b>6</b>
<b>Lista de Códigos</b>	<b>7</b>
<b>Resumo</b>	<b>8</b>
<b>Abstract</b>	<b>9</b>
<b>Introdução</b>	<b>10</b>
<b>1 Distribuição Poisson Generalizada</b>	<b>12</b>
1.1 Distribuições de Probabilidade Lagrangianas . . . . .	12
1.1.1 Distribuições Lagrangianas Gerais . . . . .	13
1.1.2 Momentos da Distribuição Lagrangiana Geral . . . . .	14
1.2 Distribuição Poisson Generalizada . . . . .	15
1.2.1 Formulação . . . . .	16
1.2.2 Média e Variância . . . . .	18
1.2.3 Função Geradora de Momentos e de Cumulantes . . . . .	18
1.2.4 Estimação . . . . .	19
1.2.5 Simulação . . . . .	20
1.2.6 Exemplo . . . . .	25
<b>2 Modelos Dinâmicos Lineares</b>	<b>28</b>

2.1	Introdução . . . . .	28
2.2	Modelos de Espaço de Estados . . . . .	34
2.3	Modelos Dinâmicos Lineares . . . . .	41
2.3.1	Formulação . . . . .	43
2.4	Estimação dos estados e predição . . . . .	44
2.4.1	Filtragem . . . . .	46
2.4.2	Suavização . . . . .	48
2.4.3	Predição . . . . .	52
<b>3</b>	<b>Especificação do Modelo</b>	<b>61</b>
3.1	Análise da série temporal por componentes . . . . .	61
3.1.1	Modelos Polinomiais . . . . .	63
3.1.2	Modelos Sazonais . . . . .	69
3.1.3	Modelos de Regressão . . . . .	79
3.2	Modelo com parâmetros desconhecidos . . . . .	82
3.2.1	Inferência Bayesiana Conjugada para matrizes de variância desconhecidas . . . . .	83
3.2.2	Especificação da matriz de variância $\mathbf{W}_t$ . . . . .	86
<b>4</b>	<b>Modelos Dinâmicos para a Distribuição Poisson Generalizada</b>	<b>87</b>
4.1	Modelo Dinâmico para a distribuição Poisson Generalizada (MDPG)	89
4.2	Cálculos para obtenção das distribuições condicionais completas a <i>posteriori</i> . . . . .	92
4.2.1	A distribuição condicional completa do vetor de estados latentes $\boldsymbol{\theta}$ , $p(\boldsymbol{\theta} \boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi, \mathbf{y})$ . . . . .	94
4.2.2	As distribuições condicionais completas das matrizes de variância $\Phi_0$ e $\boldsymbol{\Phi}$ , $p(\Phi_0 \boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\Phi}, \phi, \mathbf{y})$ e $p(\boldsymbol{\Phi} \boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \phi, \mathbf{y})$ . . . . .	95
4.2.3	Distribuição condicional completa do parâmetro estrutural $\boldsymbol{\eta}$ , $p(\boldsymbol{\eta} \boldsymbol{\theta}, \Phi_0, \boldsymbol{\Phi}, \phi, \mathbf{y})$ . . . . .	97
4.2.4	Distribuição condicional completa do parâmetro de forma $\phi$ , $p(\phi \boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \boldsymbol{\Phi}, \mathbf{y})$ . . . . .	98
4.3	Predição . . . . .	101

<b>5</b>	<b>Simulações e Aplicações</b>	<b>103</b>
5.1	Especificação do MDPG . . . . .	104
5.2	Ajuste dos dados simulados . . . . .	106

# Lista de Figuras

1.1	Função de probabilidade da Poisson Generalizada para $\lambda = 5$ e diferentes valores de $\phi$ . . . . .	17
1.2	Variáveis aleatórias simuladas segundo o modelo Poisson Generalizado para diferentes valores dos parâmetros $\lambda$ e $\phi$ . . . . .	24
1.3	Ajuste dos modelos Poisson Generalizado e Poisson padrão aos dados da Tabela (1.2). . . . .	27
2.1	(a) Totais mensais de passageiros em linhas aéreas internacionais entre 1949 a 1960; (b) Medições anuais de vazões do Rio Nilo em Ashwan entre 1871 e 1970; (c) Lucro trimestral da Johnson & Johnson entre 1960 e 1980; (d) Totais mensais de motoristas de carros na Grã-Bretanha mortos ou gravemente feridos entre 1960 a 1984. . . . .	30
2.2	Filtro linear com entrada $x(t)$ , saída $Y_t$ e função de transferência $\Psi(t)$	32
2.3	Processo estocástico $Y(t, \omega)$ com $t \in T = \{1, 2, \dots, 50\}$ e $\omega \in \Omega = \{1, \dots, 5\}$ . . . . .	35
2.4	Estrutura de dependência para um modelo de espaço de estados. . . . .	37
2.5	Ciclo de inferência sequencial para a estimação e predição. . . . .	46
2.6	Fluxo de informação para predição de $Y_{t+k}$ . . . . .	53
2.7	(a) Estimativas filtradas com razões sinal-ruído diferentes; (b) Estimativas Suavizadas e Filtradas com intervalo de credibilidade de 95%; (c) Predição das observações um passo à frente. . . . .	60
3.1	Trajetórias simuladas de acordo com um passeio aleatório para dois diferentes valores da razão sinal-ruído. . . . .	65
3.2	Estrutura dos efeitos sazonais e observações. . . . .	70

3.3	(a) Estimativas Filtradas e Suavizadas; (b) Previsão $k = 36$ passos à frente.	79
3.4	Ajustes do MDL de regressão e Modelo Padrão . . . . .	81
5.1	Série temporal com dados de contagem e $m = 10$ . . . . .	110
5.2	(a) Nível Real <i>versus</i> Nível Estimado; (b) Crescimento Linear Real <i>versus</i> Crescimento Linear Estimado; (c) Efeito Sazonal Real <i>versus</i> Efeito Sazonal Estimado. . . . .	112
5.3	(a) $\eta_t$ <i>versus</i> $\hat{\eta}_t$ – MDPG; (b) $\eta_t$ <i>versus</i> $\hat{\eta}_t$ – MDP; (c) $\lambda_t$ <i>versus</i> $\hat{\lambda}_t$ – MDPG; (d) $\lambda_t$ <i>versus</i> $\hat{\lambda}_t$ – MDP. . . . .	114
5.4	Gráficos relativos às amostras MCMC do parâmetro $\phi$ : (a) Histograma; (b) Box-Plot, onde a linha horizontal representa o valor verdadeiro de $\phi$ e as demais os limites do I.C. de 95%; (c) Amostras de $\phi$ ao longo das iterações; (d) Autocorrelação da amostra de $\phi$ . . . . .	115
5.5	Comparação da média real com média estimada via MDPG e MDP. . . . .	116
5.6	Gráficos das amostras MCMC e autocorrelações para as variância $V$ , $\mathbf{W}_1$ e $\mathbf{W}_2$ . . . . .	116
5.7	(a) Previsão $k = 6$ passos a frente com MDPG; (b) Previsão $k = 6$ passos a frente com MDP. . . . .	118
5.8	(a) Série temporal simulada do MDPG e $m = 20$ ; (b) Nível real <i>versus</i> nível estimado; (c) Crescimento linear real <i>versus</i> Crescimento linear estimado; (d) Efeito sazonal real <i>versus</i> Efeito sazonal estimado. . . . .	120
5.9	(a) Estimativas de $\eta_t$ via MDPG e MDP; (b) Estimativas de $\lambda_t$ via MDPG e MDP; (c) Estimativas de $\mu_t$ via MDPG e MDP. . . . .	121
5.10	Gráficos relativos às amostras MCMC do parâmetro $\phi$ : (a) Histograma; (b) Box-Plot, onde a linha horizontal representa o valor verdadeiro de $\phi$ e as demais os limites do I.C. de 95%; (c) Amostras de $\phi$ ao longo das iterações; (d) Autocorrelação da amostra de $\phi$ . . . . .	122
5.11	(a) Previsão a $k = 6$ passo a frente com o MDPG; (b) Previsão a $k = 6$ passos a frente com o MDP. . . . .	123

# Lista de Tabelas

1.1	Estimação dos parâmetros das amostras simuladas com DPG . . . . .	23
1.2	Contagem de tumores e frequência de pacientes com NF2 . . . . .	25
1.3	Estatísticas resumo para a contagem de número de tumores em pacientes com $NF2$ . . . . .	26
1.4	Medidas de Ajuste para seleção de modelo . . . . .	27
5.1	Estatísticas resumo das amostras MCMC de $\phi$ . . . . .	113
5.2	Medidas de qualidade de ajuste considerando previsões $k = 6$ passos à frente do MDPG e MDP . . . . .	117

# Lista de Códigos

5.1	Estimação dos parâmetros da distribuição Poisson generalizada . . . .	130
5.2	Simulação de amostras aleatórias do modelo Poisson generalizado . .	131
5.3	Estimação dos estados e Predição . . . . .	132
5.4	Ajuste do modelo com crescimento linear e efeito sazonal na forma de <i>Fourier</i> . . . . .	134
5.5	Comparação do MDL de regressão com o modelo de regressão padrão	135
5.6	Geração de dados com base na estrutura do MDPG . . . . .	136
5.7	Processo de estimação do MDPG via MCMC . . . . .	138

# Resumo

Nessa dissertação propomos um modelo dinâmico para casos em que a série temporal é composta por dados de contagem. O modelo dinâmico para a distribuição Poisson Generalizada combina a classe dos modelos dinâmicos condicionalmente Gaussianos que, por sua vez, fornece uma estrutura flexível, permitindo que os parâmetros da distribuição dos dados possam ser modelados via MDLs normais e o esquema MCMC que une o amostrador de Gibbs com o algoritmo de Metropolis-Hastings de modo a proporcionar a amostragem das distribuições condicionais completas a *posteriori*. Desse modo nossa metodologia é capaz de tratar dados discretos correlacionados no tempo sendo possível realizar a estimação dos estados latentes e previsão do desenvolvimento futuro.

**Palavras Chave:** *modelos dinâmicos lineares, distribuição Poisson generalizada, modelos dinâmicos condicionalmente Gaussianos.*

# Abstract

In this dissertation, we propose a dynamic model for cases in which the time series is composed of count data. The dynamic model for generalized Poisson distribution combines the class of conditionally Gaussian dynamic models, which in turn provides a flexible structure, allowing the parameters of the data distribution to be modeled via normal DLMS and the MCMC scheme that combines the Gibbs sampler with the Metropolis-Hastings algorithm to provide a complete sampling of the *posteriori conditional distributions*. Thus our methodology is capable of handling correlated discrete data in time and make the estimation of latent states and prediction of future development.

**key words:** *dynamic linear models, generalized Poisson distribution, conditionally Gaussian dynamic models.*

# Introdução

Os Modelos Dinâmicos Lineares (MDL) têm se apresentado uma alternativa bastante flexível e aplicável para o estudo de observações correlacionadas ao longo do tempo. Grande parte dessa versatilidade é devida à abordagem *Bayesiana* incorporada a esses modelos, que proporciona a modelagem de toda incerteza do fenômeno de interesse através de distribuições de probabilidade, além de permitir que o sistema de aprendizado sobre as observações se atualize à medida que mais informações são adicionadas ao estudo. Essa classe pode ser vista como um modelo de regressão cujos parâmetros variam com o tempo e estocasticamente.

Apesar de os Modelos Dinâmicos Lineares serem aplicáveis a uma diversidade de problemas, a estrutura desses modelos baseia-se na suposição de normalidade dos dados que, por sua vez, nem sempre é admissível. Estudos fundamentados nessa classe têm se estendido para dados não-normais, sendo que a abordagem mais empregada é aquela cuja distribuição das observações pertence à família exponencial. Essa vertente de estudo é corrente nos modelos dinâmicos, e é denominada de Modelos Dinâmicos Lineares Generalizados (MDLG). As principais referências sobre esse assunto são Migon (1984) e West et al. (1985). Outro texto na literatura que aborda e documenta muito bem as etapas e extensões dos MDLs, inclusive, o MDLG é o livro de West e Harrison (1997).

Neste trabalho, focamos numa modelagem mais ampla, no sentido de ser adaptável a diversas estruturas e características das observações. Isto é, nosso interesse reside em um modelo que acomode séries não-normais, que proporcione bom ajuste e seja de fácil implementação.

Um trabalho que permite o estudo de séries temporais que admite tais estruturas é o de Cargnoni et al. (1997), que sugerem a predição *Bayesiana* de séries multivariadas

multinomiais através dos modelos dinâmicos lineares normais. A ideia consiste em realizar transformações não lineares nos parâmetros do modelo e, fazendo uso da modelagem hierárquica introduzida por Gamerman e Migon (1993), fazer inferências baseadas no MDL tradicional. Estes autores definem tal metodologia por modelos dinâmicos condicionalmente Gaussianos.

Com base nesses modelos dinâmicos condicionalmente Gaussianos, Da-Silva e Migon (2012) propõem o modelo dinâmico hierárquico beta para séries temporais de taxas ou proporções com os parâmetros média e dispersão variando com o tempo e através das diversas séries.

Nesse contexto, sugerimos uma modelagem dinâmica para dados de contagem correlacionados ao longo do tempo, considerando que cada observação segue uma distribuição Poisson Generalizada, ou seja, construímos o Modelo Dinâmico para a distribuição Poisson Generalizada (MDPG) com o objetivo de disponibilizar uma estrutura flexível para dados dessa natureza, de modo a descrever as características latentes e realizar previsões de perspectivas futuras da série.

Em nosso modelo, consideramos que o parâmetro de locação da distribuição dos dados evolui com o tempo e, a partir de uma transformação apropriada, este é modelado via MDLs normais, o que proporciona mais facilidade nos cálculos inferenciais e na análise dos dados.

Essa dissertação está organizada da seguinte forma: No Capítulo 1, introduzimos a distribuição Poisson Generalizada e algumas das suas propriedades. Dedicamos o Capítulo 2 para descrever conceitos e o processo de inferência do modelo dinâmico linear. No Capítulo 3, abordamos sobre a especificação de um MDL e seus componentes. No Capítulo 4, introduzimos o Modelo Dinâmico para a distribuição Poisson Generalizada. Nesse capítulo, descrevemos, inicialmente, a estrutura do modelo, em seguida, apresentamos os passos necessários para a estimação via MCMC e, por fim, expomos as etapas para realizar as previsões futuras da série. No Capítulo 5 (Aplicação) discutimos dois cenários de séries temporais simuladas e ajustadas pelo MDPG. No Capítulo 6 fazemos algumas discussões sobre o processo de construção do MDPG e trabalhos futuros.

Os cálculos computacionais inerentes a cada uma dessas etapas foram implementados no *software* R e são exibidos no decorrer do texto.

# Capítulo 1

## Distribuição Poisson Generalizada

Em muitas aplicações é conveniente modelar dados de contagem através da distribuição Poisson, que depende apenas do parâmetro  $\lambda$ , o qual representa tanto a média quanto a variância. No entanto, alguns autores, como Consul (1989), sugerem o uso da distribuição Poisson Generalizada (DPG) como um modelo alternativo, e menos restritiva, uma vez que a variabilidade dos dados pode diferir da média, proporcionando a modelagem de uma ampla gama de fenômenos aleatórios. Neste capítulo introduzimos a DPG como um modelo membro da família de distribuições Lagrangianas e descreveremos suas propriedades de acordo com essa classe. O estudo da DPG é de primordial interesse nesta dissertação, visto que o foco principal deste trabalho é o desenvolvimento de um modelo dinâmico para a distribuição Poisson Generalizada de modo a fornecer uma nova ferramenta para a modelagem de séries temporais discretas.

### 1.1 Distribuições de Probabilidade Lagrangianas

A necessidade de encontrar distribuições de probabilidade para ajustar dados de contagem de modo a encontrar padrões específicos e resumir características importantes levou diversos pesquisadores a desenvolverem modelos discretos mais complexos que formam uma família de distribuições apropriadas para modelar um fenômeno aleatório de natureza variada.

Uma classe muito ampla e importante é a família de distribuições de probabili-

dade Lagrangianas que possui diferentes distribuições de probabilidade discretas que desfrutam de propriedades que se fundamentam na poderosa expansão de Lagrange (para mais detalhes veja Apêndice A). Esta classe de distribuições discretas foram extensivamente estudada por Consul e Shenton (1972) e mais atualmente por Consul e Famoye (2006).

A família de distribuições Lagrangianas pode ser dividida em três subclasses: distribuições Lagrangianas básicas, distribuições Lagrangianas delta e distribuições Lagrangianas gerais. Como os modelos probabilísticos destas subclasses pertencem a uma mesma família de distribuições, estes possuem relações entre si e com outros modelos.

Apesar de algumas relações entre estas subclasses serem importantes, nesta dissertação, abordaremos apenas as distribuições Lagrangianas gerais, que possui como membro a distribuição Poisson Generalizada, que é o principal foco deste capítulo.

### 1.1.1 Distribuições Lagrangianas Gerais

As distribuições Lagrangianas gerais baseiam-se em duas funções analíticas  $f(z)$  e  $g(z)$ , que são sucessivamente diferenciáveis em relação a  $z$  com  $g(0) \neq 0$ ,  $g(1) = 1$ ,  $f(1) = 1$  e

$$D^{x-1}\{(g(z))^x f'(z)\}|_{z=0} \geq 0 \quad \text{com } x \in \mathcal{N}, \quad (1.1)$$

onde  $D^{x-1}$  representa a derivada de ordem  $x - 1$  em relação a  $z$ .

A função geradora de probabilidade (fgp) da distribuição discreta Lagrangiana geral, considerando a transformação de Lagrange  $z = ug(z)$ , expandida em série de potências de  $u$  é

$$f(z) = f(\psi(u)) = \sum_{x=0}^{\infty} \frac{u^x}{x!} D^{x-1} \left\{ (g(z))^x f'(z) \right\}_{z=0}, \quad (1.2)$$

em que  $z = \psi(u) = \sum_{x=0}^{\infty} (u^x/x!) \{D_{x-1}(g(z))^x\}_{z=0}$  é considerada a fgp na distribuição Lagrangiana básica quando  $z = ug(z)$  é expandida em série de potências de  $u$ . Assim, a função de probabilidade desta distribuição geral é dada por

$$P(X = x) = \begin{cases} f(0), & \text{se } x = 0 \\ (1/x!) D^{x-1} \{(g(z))^x f'(z)\}_{z=0}, & \text{se } x \in \mathcal{N}. \end{cases} \quad (1.3)$$

Apesar de Consul e Famoye (2006) afirmar que as funções  $f(z)$  e  $g(z)$  não são necessariamente fgp's de alguma distribuição, é interessante notar que qualquer função geradora de probabilidade de uma variável aleatória discreta obedece às condições para dedução da distribuição Lagrangiana geral, que são,  $g(0) \neq 0$ ,  $g(1) = 1$  e  $f(1) = 1$ . Para mais detalhes sobre a fgp e suas principais propriedades vide Apêndice B.

### 1.1.2 Momentos da Distribuição Lagrangiana Geral

Seja  $z = \psi(u) = ug(z)$  uma série de potência de  $u$ , os momentos da distribuição Lagrangiana geral podem ser obtidos a partir da função geradora de probabilidade  $f(z) = f(\psi(u))$  tomando sucessivas diferenças desta função em relação a  $u$  e em seguida fazendo  $z = u = 1$  (veja Apêndice B). Portanto, considere inicialmente que a derivada de  $z$  com relação a  $u$  é

$$\begin{aligned}\frac{dz}{du} &= g(z) + ug'(z)\frac{dz}{du} \\ &= g(z)[1 - ug'(z)]^{-1}.\end{aligned}\tag{1.4}$$

Com base em (1.4), as duas primeiras derivadas de  $f(\psi(u))$  com respeito a  $u$  são

$$\frac{\partial f(\psi(u))}{\partial u} = f'(\psi(u))\frac{d\psi(u)}{du} = f'(z)\frac{dz}{du} = f'(z)g(z)[1 - ug'(z)]^{-1},$$

e

$$\begin{aligned}\frac{\partial^2 f(\psi(u))}{\partial u^2} &= \left[ \frac{f'(z)g(z)}{1 - ug'(z)} \right]' \\ &= \frac{[f'(z)g(z)]'(1 - ug'(z)) - f'(z)g(z)[1 - ug'(z)]'}{(1 - ug'(z))^2} \\ &= \frac{[f''(z)g(z) + f'(z)g'(z)](1 - ug'(z))\frac{dz}{du} + f'(z)g(z)[g'(z) + ug''(z)\frac{dz}{du}]}{(1 - ug'(z))^2} \\ &= \frac{[f''(z)g(z) + f'(z)g'(z)]g(z)}{(1 - ug'(z))^2} + \frac{f'(z)g(z)}{(1 - ug'(z))^2} \left[ g'(z) + ug''(z)\frac{dz}{du} \right].\end{aligned}$$

Fazendo  $z = u = 1$ , os dois primeiros momentos fatoriais  $\mu'_{(1)}$  e  $\mu'_{(2)}$  da distribuição Lagrangiana geral com base em (1.4) são dados por

$$\mu'_{(1)} = \mu' = f'(1 - g')^{-1}\tag{1.5}$$

$$\mu'_{(2)} = \frac{f'' + 2f'g'}{(1-g')^2} + \frac{f'g''}{(1-g')^3}, \quad (1.6)$$

em que  $f', g', f''$  e  $g''$  denotam os valores das derivadas de  $f(z)$  e  $g(z)$ , respectivamente, para  $z = 1$ . Assim, temos que

$$E(X) = \mu'_{(1)} \quad (1.7)$$

e

$$\begin{aligned} V(X) &= \mu'_{(2)} + \mu'_{(1)} - (\mu'_{(2)})^2 \\ &= \frac{f'g''}{(1-g')^3} + \frac{f'' + f'g' + f' - (f')^2}{(1-g')^1}. \end{aligned} \quad (1.8)$$

## 1.2 Distribuição Poisson Generalizada

Uma das mais relevantes distribuições de probabilidade discretas é a distribuição Poisson que geralmente é utilizada para modelar o número de eventos que ocorrem em um determinado intervalo de tempo ou espaço fixo, com função de massa de probabilidade dada por

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} I_{\{0,1,\dots\}}(x) \quad (1.9)$$

com  $\lambda > 0$  sendo a frequência média de ocorrências conhecida e constante ao longo do tempo. O parâmetro  $\lambda$  também representa a variância desta distribuição.

Um dos pressupostos desse modelo diz respeito à independência, isto é, o número de eventos durante intervalos sucessivos de tempo são independentes. Quando esta suposição não é assumida, a ocorrência de um evento, em um determinado período, influencia na ocorrência de outro evento, de modo que a média e a variância podem variar com tempo, assumindo valores distintos, ou seja, a média pode diferir da variância.

Para contornar essa divergência, diversas derivações da distribuição Poisson foram desenvolvidas, algumas destas podem ser encontradas em Johnson et al. (1992).

Uma generalização do modelo Poisson foi inicialmente proposta por Satterthwaite (1942) com o objetivo de modelar dados com a lei de probabilidade Poisson considerando pesos distintos para os diferentes eventos.

Algumas décadas depois, Consul e Jain (1973) apresentaram a distribuição Poisson Generalizada que sugere a inserção de um parâmetro no modelo Poisson, doravante denotado por  $\phi$ , com o objetivo de fazer com que a variância desta distribuição possa assumir qualquer valor real positivo menor, igual ou maior que a média, conforme  $\phi$  seja negativo, nulo ou positivo. Dessa forma, a distribuição Poisson Generalizada é capaz de ajustar dados de contagem de qualquer natureza. Neste modelo, a probabilidade de ocorrência de um evento em um certo período pode depender de outros eventos realizados anteriormente.

### 1.2.1 Formulação

A variável aleatória discreta  $X$  é dita ter *distribuição Poisson Generalizada (DPG)* definida sob os inteiros não negativos com parâmetros  $\lambda$  e  $\phi$  se

$$P_x(\lambda, \phi) = \begin{cases} \frac{\lambda(\lambda+\phi x)^{x-1}}{x!} e^{-\lambda-\phi x}, & x = 0, 1, 2, \dots, \\ 0, & \text{para } x > m \text{ se } \phi < 0, \end{cases} \quad (1.10)$$

onde  $\lambda > 0$ ,  $\max(-1, -\lambda/m) \leq \phi \leq 1$ , com  $m \geq 4$ , sendo o maior inteiro positivo para que  $\lambda + \phi x > 0$  quando  $\phi < 0$ . O parâmetro  $\lambda$  é um parâmetro de localização, enquanto  $\phi$  é um parâmetro de forma.

Para  $\phi = 0$ , o modelo em (1.10) reduz-se a distribuição Poisson padrão com parâmetro  $\lambda$ . A Figura 1.1 ilustra a forma da função de massa de probabilidade para  $\phi = -0,5; -0,1; 0; 0,1; 0,5$  e  $\lambda = 5$ . É interessante notar que, quanto maior o valor de  $\phi$  para  $\lambda$  fixo, mais platicúrtica é a forma da DPG.

Quando  $\phi < 0$ , considera-se um truncamento no modelo devido a  $P_x(\lambda, \phi) = 0$  para valores de  $x > m$  e a soma  $\sum_{x=0}^m P_x(\lambda, \phi)$  aproxima-se da unidade. Isso gera um erro de truncagem que é inferior a 0,5% quando  $m \geq 4$  sendo, geralmente, negligenciável em aplicações práticas.

Uma sugestão para eliminar esse erro de truncamento é multiplicar cada probabilidade  $P_x(\lambda, \phi)$  por  $[F_m(\lambda, \phi)]^{-1}$ , em que

$$F_m(\lambda, \phi) = \sum_{x=0}^m P_x(\lambda, \phi). \quad (1.11)$$

Lerner et al. (1997) provam por expansão de Lagrange que a soma da DPG é igual a 1. Detalhes sobre a especificação dos limites de  $\phi$  e o erro de truncagem dessa

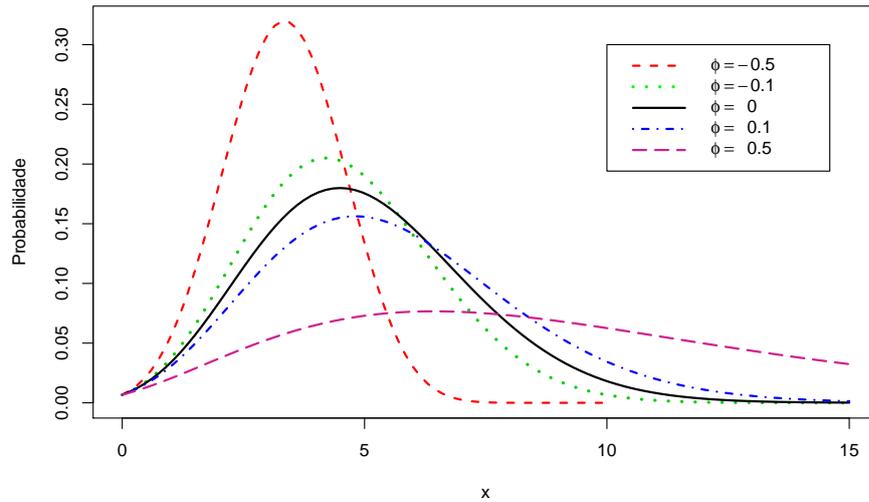


Figura 1.1: Função de probabilidade da Poisson Generalizada para  $\lambda = 5$  e diferentes valores de  $\phi$ .

distribuição podem ser encontrados em Consul e Shoukri (1985) e Ramalho (1996).

Quando  $f(z) = e^{\lambda(z-1)}$  e  $g(z) = e^{\phi(z-1)}$  são duas funções geradoras de probabilidade com  $\lambda > 0$  e  $0 < \phi < 1$ , temos, baseado em (1.3), que a DPG pertence à família de distribuições Lagrangianas gerais. Algumas propriedades da DPG podem ser derivadas a partir desta classe de distribuições. Consul (1989) estuda com detalhes as propriedades e apresenta algumas aplicações da DPG.

Aplicações dessa distribuição podem ser encontradas no processo de produção e restituição de cromossomos, mais especificamente em Janardan e Schaeffer (1977) (*apud* Consul e Famoye (2006)) que mostram que o número de aberrações cromossômicas aguardando a restituição na fila é uma variável aleatória com distribuição Poisson Generalizada. Satterthwaite (1942) e Denuit et al. (2007) utilizam a DPG para modelar dados relacionados a seguros de automóveis quando estar-se-á interessado no número de sinistros em um determinado período de tempo. Outras importantes aplicações desta distribuição estão disponíveis em Consul (1989).

## 1.2.2 Média e Variância

A função geradora de probabilidade da distribuição Poisson Generalizada é

$$f(z) = e^{\lambda(z-1)}, \quad \text{onde } z = ug(z) = ue^{\phi(z-1)}. \quad (1.12)$$

Note também, que  $f(z)$  é a fgp da distribuição Poisson padrão. Temos que  $f'(z) = \lambda e^{\lambda(z-1)}$ ,  $g'(z) = \phi e^{\phi(z-1)}$  e com base em (1.4)

$$\begin{aligned} \frac{dz}{du} &= e^{\phi(z-1)} [1 - u\phi e^{\phi(z-1)}]^{-1} \\ &= [e^{-\phi(z-1)} - u\phi]^{-1}. \end{aligned} \quad (1.13)$$

Portanto, temos que a média da DPG, segundo (1.5), é

$$E(X) = \frac{\lambda}{1 - \phi}, \quad (1.14)$$

e a variância da DPG, de acordo com (1.8), é:

$$V(X) = \frac{\lambda}{(1 - \phi)^3}. \quad (1.15)$$

Portanto, observa-se que para  $\phi < 0$ ,  $V(X) < E(X)$ , para  $\phi = 0$ ,  $V(X) = E(X)$  e para  $\phi > 0$ ,  $V(X) > E(X)$ .

Quando  $\phi$  é positivo com  $0 < \phi \leq 1$ , à medida que a média aumenta a variância também aumenta, no entanto a variância aumenta numa proporção bem maior devido ao seu denominador. Dessa forma, pode-se considerar o modelo Poisson Generalizado adequado para ajustar dados que apresentam sobredispersão quando a variância dos dados for relativamente grande.

## 1.2.3 Função Geradora de Momentos e de Cumulantes

Considere  $z = e^\beta$  e  $u = e^\alpha$  em (1.12). A função geradora de momentos (fgm) da DPG pode ser obtida a partir da fgp,

$$M_X(\beta) = f(e^\beta) = e^{\lambda(e^\beta - 1)}, \quad \text{onde } \beta = \alpha + \phi(e^\beta - 1). \quad (1.16)$$

Então, a função geradora de cumulantes (fgc) é dada por

$$K_X(\beta) = \ln M_X(\beta) = \ln f(e^\beta) = \lambda(e^\beta - 1), \quad \text{onde } \beta = \alpha + \phi(e^\beta - 1). \quad (1.17)$$

## 1.2.4 Estimação

Sejam  $X_1, X_2, \dots, X_n$  uma amostra aleatória da distribuição Poisson Generalizada dada em (1.10), os parâmetros  $\lambda$  e  $\phi$  serão estimados considerando os métodos dos momentos e de máxima verossimilhança (MV).

### Método dos Momentos

É conhecido que a média e a variância da DPG são

$$E(X) = \frac{\lambda}{1 - \phi} \quad \text{e} \quad \text{Var}(X) = \frac{\lambda}{(1 - \phi)^3}.$$

Comparando-se estes momentos com os momentos amostrais, tem-se que os estimadores para  $\lambda$  e  $\phi$  são obtidos como solução das equações

$$\frac{\tilde{\lambda}}{1 - \tilde{\phi}} = \frac{1}{n} \sum_{i=0}^n X_i = \bar{X}$$

e

$$\frac{\tilde{\lambda}}{(1 - \tilde{\phi})^3} + \frac{\tilde{\lambda}^2}{(1 - \tilde{\phi})^2} = \frac{1}{n} \sum_{i=0}^n X_i^2$$

que fornece os estimadores

$$\tilde{\lambda} = \sqrt{\frac{\bar{X}^3}{S^2}} \quad \text{e} \quad \tilde{\phi} = 1 - \sqrt{\frac{\bar{X}}{S^2}}, \quad (1.18)$$

com  $S^2 = \frac{1}{n} [\sum_{i=0}^n X_i^2 - n\bar{X}]$ .

### Estimação por Máxima Verossimilhança

O logaritmo da função de máxima verossimilhança (MV) para os parâmetros  $\lambda$  e  $\phi$  de uma amostra com distribuição Poisson Generalizada é dado por

$$l(\lambda, \phi|x) = n \ln \lambda + \sum_{i=1}^n (x_i - 1) \ln(\lambda + \phi x_i) - n\lambda - \phi \sum_{i=1}^n x_i - \sum_{i=1}^n \ln x_i!. \quad (1.19)$$

Para obter as estimativas de MV a partir da função de log-verossimilhança é necessário calcular as derivadas de  $l(\lambda, \phi|x)$  com respeito a  $\lambda$  e  $\phi$ . Entretanto, esta função não apresenta forma analítica explícita, de modo que é necessário recorrer a métodos

numéricos. Assim, segue que os estimadores de máxima verossimilhança para  $\lambda$  e  $\phi$  são, respectivamente, soluções das equações

$$\frac{n}{\hat{\lambda}} + \sum_{i=1}^n \frac{x_i - 1}{\hat{\lambda} + \hat{\phi}x_i} = n \quad (1.20)$$

$$\sum_{i=1}^n \frac{x_i(x_i - 1)}{\hat{\lambda} + \hat{\phi}x_i} = n\bar{X}. \quad (1.21)$$

Mas observe que  $E(X) = \lambda/(1-\phi)$  e  $\bar{X} \approx E(X)$  para  $n$  grande, pois a média amostral converge em probabilidade para  $E(X)$ . Desse modo,

$$\hat{\lambda} \approx \bar{X}(1 - \hat{\phi}). \quad (1.22)$$

Como consequência, temos que o estimador de máxima verossimilhança  $\hat{\lambda}$  de  $\lambda$  é uma função apenas da estimativa do parâmetro  $\phi$  e da média amostral. Assim, substituindo (1.22) na equação (1.21), temos que os estimadores MV  $\hat{\lambda}$  e  $\hat{\phi}$  dos parâmetros da distribuição Poisson Generalizada podem ser encontrados solucionando a equação

$$\sum_{i=1}^n \frac{x_i(x_i - 1)}{\bar{X} + \hat{\phi}(x_i - \bar{X})} = n\bar{X}. \quad (1.23)$$

De acordo com Consul e Shoukri (1984)(*apud* Consul e Famoye (2006)), quando a variância amostral é maior que a média amostral, as estimativas de máxima verossimilhança para  $\hat{\lambda} > 0$  e  $\hat{\phi} > 0$  são únicas. Consul e Famoye (1988) (*apud* Consul e Famoye (2006) ) provam que se a variância amostral é menor que a média amostral, as estimativas MV de  $\hat{\lambda} > 0$  e  $\hat{\phi} < 0$  também são únicas.

O Código 5.1 no Apêndice D.1 mostra os detalhes para estimação dos parâmetros  $\lambda$  e  $\phi$  da distribuição Poisson Generalizada pelo método dos momentos e máxima verossimilhança no *software* R.

### 1.2.5 Simulação

Existem vários métodos de simulação para a amostragem da distribuição Poisson Generalizada. Famoye (1997) cita alguns deles e aponta que a simulação do modelo DPG é muito importante na comparação dos diferentes métodos de estimação.

Em nosso trabalho, para simular valores da variável aleatória  $X$  com distribuição de probabilidade dada em (1.10) consideramos dois métodos distintos, o da transformada inversa e o método da ramificação. A seleção de ambos os procedimentos de simulação deve-se às restrições relacionadas ao suporte do parâmetro  $\phi$ , que requer a especificação de um método de simulação quando  $\phi < 0$  e outro quando  $\phi \geq 0$ .

### Método da Transformada Inversa

No processo de geração de variáveis aleatórias com distribuição Poisson Generalizada, o método da inversão revelou-se um procedimento razoável quando  $\lambda > 0$  e  $\max\{-1, -\frac{\lambda}{m}\} \leq \phi < 0$ . Observamos que a especificação correta do elemento  $m$  é determinante para a obtenção de uma amostra condizente com a DPG, ou seja, além de observar a restrição de  $m \geq 4$ , deve-se tomar  $m = \max\{X_1, \dots, X_n\}$  em que  $n$  é o tamanho da amostra.

Para valores de  $\phi \geq 0$ , o método da transformada inversa não proporciona resultados eficazes, principalmente quando  $\phi$  aproxima-se 1. Abaixo descrevemos o esquema de simulação da DPG via método da transformada inversa.

Seja  $U$  uma variável com distribuição uniforme no intervalo  $(0, 1)$ , segue que

$$X = \begin{cases} x_0 & \text{se } U < p_0^* \\ x_1 & \text{se } p_0^* \leq U < p_0^* + p_1^* \\ x_2 & \text{se } p_0^* + p_1^* \leq U < p_0^* + p_1^* + p_2^* \\ \vdots & \\ x_j & \text{se } \sum_{i=1}^{j-1} p_i^* \leq U < \sum_{i=1}^j p_i^* \\ \vdots & \end{cases}$$

em que  $p_0^*, p_1, \dots, p_j^*$  são as probabilidades da distribuição Poisson Generalizada ponderadas pelo o fator de correção  $F_m(\lambda, \phi)$ , ou seja

$$p_j^* = \frac{p_j}{F_m(\lambda, \phi)}, \quad j = 0, 1, \dots$$

Essa ponderação é considerada em consequência de  $\phi < 0$  quando  $x_j > m$  e garante que  $\sum_j p_j^* = 1$ .

Logo, como é conhecido que para  $0 < a < b < 1$ , a  $P(a \leq U < b) = b - a$ , então

$$P(X = x_j) = P\left(\sum_{i=1}^{j-1} p_i^* \leq U < \sum_{i=1}^j p_i^*\right) = p_j^*,$$

de modo que  $X$  tem distribuição Poisson Generalizada.

## Método da Ramificação

Seja  $X_i$ ,  $i = 0, \dots, n$ , o número de unidades que adquiriram uma particular característica na  $i$ -ésima geração de um processo de ramificação. Suponha que cada indivíduo em  $X_i$  torna-se um propagador da característica e seja  $X_j$ , para  $i \neq j$ , o número de indivíduos que receberam a característica da  $i$ -ésima geração. Considerando que o processo de propagação se dá até o tempo  $n$ , o número total de indivíduos que adquiriram a característica até o momento é dado por  $Y_n$ , em que

$$Y_n = X_0 + X_1 + \dots + X_n,$$

de modo que a variável aleatória  $X_0$  tem distribuição Poisson com parâmetro  $\lambda$  e  $X_1, \dots, X_n$  seguem o modelo Poisson com parâmetro  $\phi$ .

Consul e Shoukri (1988) mostraram que a distribuição de probabilidade de  $Y_n$ , número total de indivíduos que adquiriram a característica até o momento, quando não há mais propagação, tem distribuição Poisson Generalizada dada em (1.10) com  $\lambda > 0$  e  $0 \leq \phi < 1$ .

Os passos para gerar uma amostra da DPG a partir desse método é esboçado a seguir:

1. Gere  $Y$  de uma distribuição Poisson com média  $\lambda$ ;
2. Faça  $X = Y$ ;
3. Se  $Y \leq 0$ , retornar valor de  $X$ ;
4. Caso contrário, gere  $Z$  de uma distribuição Poisson com parâmetro  $\lambda Y$  e faça  
 $X = X + Z$ ;
5. Faça  $Y = Z$ ;
6. Se  $Y \leq 0$ , retornar valor de  $X$ , caso contrário volte para o passo 4.

Quando da implementação desse procedimento no *software* R, notamos que para valores de  $\phi$  próximos de 1, a amostra calculada exibiu um comportamento distinto da distribuição teórica, apresentando observações com valores bem elevados. Algumas manipulações computacionais foram realizadas para tentar corrigir esse problema, contudo verificamos que a própria distribuição apresenta limitações, sobretudo com

respeito ao parâmetro  $\phi$  e, em geral, tal procedimento de simulação para geração de variáveis aleatórias DPG é eficiente.

### Geração de Variáveis Aleatórias DPG no R

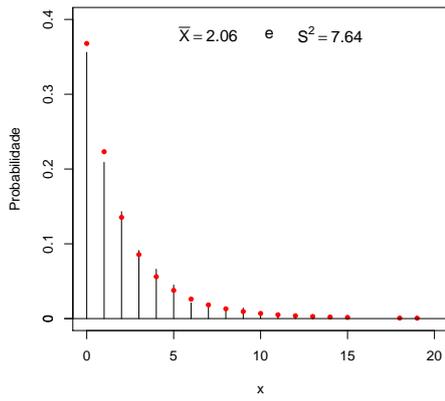
A título de ilustração, foram simuladas no *software* R, seis variáveis aleatórias de acordo com o modelo Poisson Generalizado com tamanho amostral  $n = 1000$  e diferentes valores dos parâmetros  $\lambda$  e  $\phi$ . A amostra 1 foi gerada com parâmetros  $\lambda = 1$  e  $\phi = 0,5$ , a amostra 2 com  $\lambda = 3$  e  $\phi = 0,4$ , amostra 3 com  $\lambda = 5$  e  $\phi = 0,75$ , amostra 4 com  $\lambda = 8$  e  $\phi = -0,1$ , a amostra 5 com  $\lambda = 10$  e  $\phi = 0$  e amostra 6 com  $\lambda = 30$  e  $\phi = -1$ .

Para cada variável simulada, calculamos as estimativas dos parâmetros  $\lambda$  e  $\phi$  pelo método dos momentos e de máxima verossimilhança. Ambos os métodos proporcionaram bons resultados, no sentido de que, as estimativas obtidas apresentaram valores bem próximos dos parâmetros reais. A Tabela 1.1 mostra os resultados destas estimações.

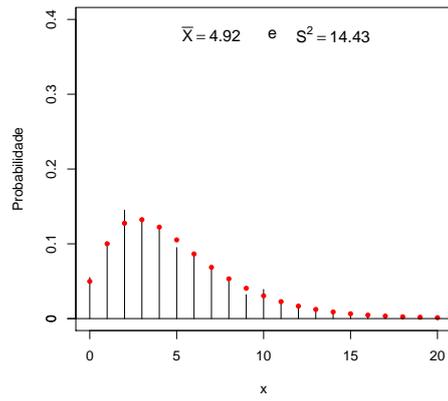
Tabela 1.1: Estimação dos parâmetros das amostras simuladas com DPG

Amostra	Valores Reais		Estimativas			
	$\lambda$	$\phi$	$\tilde{\lambda}$	$\tilde{\phi}$	$\hat{\lambda}$	$\hat{\phi}$
1	1	0,50	1,07	0,48	1,05	0,49
2	3	0,40	2,87	0,42	2,90	0,41
3	5	0,75	4,92	0,76	5,10	0,75
4	8	-0,10	7,78	-0,47	7,82	-0,48
5	10	0	10,04	0	10,04	0
6	30	-1,00	29,37	-0,96	29,44	-0,97

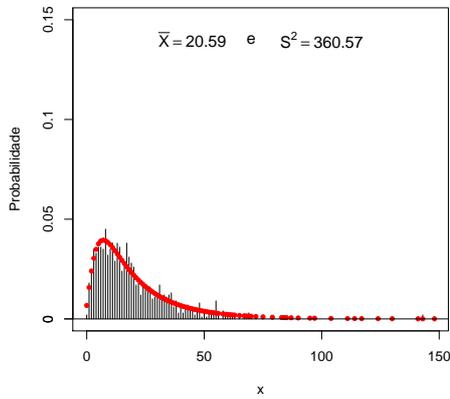
A Figura 1.2 apresenta os gráficos de barras dessas amostras simuladas com as respectivas probabilidades teóricas da DPG. A partir destes gráficos nota-se que existe uma boa concordância entre os dados simulados e o modelo teórico. É possível observar, também, que para valores do parâmetro  $\phi$  próximos de  $-1$ , a variância é menor que a média, enquanto para valores de  $\phi$  próximos de  $1$ , a variância é maior que a média. Para  $\phi = 0$ , a variável Poisson Generalizada reduz-se à distribuição Poisson padrão, cuja variância é aproximadamente igual a média.



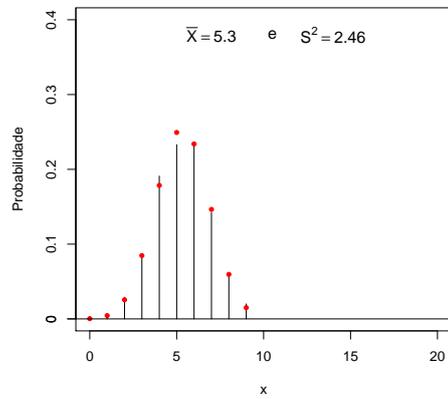
(a) Amostra 1 ( $\lambda = 1$ ,  $\phi = 0,5$ )



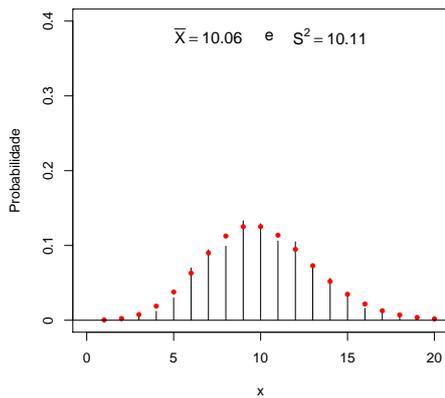
(b) Amostra 2 ( $\lambda = 3$  e  $\phi = 0,4$ )



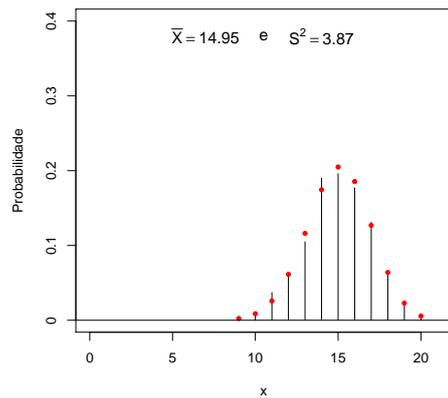
(c) Amostra 3 ( $\lambda = 5$  e  $\phi = 0,75$ )



(d) Amostra 4 ( $\lambda = 8$  e  $\phi = -0,1$ )



(e) Amostra 5 ( $\lambda = 10$  e  $\phi = 0$ )



(f) Amostra 6 ( $\lambda = 30$  e  $\phi = -1$ )

Figura 1.2: Variáveis aleatórias simuladas segundo o modelo Poisson Generalizado para diferentes valores dos parâmetros  $\lambda$  e  $\phi$ .

Quando o parâmetro  $\phi$  é positivo, a variância aumenta mais rapidamente que a média, de modo que o modelo Poisson Generalizado pode ser apropriado para ajustar dados com a presença de sobredispersão.

O Código (5.2) no Apêndice D.1 mostra as etapas para a realização da simulação no R, que consiste basicamente em três passos:

1. Criar uma função que retorne a DPG;
2. Criar outra função que retorne a função de distribuição acumulada da Poisson Generalizada;
3. Se  $\phi < 0$ , então desenvolver o algoritmo de simulação que retorna uma amostra aleatória da DPG fazendo uso do método da transformada inversa;
4. Caso  $\phi \geq 0$ , desenvolver o algoritmo de simulação que retorna uma amostra aleatória da DPG fazendo uso do método da ramificação.

### 1.2.6 Exemplo

Para ilustrar a utilidade da distribuição Poisson Generalizada, consideramos o conjunto de dados de contagem apresentados por Joe e Zhu (2005) que contabiliza o número de tumores em pacientes com neurofibromatose tipo 2 (*NF2*). Esta é uma doença hereditária de transmissão autossômica, ou seja, pode afetar qualquer sexo e em 50% dos casos, mutações germinais são herdadas dos genitores. A Tabela 1.2 apresenta a frequência de pacientes com *NF2* por registro de tumores entre  $N = 158$  pacientes que tomaram parte do estudo.

Tabela 1.2: Contagem de tumores e frequência de pacientes com *NF2*

Número de Tumores	0	1	2	3	4	5	6	7	8	9	10
Frequência	70	13	15	6	7	5	9	9	1	2	5
Número de Tumores	11	13	14	15	16	20	21	24	26	30	50
Frequência	1	1	1	1	1	3	1	3	1	1	2

Fonte: Joe e Zhu (2005)

Note que à medida que o número de tumores aumenta, o número de pacientes diminui. Essa relação é justificável em razão de que a *NF2* é uma doença rara, com

uma incidência estimada de 1 caso para cerca de 40000 nascidos vivos. A distribuição Poisson Generalizada pode ser uma boa alternativa para ajustar tais dados, uma vez que tem sido bastante utilizada na modelagem de experimentos com eventos raros, e é aplicável a uma diversidade maior de dados do que a distribuição Poisson, visto que a média não é necessariamente igual a variância. A Tabela 1.3 apresenta algumas estatísticas descritivas dos dados sob estudo.

Tabela 1.3: Estatísticas resumo para a contagem de número de tumores em pacientes com *NF2*

Mínimo	Mediana	Média	Desvio Padrão	Máximo
0	1	4,34	7,89	50

Entre os  $N = 158$  pacientes obteve-se uma média igual a  $\bar{X} = 4,34$  tumores por paciente e desvio padrão  $S = 7,89$ . Observa-se que a média amostral é menor que a variância, o que pode indicar que o modelo Poisson Generalizado com parâmetro  $\phi$  positivo pode ser uma boa alternativa para descrever os dados.

As estimativas pelo método de momentos dos parâmetros  $\lambda$  e  $\phi$  da distribuição Poisson Generalizada são  $\tilde{\lambda} = 1,143$  e  $\tilde{\phi} = 0,736$ .

Para a obter as estimativas pelo método da máxima verossimilhança (MV) é necessário estimar os parâmetros do modelo Poisson Generalizado numericamente. Como a EMV de  $\lambda$  depende da estimativa de  $\phi$ , é necessário solucionar apenas a equação (1.23) para encontrar ambas as estimativas. Logo, utilizando como valor inicial a estimativa  $\tilde{\phi}$  obtida pelo método de momentos e tomando  $m = 50$ , temos que as EMV's de  $\lambda$  e  $\phi$  são, respectivamente,  $\hat{\lambda} = 0,913$  e  $\hat{\phi} = 0,789$ .

Embora a diferença entre a média e a variância seja grande, ajustamos também a distribuição Poisson padrão aos dados, com o objetivo de comparar os resultados com os obtidos pela DPG. De acordo com o ajuste do modelo Poisson tradicional, a estimativa de máxima verossimilhança de  $\lambda$  é 4,34. A Figura 1.3 ilustra os ajustes de ambos os modelos.

O gráfico nos fornece um indicativo de que a distribuição Poisson Generalizada é mais adequada do que o modelo Poisson padrão para descrever os dados em estudo. Para verificar tal afirmação, calculamos dois critérios de seleção de modelos, bastante utilizados na literatura, o Critério de Informação de Akaike (AIC) e o Critério de

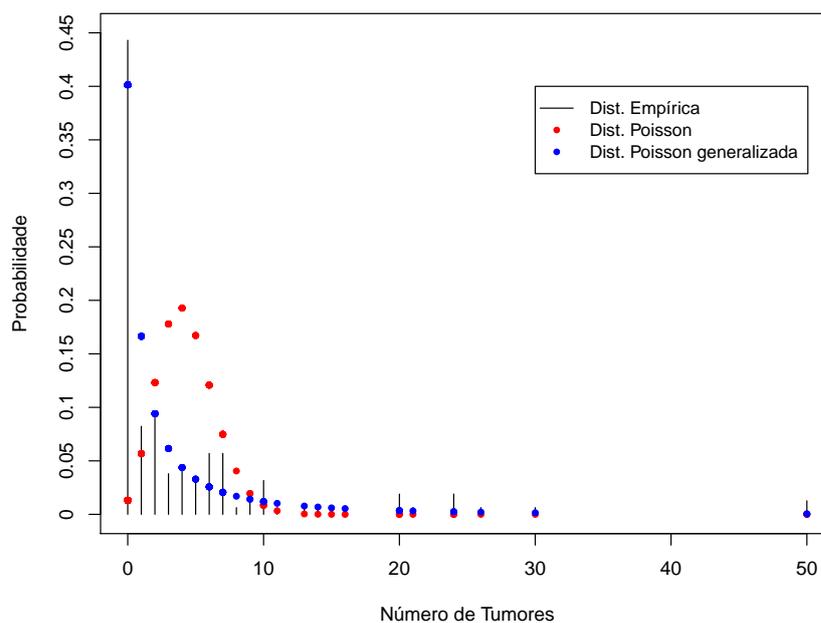


Figura 1.3: Ajuste dos modelos Poisson Generalizado e Poisson padrão aos dados da Tabela (1.2).

Tabela 1.4: Medidas de Ajuste para seleção de modelo

Modelo Ajustado	- log-verossimilhança	AIC	BIC
Poisson Generalizado	374,4	752,8	758,9
Poisson	888,7	1779,3	1782,4

Informação *Bayesiano* (BIC). Os resultados são apresentados na Tabela 1.4.

Ambos os critérios de informação, AIC e BIC, indicam que a distribuição Poisson Generalizada proporciona, de fato, melhor ajuste do que a distribuição Poisson padrão.

# Capítulo 2

## Modelos Dinâmicos Lineares

Neste capítulo descrevemos uma revisão de conceitos fundamentais da teoria dos modelos de espaço de estados para a modelagem de processos de séries temporais numa abordagem *Bayesiana*. Em particular, introduzimos o modelo dinâmico linear (MDL) como um caso especial dos modelos de espaço de estados. A forma padrão do MDL é obtida sob a suposição de normalidade dos dados, de modo que o modelo é referenciado como modelo dinâmico linear normal. Também apresentaremos aspectos e propriedades relevantes deste modelo, a fim de compreender, detalhadamente, os mecanismos que nos levam ao interesse principal deste estudo, que é a estimação e predição.

### 2.1 Introdução

Qualquer coleção de observações ordenadas sequencialmente ao longo do tempo é denominada uma série temporal. A principal característica deste tipo de dados é que as observações apresentam correlação serial, dificultando a análise estatística que, em sua maioria, baseia-se na estrutura de independência dos dados. Assim, é necessário o uso de procedimentos específicos que modelem essa dependência e incorpore toda informação relevante dos dados.

Uma série temporal é dita ser contínua quando é observada em um intervalo de tempo contínuo  $T = \{t : t_1 < t < t_2\}$  ou, discreta, quando as observações são feitas em instantes específicos,  $T = \{1, 2, \dots, n\}$ . Geralmente o tempo entre as observações

são igualmente espaçados.

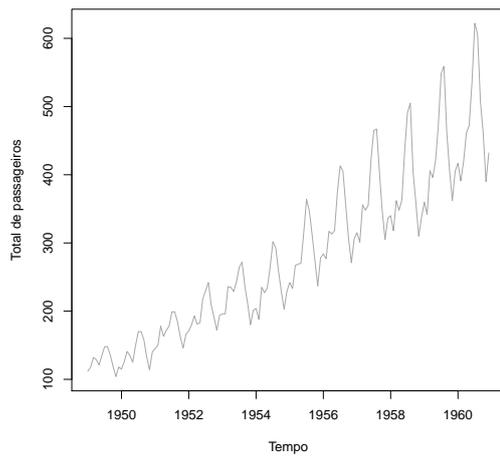
Segundo Morettin e Toloï (2006), na análise de séries temporais os principais interesses são:

- Descrever o comportamento da série através da verificação de padrões de tendência, da existência de componentes sazonais e cíclicos, observações discrepantes, mudanças estruturais e etc;
- Modelar a série temporal estimando os parâmetros e avaliando a qualidade de ajuste dos modelos propostos de forma a explicar o mecanismo gerador dos dados;
- Prever valores futuros da série temporal;
- Avaliar a adequação do modelo em termos de previsão;
- Monitorar a série com a finalidade de observar variações nas propriedades da mesma.

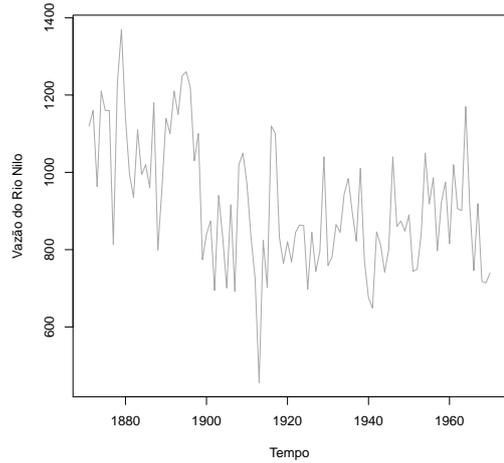
No conceito de estatística clássica, uma série temporal pode ser descrita através de três componentes: tendência, sazonalidade e um termo aleatório. Uma série apresenta tendência se existe algum crescimento ou decrescimento padrão. A sazonalidade aparece quando as observações tendem a repetir um dado comportamento a cada  $s$  períodos de tempo. Esta componente é dita *aditiva* quando a série exibe flutuações sazonais constantes, não dependendo do nível, e é dita *multiplicativa* quando as amplitudes sazonais variam com a tendência. O termo aleatório é denominado como erro ou ruído branco, e geralmente supõe-se que são não-correlacionados.

Para ilustrar a estrutura de uma série temporal e seus componentes, são apresentadas quatro séries disponíveis no *software* estatístico R, de domínio público, desenvolvido por “R Development Core Team”, acessível no site <http://cran.r-project.org>. As séries são: *AirPassengers*, *Nile*, *JohnsonJohnson* e *UKDriverDeaths*.

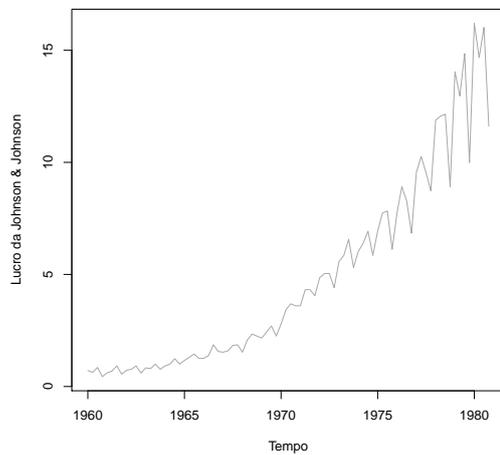
A Figura 2.1a mostra os totais mensais de passageiros em companhias aéreas internacionais entre 1949 a 1960. A série parece apresentar uma tendência crescente bem como um padrão sazonal mensal. A Figura 2.1b mostra as medições da vazão anual do rio Nilo em Ashwan entre 1871 a 1970. Parece haver uma mudança de



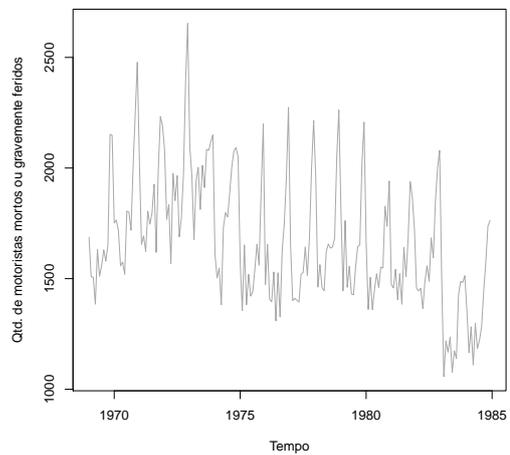
(a) *AirPassengers*



(b) *Nile*



(c) *JohnsonJohnson*



(d) *UKDriverDeaths*

Figura 2.1: (a) Totais mensais de passageiros em linhas aéreas internacionais entre 1949 a 1960; (b) Medições anuais de vazões do Rio Nilo em Ashwan entre 1871 e 1970; (c) Lucro trimestral da Johnson & Johnson entre 1960 e 1980; (d) Totais mensais de motoristas de carros na Grã-Bretanha mortos ou gravemente feridos entre 1960 a 1984.

nível na série, indicando que são necessários modelos mais flexíveis para incorporar tais modificações estruturais. A Figura 2.1c apresenta o lucro trimestral da empresa Johnson & Johnson entre 1960 a 1980. Aparentemente observa-se uma tendência de crescimento e uma variabilidade que aumenta com o tempo. Por fim, a Figura 2.1d apresenta a série de tempo que descreve os totais mensais de motoristas de carros

na Grã-Bretanha, mortos ou gravemente feridos, em janeiro de 1969 a dezembro de 1984. Observa-se também que esta série parece exibir mudanças de nível, com uma mudança ocorrendo, principalmente, no início de 1983, quando o uso obrigatório do cinto de segurança foi estabelecido.

Para fazer inferências estatísticas sobre a estrutura de um processo estocástico é necessário, em geral, fazer algumas suposições. Um dos mais importantes pressupostos é o de estacionariedade, que reflete a ideia de que as leis probabilísticas que regem o comportamento do processo não mudam ao longo do tempo, isto é, a série temporal se desenvolve aleatoriamente em torno de uma média constante. No entanto, a maioria das séries que encontramos na prática apresenta alguma forma de não-estacionariedade. Essa é uma das razões para se realizar transformações nos dados originais. A transformação mais comum consiste em tomar diferenças sucessivas da série original até a obtenção de uma série estacionária.

Outros motivos para transformar os dados são o de estabilizar a variância do processo, que pode aumentar ou diminuir à medida que o tempo evolui e tornar o efeito sazonal aditivo. Contudo, segundo alguns autores, as transformações não melhoram a qualidade da previsão e introduzem um vício em tais estimativas.

O problema está em modelar adequadamente as três componentes que, por sua vez são não observáveis, levando em conta a natureza dos dados. Os modelos utilizados para descrever essa estrutura são denominados processos estocásticos, isto é, processos regidos por leis probabilísticas. Para mais detalhes veja Morettin e Tolo (1986), Morettin e Tolo (2006) e Cryer e Chan (2008).

Existem na literatura diversas propostas de modelos que incorporam tais elementos e estruturas da série, uma abordagem bastante conhecida e utilizada são os modelos de Box e Jenkins (1970) ou modelos auto-regressivos integrados de médias móveis, ARIMA. Estes supõem que a série temporal é gerada por um filtro linear que converte uma série  $x_t$ , geralmente considerada puramente aleatória, em outra  $Y_t$  através de uma função de transferência, veja Figura 2.2. De modo mais simples,  $Y_t$  é uma combinação linear das variáveis de entrada  $x_t$ . É através desta nova variável que a análise de séries temporais é fundamentada.

Como em uma análise de regressão, a metodologia destes modelos segue um ciclo iterativo, onde os passos para a escolha do modelo, estimação dos parâmetros, previsão

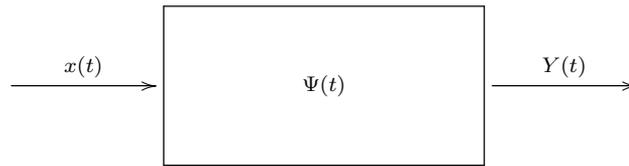


Figura 2.2: Filtro linear com entrada  $x(t)$ , saída  $Y_t$  e função de transferência  $\Psi(t)$

e diagnóstico são feitos sequencialmente. Os estágios desse procedimento são:

- i). De acordo com algumas análises descritivas da série, considera-se uma classe de modelos para a análise;
- ii). Com base nas funções de autocorrelação, autocorrelação parcial e demais critérios um modelo é identificado;
- iii). Em seguida, é feita a estimação dos parâmetros do modelo selecionado;
- iv). Por fim, a verificação do modelo ajustado é realizada, através de uma análise de resíduos, antes de usá-lo para fazer previsões.

Comumente, os modelos de Box e Jenkins são parcimoniosos uma vez que possuem um número pequeno de parâmetros e as previsões obtidas são bastante precisas. Todavia, uma das desvantagens dessa metodologia está na seleção de um modelo, uma vez que esta escolha depende também do critério de avaliação do pesquisador, visto que diferentes analistas podem especificar modelos distintos para a mesma série temporal. Uma alternativa para isso é investigar vários modelos e escolher aquele que forneça a melhor descrição dos dados com vista aos objetivos da análise.

Ocorre que sempre haverá um modelo que será mais coerente e adequado que outro, no sentido de proporcionar uma maior aprendizagem do sistema e previsão. Esta competitividade nos conduz a uma gama de modelos que podem representar eficientemente as observações em estudo, de modo que não existe um único modelo que podemos classificar como verdadeiro, ao contrário, existem diferentes visões acerca do sistema, de seus problemas e contextos. O desejável é que o modelo selecionado descreva, explique, organize as informações e experiências fornecendo processos com-

petentes de aprendizagem, que aumentam a compreensão dos dados, permitam decisões eficazes e melhoram o desempenho do processamento de informações do sistema, levando a condições e estruturas sólidas para se inferir sobre o futuro.

Muitas vezes as informações que os dados produzem não são suficientes para prever o futuro ou mudanças estruturais na série, de modo a identificar deficiências no modelo. Neste caso a intervenção de um especialista se faz necessária para a modelagem de tais dados, uma vez que tal profissional pode ter o conhecimento e experiência imprescindíveis sob o comportamento dos dados.

Em termos gerais, a estrutura do modelo deve ser capaz de incorporar toda a informação relevante contida nos dados e também informações externas que sejam indispensáveis para melhorar as estimativas, aprendizagem e previsão.

Uma metodologia que está sendo amplamente utilizada, pela sua flexibilidade na análise de séries temporais, uma vez que acomoda uma classe geral de modelos, cujos dados possuem natureza e estrutura variadas, são os modelos de espaço de estados. Tais modelos foram introduzidos por Kalman (1960), na área de engenharia, com o objetivo de estimar os estados de sistemas dinâmicos. Contudo, apesar de ter sido aplicada em diversas áreas, como, economia, física, oceanografia, esta metodologia só começou a ser explorada em econometria e estatística nos anos 70 (Harrison e Stevens, 1976), com modelos de regressão que apresentavam parâmetros com evolução temporal. O desenvolvimento do *filtro de Kalman* na análise de séries temporais consolidou-se nos anos 80, para mais detalhes veja (Harvey, 1989; West e Harrison, 1997).

A ideia básica dos modelos de espaço de estados é que qualquer modelo linear de séries temporais pode ser expresso na forma de espaço de estados, relacionando um vetor de observações  $Y_t$  a um vetor de erros  $v_t$  através de um processo de Markov  $\theta_t$ , denominado vetor de estados latentes do sistema. Assim, a estimação e predição dos estados latentes no tempo  $t$  podem ser feitas recursivamente, baseado no conjunto de informações disponíveis no instante  $t$ .

Com o avanço tecnológico nos anos 90, o uso dos modelos de espaço de estados tornou-se ainda mais viável, e a aproximação *Bayesiana* mostrou-se bastante favorável para o cálculo recursivo de tal metodologia. Trabalhos mais recentes para a análise de séries temporais baseados nos modelos de espaço de estados podem ser encontrados

em Durbin e Koopman (2001) e Petris et al. (2009).

A grande vantagem do modelo de espaço de estados é que este nos permite modelar processos univariados ou multivariados, estacionários ou não-estacionários, com mudanças estruturais e padrões irregulares sem a preocupação de fazer transformações nos dados originais.

Um caso especial deste procedimento são os modelos dinâmicos lineares. Estes modelos são bastante versáteis e são parametrizados de maneira a permitir mudanças estruturais, por exemplo, mudança no padrão de tendência ou sazonalidade, isto é, os parâmetros do modelo mudam ao longo do tempo. Quando essa metodologia é abordada com o enfoque *Bayesiano* propicia diversos benefícios, como: intervenção, filtragem e análise retrospectiva dos dados.

Estudaremos de forma abrangente um caso particular dessa metodologia que são os modelos dinâmicos lineares gaussianos que possuem propriedades relevantes e são tratáveis com maior facilidade. Abordaremos mais extensivamente estes modelos nas próximas seções.

## 2.2 Modelos de Espaço de Estados

Um processo estocástico é um conjunto de variáveis aleatórias  $\{Y(t), t \in T\}$  definidas em um conjunto  $T$  que pode ser enumerável ou não-enumerável.

Conforme James (2010), uma variável aleatória  $Y(t)$  em um espaço de probabilidade  $(\Omega, \mathcal{A}, \mathcal{P})$  é uma função real definida no espaço  $\Omega$ . Então, a variável aleatória  $Y(t)$  é, na realidade, uma função de  $t \in T$  e  $\omega \in \Omega$ , isto é,  $Y(t, \omega)$ .

Nestas condições, para cada  $t \in T$  fixo, temos uma variável aleatória  $Y(t, \omega) = Y_t(\omega)$  com uma distribuição de probabilidade e consideramos que tal distribuição é a mesma para todo  $t \in T$ .

Por outro lado, se  $\omega \in \Omega$  é fixado, temos uma função que depende apenas de  $t$  e, portanto, uma realização do processo, ou seja, uma série temporal.

Definindo cada trajetória  $Y(t, \omega)$ , para  $\omega \in \Omega$  fixo, como  $Y^\omega(t)$ , observa-se que esta é uma função de  $t$  não aleatória e que, para  $t$  fixo,  $Y^\omega(t)$  é um número real. Deste modo, um processo estocástico pode também ser pensado como o conjunto de todas as possíveis realizações ou trajetórias. A Figura 2.3 ilustra um processo estocástico

$Y(t, \omega)$  com  $t \in T = \{1, 2, \dots, 50\}$  e  $\omega \in \Omega = \{1, \dots, 5\}$ .

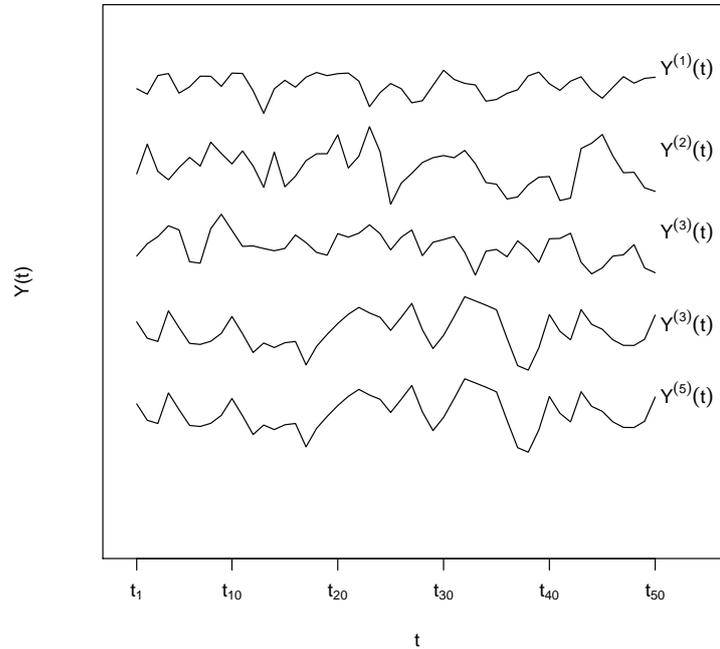


Figura 2.3: Processo estocástico  $Y(t, \omega)$  com  $t \in T = \{1, 2, \dots, 50\}$  e  $\omega \in \Omega = \{1, \dots, 5\}$ .

O conjunto  $E$  de todos os valores de  $\{Y(t), t \in T\}$  é denominado de espaço de estados do processo estocástico, em que o conjunto  $T$  é chamado de espaço do parâmetro e os valores assumidos por  $Y(t)$  são chamados de estados.

Se o conjunto  $T$  é discreto, então o processo estocástico é dito ser de tempo discreto. Se  $T$  é contínuo, então o processo é dito ser de tempo contínuo. O espaço de estados também pode ser discreto ou contínuo.

Portanto, sejam  $t_1, t_2, \dots, t_n$  elementos quaisquer de  $T$  e consideremos

$$F(y_1, \dots, y_n; t_1, \dots, t_n) = P(Y(t_1) \leq y_1, \dots, Y(t_n) \leq y_n). \quad (2.1)$$

Para caracterizar completamente o processo estocástico é necessário conhecer as distribuições finito-dimensionais em (2.1), para todo  $n \geq 1$ . Isto é, para  $n = 1$ , temos que conhecer as distribuições unidimensionais ou de primeira ordem de  $Y(t_1)$ ,  $t_1 \in T$ . Para  $n = 2$ , devemos conhecer as distribuições bidimensionais ou de segunda ordem  $(Y(t_1), Y(t_2))$ , onde  $t_1, t_2 \in T$ , e assim sucessivamente.

Caracterizar tal distribuição não é uma tarefa simples e como nesse estudo abordaremos apenas uma realização do processo estocástico  $Y(t)$  e estamos interessados apenas no caso em que os estados são discretos, descreveremos tal processo através de uma cadeia de Markov. Por questão de notação, considere a variável do processo definida, agora, como  $Y_t$ , onde  $t \in T = \{1, 2, \dots\}$ .

Um processo estocástico é dito ser uma cadeia de Markov de ordem 1, se o estado futuro depende apenas do estado atual e as variáveis aleatórias  $Y_t$ 's estão definidas no espaço de estados discreto. Ou seja,

$$P[Y_t = y_t | Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}, \dots, Y_1 = y_1] = P[Y_t = y_t | Y_{t-1} = y_{t-1}]. \quad (2.2)$$

Essa expressão significa que o estado  $Y_{t-1}$  resume as mesmas informações sobre  $Y_t$  que todas as demais variáveis,  $Y_1, Y_2, \dots, Y_{t-2}$ , sintetizam.

Apesar de a estrutura *Markoviana* ser uma forma simples de modelar a dependência entre os  $Y_t$ 's, esta nem sempre é apropriada, de modo que é conveniente fazer uso de modelos paramétricos. Ou seja, é mais simples expressar a lei probabilística que rege  $Y_1, Y_2, \dots, Y_t$  condicionalmente a alguma característica  $\theta_t$  do processo gerador dos dados. O modelo de espaço de estados define a estrutura de uma cadeia de Markov para os parâmetros  $\theta_t$ , isto é,  $\theta_t$  é considerado uma série temporal que auxilia na especificação da distribuição de probabilidade das variáveis aleatórias  $Y_t$ 's.

Os estados  $\theta_t$  são processos não observáveis e  $Y_t$  é considerado uma medida imprecisa de tais estados latentes, uma vez que depende de uma parte aleatória, também não observável: os erros.

A vantagem dessa modelagem é que esta assume uma estrutura de dependência simples que nos permite uma abordagem geral de qualquer modelo linear de série temporal.

Conforme Petris et al. (2009), o modelo de espaço de estados consiste em uma série temporal em  $\mathbb{R}^p$  ( $\theta_t : t = 0, 1, \dots$ ) e uma série temporal em  $\mathbb{R}^n$  ( $Y_t : t = 1, 2, \dots$ ), satisfazendo as seguintes suposições:

(A.1).  $(\theta_t, t = 0, 1, \dots)$  é uma cadeia de Markov;

(A.2). Condicional a  $\theta_t$ , os  $Y_t$ 's são independentes e  $Y_t$  depende apenas de  $\theta_t$ .

As consequências de ambas suposições é que a distribuição conjunta dos estados latentes  $\theta_t$  e das observações  $y_t$  é completamente especificada pela distribuição inicial  $p(\theta_0)$ , as densidades condicionais  $p(\theta_t|\theta_{t-1})$  e  $p(y_t|\theta_t)$ . Logo, para qualquer  $t > 0$

$$p(\theta_0, \theta_1, \dots, \theta_t, y_1, y_2, \dots, y_t) = p(\theta_0) \prod_{j=1}^t p(\theta_j|\theta_{j-1})p(y_t|\theta_t). \quad (2.3)$$

Essa estrutura, incorporada ao modelo de espaço de estados, é melhor representada na Figura 2.4.

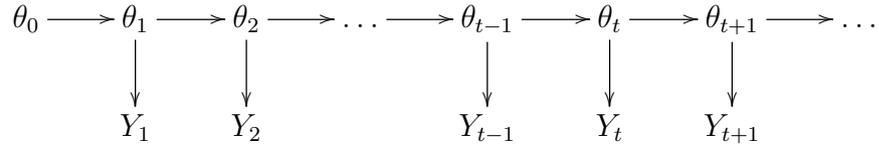


Figura 2.4: Estrutura de dependência para um modelo de espaço de estados.

Outros resultados fundamentais decorrentes dessa estrutura de independência condicional são dados a partir das propriedades listadas a seguir:

- (P.1).  $p(\theta_{t+1}|\theta_t, y_1, \dots, y_t) = p(\theta_{t+1}|\theta_t)$ ;
- (P.2).  $p(y_t|y_{t-1}, \dots, y_1, \theta_0, \theta_1, \dots, \theta_t) = p(y_t|\theta_t)$ ;
- (P.3).  $p(\theta_0, \theta_1, \dots, \theta_t, y_1, y_2, \dots, y_t) = p(\theta_0) \prod_{j=1}^t p(\theta_j|\theta_{j-1})p(y_t|\theta_t)$ ;
- (P.4).  $p(y_t, y_{t-1}, \dots, y_1|\theta_t, \theta_{t-1}, \dots, \theta_0) = \prod_{j=1}^t p(y_j|\theta_j)$ ;
- (P.5).  $p(\theta_t|\theta_{t+1}, y_1, \dots, y_t, y_{t+1}, \dots, y_T) = p(\theta_t|\theta_{t+1}, y_1, \dots, y_t)$ , em que  $t = \{1, \dots, T-1\}$ ;
- (P.6).  $p(\theta_t|\theta_{t+1}, \dots, \theta_T, y_1, \dots, y_t, y_{t+1}, \dots, y_T) = p(\theta_t|\theta_{t+1}, y_1, \dots, y_t)$ , em que  $t = \{1, \dots, j, j+1, \dots, T-1\}$ .

Para a demonstração dessas propriedades é razoável considerar um modelo de espaço de estados geral especificado por uma distribuição inicial  $p(\theta_0)$  e pelas seguintes equações

$$Y_t = h_t(\boldsymbol{\theta}_t, v_t) \quad (2.4a)$$

$$\boldsymbol{\theta}_t = g_t(\boldsymbol{\theta}_{t-1}, w_t) \quad (2.4b)$$

em que  $h_t$  e  $g_t$  são funções arbitrárias. Assume-se que  $(v_t)$  e  $(w_t)$  são erros aleatórios internamente e mutuamente independentes. Além disso, é suposto que  $\theta_0$  é independente de  $(v_t)$  e  $(w_t)$ .

**Prova:**

(P.1.) Considere a equação (2.4a), de maneira que pode-se escrever

$$p(\theta_{t+1}|\theta_t, y_1, y_2, \dots, y_t) = p(\theta_{t+1}|\theta_t, h_1(\theta_1, v_1), h_2(\theta_2, v_2), \dots, h_t(\theta_t, v_t)).$$

Desse modo, pela estrutura *Markoviana* assumida para o vetor de estados em (A.1), temos que

$$p(\theta_{t+1}|\theta_t, y_1, y_2, \dots, y_t) = p(\theta_{t+1}|\theta_t, h_t(\theta_t, v_t)).$$

Portanto, condicionalmente a  $\theta_t$ , a informação fornecida por  $\theta_1, \dots, \theta_{t-1}$  e, consequentemente, por  $Y_1, \dots, Y_{t-1}$  é irrelevante para descrever a distribuição de probabilidade de  $\theta_{t+1}$ . A Figura (2.4) ilustra bem tal estrutura de independência condicional. No entanto, como  $Y_t = h_t(\theta_t, v_t)$  é uma medida imprecisa do vetor de estados  $\theta_t$ , dado que possui uma parte aleatória não controlável  $v_t$ , a informação carregada por esta variável é insignificante quando na presença de  $\theta_t$  que resume completamente o conhecimento necessário do sistema para estimar  $\theta_{t+1}$ . Então, fica provado que

$$p(\theta_{t+1}|\theta_t, y_1, y_2, \dots, y_t) = p(\theta_{t+1}|\theta_t).$$

(P.2.) De modo análogo à prova da propriedade (P.1), podemos escrever a probabilidade condicional em P.2 como

$$\begin{aligned} p(y_t|y_{t-1}, \dots, y_1, \theta_t, \dots, \theta_0) &= p(y_t|(h_{t-1}(\theta_t, v_{t-1}), \dots, h_1(\theta_1, v_1), \theta_t, \dots, \theta_1, \theta_0)) \\ &= p(y_t|\theta_0, \theta_1, \dots, \theta_t). \end{aligned}$$

Dessa forma, dado a presença dos vetores de estados  $\theta_1, \dots, \theta_{t-1}$ , o conhecimento das observações  $Y_1, \dots, Y_{t-1}$  é irrelevante, uma vez que estas variáveis são medidas imprecisas de tais estados. Considere, ainda, pela propriedade (A.2) que,  $Y_t$  depende apenas de  $\theta_t$ . Baseado nesses argumentos, fica provado que

$$p(y_t|y_{t-1}, \dots, y_1, \theta_0, \dots, \theta_t) = p(y_t|\theta_t).$$

(P.3.) Escrevendo a distribuição conjunta dos  $\theta_0, \theta_1, \dots, \theta_t$  e  $Y_1, \dots, Y_t$  como um produto de distribuições condicionais, obtemos

$$\begin{aligned} p(\theta_0, \dots, \theta_t, y_1, \dots, y_t) &= p(y_t, y_{t-1}, \dots, y_1, \theta_t, \dots, \theta_0) \\ &= p(y_t|y_{t-1}, \dots, y_1, \theta_t, \dots, \theta_0)p(y_{t-1}|y_{t-2}, \dots, y_1, \theta_t, \dots, \theta_0) \\ &\dots p(y_1|\theta_t, \dots, \theta_0)p(\theta_t, \dots, \theta_0). \end{aligned}$$

Pela propriedade **(P.2)** podemos simplificar a expressão acima em

$$\begin{aligned} p(\theta_0, \dots, \theta_t, y_1, \dots, y_t) &= p(y_t|\theta_t)p(y_{t-1}|\theta_{t-1}) \dots p(y_1|\theta_1)p(\theta_t, \dots, \theta_0) \\ &= \prod_{j=1}^t p(y_j|\theta_j)p(\theta_t|\theta_{t-1}, \dots, \theta_0) \dots p(\theta_1|\theta_0)p(\theta_0). \end{aligned}$$

E, finalmente, utilizando a suposição **(A.1)**, temos que

$$\begin{aligned} p(\theta_0, \dots, \theta_t, y_1, \dots, y_t) &= \prod_{j=1}^t p(y_j|\theta_j)p(\theta_t|\theta_{t-1})p(\theta_{t-1}|\theta_{t-2}) \dots p(\theta_1|\theta_0)p(\theta_0) \\ &= p(\theta_0) \prod_{j=1}^t p(y_j|\theta_j)p(\theta_j|\theta_{j-1}). \end{aligned}$$

**(P.4).** Aplicando a definição de probabilidade condicional, provamos que

$$\begin{aligned} p(y_t, y_{t-1}, \dots, y_1|\theta_t, \theta_{t-1}, \dots, \theta_0) &= \frac{p(y_t, \dots, y_1, \theta_t, \dots, \theta_1, \theta_0)}{p(\theta_t, \dots, \theta_1, \theta_0)} \\ &= \frac{\left[ \prod_{j=1}^t p(y_j|\theta_j) \right] p(\theta_t, \dots, \theta_1, \theta_0)}{p(\theta_t, \dots, \theta_1, \theta_0)} \\ &= \prod_{j=1}^t p(y_j|\theta_j). \end{aligned}$$

**(P.5).** Para a prova desta propriedade e da **(P.6)** é necessário escrever os  $\theta'_j$ s e  $Y'_j$ s como funções dos erros  $w'_j$ s e  $v'_j$ s, de maneira que

- para  $j = 1$ :

$$\begin{aligned} \theta_1 &= g_1(\theta_0, w_1) = g'_1(w_1); \\ Y_1 &= h'_1(\theta_1, v_1) = h_1(g'_1(w_1), v_1) = h'_1(w_1, v_1); \end{aligned}$$

- para  $j = 2$ :

$$\begin{aligned} \theta_2 &= g_2(\theta_1, w_2) = g_2(g'_1(w_1), w_2) = g'_2(w_1, w_2); \\ Y_2 &= h_2(\theta_2, v_1) = h_2(g'_2(w_1, w_2), v_1) = h'_2(w_1, w_2, v_1); \\ &\vdots \end{aligned}$$

- para  $j = t$ :

$$\begin{aligned} \theta_t &= g_t(\theta_{t-1}, w_t) = g_t(g'_{t-1}(w_1, \dots, w_{t-1}), w_t) = g'_t(w_1, \dots, w_t); \\ Y_t &= h_t(\theta_t, v_t) = h_t(g'_t(w_1, \dots, w_t), v_t) = h'_t(w_1, \dots, w_t, v_t); \end{aligned}$$

- para  $j = t + 1$ :

$$\begin{aligned} \theta_{t+1} &= g_{t+1}(\theta_t, w_t); \\ Y_{t+1} &= h_{t+1}(\theta_{t+1}, v_{t+1}); \end{aligned}$$

- para  $j = t + 2$ :  
 $\theta_{t+2} = g_{t+2}(\theta_{t+1}, w_{t+2});$   
 $Y_{t+2} = h_{t+2}(\theta_{t+2}, v_{t+2}) = h'_{t+2}(\theta_{t+1}, w_{t+2}, v_{t+2});$
- para  $j = t + 3$ :  
 $\theta_{t+3} = g_{t+3}(\theta_{t+2}, w_{t+3}) = g'_{t+3}(\theta_{t+1}, w_{t+2}, w_{t+3});$   
 $Y_{t+3} = h_{t+3}(\theta_{t+3}, v_{t+3}) = h'_{t+3}(\theta_{t+1}, w_{t+2}, w_{t+3}, v_{t+3});$   
 $\vdots$
- para  $j = T$ :  
 $\theta_T = g_T(\theta_{T-1}, w_T) = g'_T(\theta_{t+1}, w_{t+2}, \dots, w_T);$   
 $Y_T = h_T(\theta_T, v_T) = h'_T(\theta_{t+1}, w_{t+2}, \dots, w_T, v_T).$

Então, a probabilidade condicional fica

$$p(\theta_t | \theta_{t+1}, y_1, \dots, y_t, y_{t+1}, \dots, y_T) = p(\theta_t | \theta_{t+1}, y_1, \dots, y_t, h_{t+1}(\theta_{t+1}, v_{t+1}), h'_{t+2}(\theta_{t+1}, w_{t+2}, v_{t+2}), \dots, h'_T(\theta_{t+1}, w_{t+2}, \dots, w_T, v_T)).$$

Como as observações  $Y_{t+1} = h_{t+1}(\theta_{t+1}, v_{t+1}), Y_{t+2} = h'_{t+2}(\theta_{t+1}, w_{t+2}, v_{t+2}), \dots, Y_T = h'_T(\theta_{t+1}, w_{t+2}, \dots, w_T, v_T)$  resumem as mesmas informações que  $\theta_{t+1}$ , adicionando somente informações sobre alguns erros que não influenciam na estimação de  $\theta_t$ , o conhecimento acerca destas observações é irrelevante quando na presença de  $\theta_{t+1}$ . Dessa maneira,

$$p(\theta_t | \theta_{t+1}, y_1, \dots, y_t, y_{t+1}, \dots, y_T) = p(\theta_t | \theta_{t+1}, y_1, y_2, \dots, y_t). \quad (2.5)$$

Ademais, em (2.5) nenhuma simplificação adicional pode ser possível, além das descritas abaixo:

$$\begin{aligned} p(\theta_t | \theta_{t+1}, y_1, \dots, y_t) &= \frac{p(\theta_{t+1} | \theta_t, y_1, \dots, y_t) p(\theta_t | y_1, \dots, y_t)}{p(\theta_{t+1} | y_1, \dots, y_t)} \\ &= \frac{p(\theta_{t+1} | \theta_t) p(\theta_t | y_1, \dots, y_t)}{p(\theta_{t+1} | y_1, \dots, y_t)}. \end{aligned}$$

**(P.6).** A prova desta propriedade é similar a da propriedade **(P.5)**, a diferença é que os estados  $\theta_{t+2}, \dots, \theta_T$  também estão fixos. Assim sendo, a probabilidade pode ser

escrita como

$$p(\theta_t | \theta_{t+1}, \theta_{t+2}, \dots, \theta_T, y_1, \dots, y_t, y_{t+1}, \dots, y_T) = p(\theta_t | \theta_{t+1}, g_{t+2}(\theta_{t+1}, w_{t+2}), \dots, \\ g'_T(\theta_{t+1}, w_{t+2}, \dots, w_T, v_T), y_1, \dots, y_t, h_{t+1}(\theta_{t+1}, v_{t+1}), h'_{t+2}(\theta_{t+1}, w_{t+2}, v_{t+2}), \\ \dots, h'_T(\theta_{t+1}, w_{t+2}, \dots, w_T, v_T)).$$

Note que os estados  $\theta_{t+2}, \dots, \theta_T$ , bem como as observações  $Y_{t+1}, Y_{t+2}, \dots, Y_T$ , carregam as mesmas informações sobre  $\theta_{t+1}$ , acrescentando apenas algumas informações sobre outros erros que são irrelevantes na medição de  $\theta_t$ . Logo, assim como pensado em (P.5), o conhecimento desses estados e observações são insignificantes, visto que  $\theta_{t+1}$  e  $y_1, y_2, \dots, y_t$  são suficientes para estimar  $\theta_t$ . Conseqüentemente,

$$p(\theta_t | \theta_{t+1}, \dots, \theta_T, y_1, \dots, y_t, y_{t+1}, \dots, y_T) = p(\theta_t | \theta_{t+1}, y_1, \dots, y_t).$$

## 2.3 Modelos Dinâmicos Lineares

Uma importante classe de modelos para análise de dados correlacionados ao longo do tempo é o modelo dinâmico linear numa abordagem *Bayesiana*. Este modelo pode ser considerado como uma generalização de um modelo de regressão linear simples adotando a premissa de que os parâmetros também evoluem temporalmente e estocasticamente.

Considere um modelo em que uma variável  $y$  está relacionada a uma outra variável  $X$  pela seguinte forma paramétrica

$$y = X\theta + \epsilon.$$

O objetivo em uma regressão linear é encontrar uma relação funcional entre a resposta  $y$  e a variável  $X$  de modo a conhecer características importantes de  $\theta$ . Contudo, como  $\theta$  é desconhecido, a incerteza a respeito deste parâmetro pode assumir magnitudes diversas. No paradigma *Bayesiano*, toda incerteza que se tem sobre o fenômeno em estudo, sintetizado por  $\theta$ , é descrito através de uma distribuição de probabilidade.

Nesta perspectiva, tal probabilidade tem uma interpretação subjetiva, uma vez que formaliza a informação incompleta que o pesquisador tem sob o evento ou fenômeno de interesse. Entretanto, como em modelos dinâmicos lineares a parametrização é

dinâmica, à medida que o tempo evolui mais informações vão sendo incorporadas ao sistema, visto que os dados estão sendo observados, de forma que estas podem ser utilizadas para rever os preceitos sobre a estrutura geral do modelo.

Deste modo, pode-se visualizar que o processo inferencial é sequencial, ou seja, concentra-se na ideia de que toda informação relevante para prever o futuro baseia-se na informação atual.

Então, considere que toda informação inicial disponível antes de observar os dados, quando  $t = 0$ , incluindo a própria definição do modelo, está concentrada no conjunto  $D_0$ . Assim sendo, para qualquer tempo  $t$ , toda afirmação sobre o futuro será baseada no conjunto  $D_t$ , que dispõe de toda informação histórica acerca dos dados até o momento atual. Este conjunto é suficiente para fazer previsões tanto para um futuro próximo como também de horizontes mais distantes.

Para simplificar, podemos denotar  $D_t = \{y_t, D_{t-1}\}$ , e caracterizamos esse conjunto como *fechado*, quando este não contém informações externas à série temporal, ou seja,  $D_t$  resume apenas as informações provenientes das observações. Porém, geralmente, existem situações em que informações advindas de fontes externas às observações devem ser introduzidas ao modelo, deste modo, definimos  $D_t = \{I_t, D_{t-1}\}$ , onde  $I_t$  indica toda informação adicional no tempo  $t$ .

De modo geral, pode-se sumarizar o processo de aprendizado do sistema por meio do modelo paramétrico:

$$p(y_t | \boldsymbol{\theta}_t, D_{t-1}) \quad (2.6)$$

em que  $\boldsymbol{\theta}_t$  é um vetor de parâmetros no tempo  $t$ , especificando que a parametrização é dinâmica. Portanto, o vetor de estados  $\boldsymbol{\theta}_t$  é representado estocasticamente por

$$p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, D_{t-1}). \quad (2.7)$$

Conforme foi discutido na Seção 2.2, ambas distribuições são suficientes para caracterizar a distribuição conjunta dos estados  $(\boldsymbol{\theta}_t)_{t \geq 0}$  e das observações  $(Y_t)_{t \geq 1}$ .

Com base nesse contexto, em um tempo  $t$  qualquer, a informação histórica  $D_{t-1}$  é sintetizada através de uma distribuição a *priori*  $p(\boldsymbol{\theta}_t | D_{t-1})$ . E em seguida, dado que  $Y_t$  foi observado, as informações contidas em  $D_t$  são condensadas na distribuição a *posteriori*  $p(\boldsymbol{\theta}_t | D_t)$ .

A inferência em modelos dinâmicos sob o aspecto *Bayesiano* segue os procedimentos padrões dessa abordagem. Além disso, tal estrutura fornece modelos robustos, com parâmetros significativos, viabiliza a previsão do desenvolvimento futuro da série temporal via distribuições de probabilidade, permite uma análise retrospectiva dos dados e incorpora informações de fontes externas relevantes, de modo que, a formulação probabilística atribuída a  $\boldsymbol{\theta}_t$  absorve, rapidamente, tal informação, adaptando-se às alterações especificadas e às incertezas associadas.

### 2.3.1 Formulação

Uma das classes mais importantes dos modelos de espaço de estados é o modelo dinâmico linear normal que será denominado apenas por modelo dinâmico linear (MDL). Tal modelo consiste, de acordo com a definição de modelos de espaço de estados, em uma série temporal  $Y_t$  com dimensão  $(n \times 1)$  e uma série temporal  $\theta_t$   $p$ -dimensional. Assim, segue que um MDL normal para um tempo  $t \geq 1$  é especificado pelas seguintes equações:

**Equação das observações:**

$$Y_t = \mathbf{F}'_t \boldsymbol{\theta}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim N(0, \mathbf{V}_t) \quad (2.8a)$$

**Equação do sistema:**

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(0, \mathbf{W}_t) \quad (2.8b)$$

**Informação Inicial:**

$$(\boldsymbol{\theta}_0 | D_0) \sim N(m_0, C_0),$$

em que

- $\mathbf{F}_t$  é uma matriz de delineamento conhecida com dimensão  $(n \times p)$ ;
- $\mathbf{G}_t$  é uma matriz de evolução dos estados  $(p \times p)$ ;
- $\boldsymbol{\mu}_t = \mathbf{F}'_t \boldsymbol{\theta}_t$  é a resposta média ou nível da série;
- $\mathbf{v}_t$  e  $\mathbf{w}_t$  são os erros observacionais e de evolução, respectivamente.

O erro  $\mathbf{v}_t$  é apenas uma perturbação aleatória no processo de medição que afeta a observação  $Y_t$ , porém não influencia a série temporal. Enquanto que, o erro  $\mathbf{w}_t$  influencia no desenvolvimento futuro do sistema. Tais erros  $\mathbf{v}_t$  e  $\mathbf{w}_t$  são assumidos mutuamente e serialmente não correlacionados com distribuição normal de média zero e matrizes de variância  $\mathbf{V}_t$  e  $\mathbf{W}_t$ , respectivamente. Além disso, é admitido que  $(\theta_0|D_0)$  é independente de  $\mathbf{v}_t$  e  $\mathbf{w}_t$ .

Observe que todos os componentes de ambas as equações são indexados ao tempo  $t$ , descrevendo portanto, uma estrutura de aprendizagem dinâmica e sequencial, uma vez que, a cada informação nova adicionada, o sistema é atualizado.

A equação de evolução em (2.8b) é caracterizada por um processo de Markov e é conveniente notar que à medida que o tempo evolui a incerteza que se tem sobre  $\theta_t$  também muda.

Toda informação inicial antes de observar os dados, quando  $t = 0$ , é especificada pela distribuição a *priori* normal com média  $m_0$  e variância  $C_0$ .

Nesse cenário, é razoável considerar que as equações (2.8a) e (2.8b) podem ser também definidas por

$$\begin{aligned} (Y_t|\boldsymbol{\theta}_t) &\sim N_n(\mathbf{F}'_t\boldsymbol{\theta}_t, \mathbf{V}_t), \\ (\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) &\sim N_p(\mathbf{G}'_t\boldsymbol{\theta}_{t-1}, \mathbf{W}_t). \end{aligned}$$

Uma propriedade derivada dessa estrutura é que qualquer combinação linear de MDLs normais é um MDL normal. Ademais, se a distribuição a *posteriori* de  $(\boldsymbol{\theta}_t|D_t)$  segue uma normal, a média e variância desta contém toda informação relevante sobre o futuro.

## 2.4 Estimação dos estados e predição

Em qualquer análise de séries temporais o principal objetivo é a estimação dos parâmetros do modelo e a predição. Uma vez que o modelo é postulado na forma de espaço de estados é possível empregar o filtro de Kalman, que é um procedimento recursivo com a finalidade de fazer inferências sobre a estrutura dos estados e previsão de observações futuras da série, baseado nas informações disponíveis. As estimativas

e previsões podem ser obtidas pelo cálculo das distribuições condicionais marginais dos estados e das observações, respectivamente.

A estimativa do vetor de estados é de interesse fundamental no estudo, visto que, a partir de tal vetor estimado, pode-se derivar as demais estimativas, bem como a previsão de valores futuros da série. Assim, é conveniente especificar que uma determinada distribuição condicional  $p(\boldsymbol{\theta}_s|D_t)$  é dita ser um *problema de filtragem* quando  $s = t$ , de *predição do estado latente* quando  $s > t$  e de *suavização* quando  $s < t$ .

West e Harrison (1997) descrevem a filtragem como um processo que utiliza dados recentes para revisar a inferência sobre os valores anteriores do vetor de estados, especificando  $(\boldsymbol{\theta}_{t-s}|D_t)$ , para  $s \geq 1$ , como a distribuição a  $s$ -passos filtrados. Enquanto Harvey (1989), denomina filtragem como a previsão do estado atual  $\boldsymbol{\theta}_t$  com base em uma determinada amostra das observações.

Neste trabalho, assim como em Petris et al. (2009), designamos filtragem como o processo utilizado para estimar o vetor de estados atual,  $\boldsymbol{\theta}_t$ , baseado nas informações condensadas até o instante  $t$ ,  $D_t$ . Simplificando, o processo de filtragem consiste na estimação de  $(\boldsymbol{\theta}_t|D_t)$ . Logo, à medida que novas informações tornam-se disponíveis, a estimativa é atualizada, incorporando todas as mudanças e incertezas vinculadas aos dados.

A *análise retrospectiva*, ou *suavização*, pode ser interpretada como a estimação do estado  $\boldsymbol{\theta}_s$  no tempo  $s$ , para  $s = t - 1, t - 2, \dots, 1$ , quando  $D_t$  está disponível, a fim de compreender melhor o comportamento dos dados e fazer previsões mais precisas.

A estimação do vetor de estados é apenas um meio para se obter a previsão do desenvolvimento futuro da série, que consiste no principal interesse no MDL. Portanto, podemos sintetizar o processo de inferência sequencial da seguinte maneira:

- Em um tempo  $t - 1$  dispomos da distribuição a *posteriori*  $p(\boldsymbol{\theta}_{t-1}|D_{t-1})$ ;
- Dado que conhecemos  $D_{t-1}$ , podemos resumir a informação contida neste conjunto através da distribuição a *priori*  $p(\boldsymbol{\theta}_t|D_{t-1})$  ;
- Baseado numa estimativa de  $\boldsymbol{\theta}_t$ , calculamos a distribuição preditiva  $p(y_t|D_{t-1})$ .
- Por conseguinte, uma vez que  $Y_t$  foi observado, o sistema é atualizado com a



(ii). A distribuição a priori de  $\boldsymbol{\theta}_t$  dado  $D_{t-1}$  é Gaussiana com parâmetros:

$$(\boldsymbol{\theta}_t|D_{t-1}) \sim N_p(\mathbf{a}_t, \mathbf{R}_t),$$

onde

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1} \quad \text{e} \quad \mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t.$$

(iii). Distribuição preditiva de  $Y_t$  dado  $D_{t-1}$  é Gaussiana com parâmetros:

$$(Y_t|D_{t-1}) \sim N_n(f_t, Q_t),$$

onde

$$f_t = \mathbf{F}_t' \mathbf{a}_t \quad \text{e} \quad Q_t = \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \mathbf{V}_t.$$

(iv). Distribuição filtrada, ou a posteriori, de  $\boldsymbol{\theta}_t$  dado  $D_t$  é Gaussiana com parâmetros:

$$(\boldsymbol{\theta}_t|D_t) \sim N_p(\mathbf{m}_t, \mathbf{C}_t),$$

com

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{R}_t \mathbf{F}_t Q_t^{-1} e_t \quad \text{e} \quad \mathbf{C}_t = \mathbf{R}_t - \mathbf{R}_t \mathbf{F}_t Q_t^{-1} \mathbf{F}_t' \mathbf{R}_t,$$

em que

$$\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t Q_t^{-1} \quad \text{e} \quad e_t = Y_t - f_t.$$

A prova dessas quatro etapas é obtida a partir da distribuição conjunta do vetor  $(\theta_0, \theta_1, \dots, \theta_t, Y_1, \dots, Y_t)$  dada em (2.3). Como trata-se de um MDL normal, temos que esse vetor tem distribuição Gaussiana multivariada, de maneira que as distribuições marginais também são normais. Para a prova do item (iv) é necessário aplicar a fórmula de Bayes.

Como a distribuição filtrada foi derivada pelo filtro de Kalman, uma estimativa pontual ótima é dada pela média de  $p(\boldsymbol{\theta}_t|D_t)$ , no sentido de que minimiza o erro quadrático médio (EQM). Conforme Harvey (1989), quando a suposição de normalidade é rejeitada, não existe garantias de que a estimativa pontual do filtro de Kalman será a média condicional de  $\boldsymbol{\theta}_t$  dado  $D_t$ , embora ainda pode ser considerado um estimador ideal, uma vez que apresenta o menor EQM na classe dos estimadores lineares.

A média e a variância da distribuição filtrada  $p(\boldsymbol{\theta}_t|D_t)$  podem ser reformuladas como  $\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t e_t$  e  $\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t Q_t \mathbf{A}_t'$ . Note que a estimativa média  $\mathbf{m}_t$  é uma soma da média a *priori*, quando  $Y_t$  ainda não tinha sido observado, mais uma correção que retrata o quanto a previsão  $f_t$  está distante da nova observação. O peso dessa correção é dado por  $\mathbf{A}_t$ , que depende da variância do erros observacionais  $\mathbf{V}_t$  através da variância da distribuição preditiva  $Q_t$ .

Observe, também, que podemos descrever  $\mathbf{m}_t$  como uma soma ponderada na forma  $\mathbf{m}_t = \mathbf{A}_t Y_t + (\mathbf{I} + \mathbf{A}_t \mathbf{F}_t') \mathbf{a}_t$ . O peso  $\mathbf{A}_t$  é denominado de *coeficiente adaptativo*, e este pode ser melhor interpretado quando o MDL é univariado, para mais detalhes veja West e Harrison (1997, cap.2).

Em um modelo de nível local, a razão entre as variâncias  $W$  e  $V$  exerce um papel importante na filtragem, uma vez que a trajetória do processo,  $Y_t$ , é fortemente influenciada por essa relação que é denominada de razão sinal-ruído. Quando a razão  $r = W/V$  é pequena, considera-se que grande parte do movimento da série é controlado pelas observações, enquanto que, quando  $r$  é grande, acredita-se que os movimentos da série são devidos tanto as observações quanto às variações no nível.

O filtro de Kalman pode ser implementado computacionalmente no *software* R pelo comando `dImFilter` do pacote `dIm`, que foi desenvolvido (Petris, 2010). Essa função utiliza os dados originais e um modelo ajustado, retornando os valores  $m_t, C_t, a_t, R_t$  e  $f_t$ . As matrizes de variância são expressas em termos de decomposições em valores singulares.

Na Figura 2.7 ilustramos o processo de filtragem para a série temporal *AirPassengers* do *software* R.

## 2.4.2 Suavização

Em diversas aplicações, é conveniente analisar o comportamento de observações passadas de uma série temporal, dispondo de toda a informação acessível, como forma de entender melhor o comportamento do sistema retrospectivamente.

Imagine uma situação hipotética em que o Departamento Estadual de Trânsito do Ceará dispõe de dados, que conta o número de acidentes de trânsito nas rodovias intermunicipais do estado, durante alguns anos e deseja estudar o motivo pelo qual

em determinados períodos ao longo de um ano, a incidência de acidentes é maior. O estatístico responsável pelo estudo observou que esses picos coincidem com feriados prolongados (como Carnaval, Páscoa, Natal e etc.), quando há um grande deslocamento de veículos para o litoral e interior do estado. Uma ferramenta muito útil para compreender melhor o comportamento desse sistema é a *análise retrospectiva* que permite fazer inferências sobre valores passados da série temporal.

Nesse sentido, é apropriado aplicar um procedimento recursivo para estimar o vetor de estados  $\boldsymbol{\theta}_t$ , considerando toda a informação contida na amostra. Indicaremos tal informação pelo conjunto  $D_T$ , em que  $t < T$ . O interesse reside em conhecer a distribuição condicional  $p(\boldsymbol{\theta}_t|D_T)$  a partir do tempo  $t = T - 1$ , em seguida  $t = T - 2$  e assim por diante.

Em termos gerais, a suavização consiste em construir, retrospectivamente, ou regressivamente, o comportamento do sistema, a fim de melhorar a compreensão acerca dos dados, viabilizando a tomada de decisões futuras.

Apresentamos, a seguir, alguns resultados importantes provenientes da estrutura de independência condicional do modelo de espaço de estados.

**Proposição 2.2.** *Considere um modelo de espaço de estados dado pelas suposições (A.1) e (A.2).*

(i). *A probabilidade de transição retrospectiva dos estados  $(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T)$  condicional a  $D_T$  é*

$$p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, D_T) = \frac{p(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t|D_t)}{p(\boldsymbol{\theta}_{t+1}|D_t)}.$$

(ii). *A distribuição suavizada  $p(\boldsymbol{\theta}_t|D_T)$  pode ser obtida pelo cálculo recursivo retrospectivo dado por*

$$p(\boldsymbol{\theta}_t|D_T) = p(\boldsymbol{\theta}_t|D_t) \int \frac{p(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t)}{p(\boldsymbol{\theta}_{t+1}|D_t)} p(\boldsymbol{\theta}_{t+1}|D_T) d\boldsymbol{\theta}_{t+1}.$$

**Prova:** Para a prova do item (i), considere as propriedades (P.5) e (P.1).

$$\begin{aligned} P(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, D_T) &\stackrel{(P.5)}{=} p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, D_t) \\ &= \frac{p(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t, D_t)p(\boldsymbol{\theta}_t|D_t)}{p(\boldsymbol{\theta}_{t+1}|D_t)} \\ &\stackrel{(P.1)}{=} \frac{p(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t|D_t)}{p(\boldsymbol{\theta}_{t+1}|D_t)}. \end{aligned}$$

Para a prova do item (ii), basta encontrar a distribuição marginal  $(\boldsymbol{\theta}_t|D_T)$  integrando  $p(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1}|D_T)$  com respeito a  $\boldsymbol{\theta}_{t+1}$ . Assim,

$$\begin{aligned} p(\boldsymbol{\theta}_t|D_T) &= \int p(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1}|D_T)d\boldsymbol{\theta}_{t+1} \\ &= \int p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, D_T)p(\boldsymbol{\theta}_{t+1}|D_T)d\boldsymbol{\theta}_{t+1} \\ &= \int \frac{p(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t|D_t)}{p(\boldsymbol{\theta}_{t+1}|D_t)}p(\boldsymbol{\theta}_{t+1}|D_T)d\boldsymbol{\theta}_{t+1} \\ &= p(\boldsymbol{\theta}_t|D_t) \int \frac{p(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t)}{p(\boldsymbol{\theta}_{t+1}|D_t)}p(\boldsymbol{\theta}_{t+1}|D_T)d\boldsymbol{\theta}_{t+1}. \end{aligned}$$

A prova de ambos os itens também pode ser encontrada em Petris et al. (2009).

Note que para realizar a análise retrospectiva é necessário antes conhecer a distribuição filtrada  $p(\boldsymbol{\theta}_t|D_t)$ , visto que o ponto de partida para este procedimento são os momentos da densidade a *posteriori*, isto é,  $\mathbf{s}_T = \mathbf{m}_T$  e  $\mathbf{S}_T = \mathbf{C}_T$ .

Devido ao pressuposto de normalidade do MDL em (2.8) e a suposição de conhecimento das matrizes de covariância  $\mathbf{V}_t$  e  $\mathbf{W}_t$ , a distribuição suavizada pode ser especificada apenas pela sua média e variância.

**Proposição 2.3.** *Para um MDL definido em (2.8), se  $(\boldsymbol{\theta}_{t+1}|D_T) \sim N_p(\mathbf{s}_{t+1}, \mathbf{S}_{t+1})$ , então  $(\boldsymbol{\theta}_t|D_T) \sim N_p(\mathbf{s}_t, \mathbf{S}_t)$ , onde*

$$\mathbf{s}_t = \mathbf{m}_t + \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} (\mathbf{s}_{t+1} - \mathbf{a}_{t+1}),$$

e

$$\mathbf{S}_t = \mathbf{C}_t - \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} (\mathbf{R}_{t+1} - \mathbf{S}_{t+1}) \mathbf{R}_{t+1}^{-1} \mathbf{G}_{t+1} \mathbf{C}_t.$$

Os cálculos recursivos desse algoritmo podem ser implementados no *software* R através do comando *dlmSmooth* do pacote *dlm*. Essa função pode ser executada a partir da saída do *dlmFilter* ou usando as próprias observações e o modelo ajustado. O *dlmSmooth* retorna uma lista com a média da distribuição suavizada  $s$  e a variância desta em termos de suas decomposições em valores singulares.

**Prova:** Para provar que a média  $\mathbf{s}_t$ , e a variância  $\mathbf{S}_t$ , da distribuição suavizada são tais como descritas acima, é necessário o uso de duas propriedades importantes de esperança condicional:

1.  $E(\boldsymbol{\theta}_t^k | D_T) = E[E(\boldsymbol{\theta}_t^k | \boldsymbol{\theta}_{t+1}, D_T) | D_T]$ ;
2.  $V(\boldsymbol{\theta}_t | D_T) = E[V(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_T) | D_T] + V[E(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_T) | D_T]$ .

A demonstração de ambas as propriedades são apresentadas no Apêndice C. Deste modo, temos que a esperança e a variância de  $\boldsymbol{\theta}_t$  dado  $D_T$  podem ser obtidas calculando

$$\mathbf{s}_t = E(\boldsymbol{\theta}_t | D_T) = E[E(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_T) | D_T]$$

e

$$\mathbf{S}_t = V(\boldsymbol{\theta}_t | D_T) = E[V(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_T) | D_T] + V[E(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_T) | D_T].$$

Baseado na propriedade **(P.5)**, sabe-se que  $p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_T) = p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_t)$ . De modo que, considerando o MDL dado em (2.8), a média e a variância da distribuição de  $(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_t)$  são obtidas a partir do cálculo da distribuição conjunta de  $\boldsymbol{\theta}_t$  e  $\boldsymbol{\theta}_{t+1}$  e pelas propriedades de distribuição condicional de variáveis aleatórias normais multivariadas. Assim, considere a distribuição conjunta de  $\boldsymbol{\theta}_t$  e  $\boldsymbol{\theta}_{t+1}$  dada a seguir

$$\left( \begin{array}{c} \boldsymbol{\theta}_t \\ \boldsymbol{\theta}_{t+1} \end{array} \middle| D_t \right) \sim N \left[ \left( \begin{array}{c} \mathbf{m}_t \\ \mathbf{G}_{t+1} \mathbf{m}_t \end{array} \right), \left( \begin{array}{cc} \mathbf{C}_t & \mathbf{C}_t \mathbf{G}'_{t+1} \\ \mathbf{G}_{t+1} \mathbf{C}'_t & \mathbf{R}_{t+1} \end{array} \right) \right].$$

Aplicando as propriedades da distribuição condicional da normal multivariada, tem-se que

$$(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_t) \sim N(\mathbf{b}_t, \mathbf{B}_t),$$

onde  $\mathbf{b}_t = \mathbf{m}_t + \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} (\boldsymbol{\theta}_{t+1} - \mathbf{G}_{t+1} \mathbf{m}_t)$  e  $\mathbf{B}_t = \mathbf{C}_t - \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} \mathbf{G}_{t+1} \mathbf{C}'_t$ . Para mais detalhes desta propriedade veja Johnson (2007).

Portanto, segue que

$$\begin{aligned} E(\boldsymbol{\theta}_t | D_T) &= E[E(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_T) | D_T] \\ &= E[E(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_t) | D_T] \\ &= E[\mathbf{m}_t + \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} (\boldsymbol{\theta}_{t+1} - \mathbf{G}_{t+1} \mathbf{m}_t) | D_T] \\ &= \mathbf{m}_t + \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} [E(\boldsymbol{\theta}_{t+1} | D_T) - \mathbf{G}_{t+1} \mathbf{m}_t], \end{aligned}$$

denotando  $\mathbf{s}_{t+1} = E(\boldsymbol{\theta}_{t+1}|D_T)$ , tem-se que

$$\mathbf{s}_t = E(\boldsymbol{\theta}_t|D_T) = \mathbf{m}_t + \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} (\mathbf{s}_{t+1} - \mathbf{a}_{t+1}),$$

com  $\mathbf{a}_{t+1} = \mathbf{G}_{t+1} \mathbf{m}_t$ . Analogamente,

$$\begin{aligned} \mathbf{S}_t &= E[V(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, D_t)|D_T] + V[E(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, D_t)|D_T] \\ &= E[\mathbf{C}_t - \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} \mathbf{G}_{t+1} \mathbf{C}'_t | D_T] + V[\mathbf{m}_t + \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} (\boldsymbol{\theta}_{t+1} - \mathbf{G}_{t+1} \mathbf{m}_t) | D_T] \\ &= \mathbf{C}_t - \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} \mathbf{G}_{t+1} \mathbf{C}'_t + \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} V(\boldsymbol{\theta}_{t+1}|D_T) \mathbf{R}_{t+1}^{-1} \mathbf{G}_{t+1} \mathbf{C}'_t \\ &= \mathbf{C}_t - \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} \mathbf{R}_{t+1} \mathbf{R}_{t+1}^{-1} \mathbf{G}_{t+1} \mathbf{C}'_t + \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} \mathbf{S}_{t+1} \mathbf{R}_{t+1}^{-1} \mathbf{G}_{t+1} \mathbf{C}'_t \\ &= \mathbf{C}_t - \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} (\mathbf{R}_{t+1} - \mathbf{S}_{t+1}) \mathbf{R}_{t+1}^{-1} \mathbf{G}_{t+1} \mathbf{C}'_t, \end{aligned}$$

em que  $V(\boldsymbol{\theta}_{t+1}|D_T) = \mathbf{S}_{t+1}$ .

Na Figura 2.7 apresentamos o processo de suavização para séries temporal *Air-Passengers*.

### 2.4.3 Predição

Suponha que o departamento de recursos humanos de uma empresa deseja fazer o planejamento dos gastos mensais com seus funcionários para o próximo ano, baseado nas informações disponíveis até o momento atual  $t$ .

Considere que esta empresa formulou um MDL normal para ajustar a série temporal custo mensal com funcionários. A fim de fazer inferências sobre os estados dos gastos (como tendência dos gastos, periodicidade de comportamentos e etc.) foi realizado o procedimento de filtragem, ou seja, à medida que um gasto mensal tornava-se conhecido, o sistema era atualizado, incorporando toda a informação dessa nova observação, de modo a resumir características importantes de todo o processo. Assim, foram obtidas a distribuição filtrada  $p(\boldsymbol{\theta}_t|D_t)$  e a preditiva a um passo à frente  $p(y_{t+1}|D_t)$ .

Porém, como o foco principal da empresa reside em observar o desenvolvimento futuro da série para horizontes mais distantes, o pesquisador responsável pelo estudo decidiu fazer previsões sobre os valores futuros sequencialmente, isto é, à medida que novas observações tornam-se disponíveis. Tal decisão se apresenta bastante coerente, uma vez que, devida a natureza *Markoviana* do modelo de espaço de estados, a

distribuição filtrada atua como uma primeira distribuição, ou informação, para a evolução do modelo. Tal informação pode ser suficientemente resumida pela sua média  $\mathbf{m}_t$  que, por sua vez, é um estimador ótimo para  $(\boldsymbol{\theta}_t|D_t)$ .

Podemos pensar nesse procedimento como o fluxo de estimação sequencial dos estados, considerando, que agora, deseja-se estimar o estado  $\boldsymbol{\theta}_{t+k}$  e prever a observação  $Y_{t+k}$ , para  $k \geq 1$ , com base no conjunto de informações  $D_t$ . A figura (2.6) mostra o processo de predição dos estados e observações.

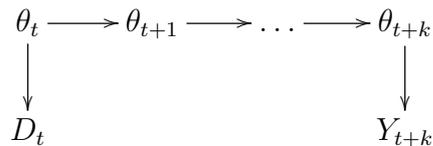


Figura 2.6: Fluxo de informação para predição de  $Y_{t+k}$ .

Observe que essas predições, além de apresentarem os possíveis cenários sobre o comportamento futuro da série, também podem ser um indicativo da adequabilidade do modelo utilizado para ajustar os dados. Ou seja, tais predições podem ser consideradas boas representações do desenvolvimento futuro, se as previsões de observações passadas forem bem próximas dos valores reais da série temporal, de modo que seu desvio,  $e_t = Y_t - f_t$ , seja pequeno.

Nesse contexto, é pertinente notar que a senescência da informação é muito relevante, no sentido de que, à medida que se faz predições com horizontes mais e mais distantes das informações contidas em  $D_t$ , mais incerteza é adicionada ao sistema, provocando previsões menos precisas.

### Predições $k$ passos a frente

No MDL, as distribuições preditivas dos estados e das observações a  $k$  passos à frente também são Gaussianas, de maneira que estas podem ser sumarizadas pelas suas médias e variâncias.

**Proposição 2.4.** *Para um MDL definido por (2.8), seja  $\mathbf{a}_t(0) = \mathbf{m}_t$  e  $\mathbf{R}_t(0) = \mathbf{C}_t$ . Então, para  $k \geq 1$ , seguem as seguintes afirmações:*

(i). A distribuição de  $(\boldsymbol{\theta}_{t+k}|D_t)$  é Gaussiana, com

$$(\boldsymbol{\theta}_{t+k}|D_t) \sim N_p(\mathbf{a}_t(k), \mathbf{R}_t(k)),$$

onde

$$\mathbf{a}_t(k) = \mathbf{G}_{t+k}\mathbf{a}_t(k-1) \quad \text{e} \quad \mathbf{R}_t(k) = \mathbf{G}_{t+k}\mathbf{R}_t(k-1)\mathbf{G}'_{t+k} + \mathbf{W}_{t+k}.$$

(ii). A distribuição de  $(Y_{t+k}|D_t)$  é Gaussiana, com

$$(Y_{t+k}|D_t) \sim N_n(f_t(k), Q_t(k)),$$

onde

$$f_t(k) = \mathbf{F}'_{t+k}\mathbf{a}_t(k) \quad \text{e} \quad Q_t(k) = \mathbf{F}'_{t+k}\mathbf{R}_t(k)\mathbf{F}_{t+k} + \mathbf{V}_{t+k}.$$

**Prova:** Decompondo o vetor de estados  $\boldsymbol{\theta}_{t+k}$ , encontramos que

$$\begin{aligned} \boldsymbol{\theta}_{t+k} &= \mathbf{G}_{t+k}\boldsymbol{\theta}_{t+k-1} + w_{t+k} \\ &= \mathbf{G}_{t+k}(\mathbf{G}_{t+k-1}\boldsymbol{\theta}_{t+k-2} + w_{t+k-1}) + w_{t+k} \\ &= \mathbf{G}_{t+k}(\mathbf{G}_{t+k-1}(\mathbf{G}_{t+k-2}\boldsymbol{\theta}_{t+k-3} + w_{t+k-2}) + w_{t+k-1}) + w_{t+k} \\ &\quad \vdots \\ &= \mathbf{G}_{t+k}\mathbf{G}_{t+k-1}\mathbf{G}_{t+k-2}\dots\mathbf{G}_{t+1}\boldsymbol{\theta}_t + \mathbf{G}_{t+k}\mathbf{G}_{t+k-1}\mathbf{G}_{t+k-2}\dots\mathbf{G}_{t+2}w_{t+1} \\ &\quad + \mathbf{G}_{t+k}\mathbf{G}_{t+k-1}\mathbf{G}_{t+k-2}\dots\mathbf{G}_{t+3}w_{t+2} + \dots + \mathbf{G}_{t+k}w_{t+k-1} + w_{t+k}, \end{aligned}$$

denotando  $\mathbf{H}_{t+k}(r) = \mathbf{G}_{t+k}\mathbf{G}_{t+k-1}\mathbf{G}_{t+k-2}\dots\mathbf{G}_{t+k-(r-1)}$ ,  $\forall t$  e  $r \leq k$  inteiro, com  $\mathbf{H}_{t+k}(0) = \mathbf{I}$ , tem-se que

$$\begin{aligned} \boldsymbol{\theta}_{t+k} &= \mathbf{H}_{t+k}(k)\boldsymbol{\theta}_t + \mathbf{H}_{t+k}(k-1)w_{t+1} + \mathbf{H}_{t+k}(k-2)w_{t+2} + \dots + \mathbf{H}_{t+k}(1)w_{t+k-1} \\ &\quad + \mathbf{H}_{t+k}(0)w_{t+k} \\ &= \mathbf{H}_{t+k}(k)\boldsymbol{\theta}_t + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)w_{t+r}. \end{aligned}$$

Então,

$$(\boldsymbol{\theta}_{t+k}|D_t) = (\mathbf{H}_{t+k}(k)\boldsymbol{\theta}_t + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)w_{t+r}|D_t).$$

Nota-se que  $(\boldsymbol{\theta}_{t+k}|D_t)$  é uma combinação linear de variáveis aleatórias normais, de maneira que pelas propriedades de linearidade da distribuição normal, tem-se que  $\boldsymbol{\theta}_{t+k}$  dado  $D_t$  é também Gaussiana, com parâmetros

$$\begin{aligned}
\mathbf{a}_t(k) &= E(\boldsymbol{\theta}_{t+k}|D_t) \\
&= E(\mathbf{H}_{t+k}(k)\boldsymbol{\theta}_t + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)w_{t+r}|D_t) \\
&= \mathbf{H}_{t+k}(k)E(\boldsymbol{\theta}_t|D_t) + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)E(w_{t+r}|D_t) \\
&= \mathbf{H}_{t+k}(k)\mathbf{m}_t.
\end{aligned}$$

Definindo,

$$\begin{aligned}
\mathbf{a}_t(1) &= \mathbf{G}_{t+1}\mathbf{m}_t \\
\mathbf{a}_t(2) &= \mathbf{G}_{t+2}\mathbf{G}_{t+1}\mathbf{m}_t = \mathbf{G}_{t+2}\mathbf{a}_t(1) \\
&\vdots \\
\mathbf{a}_t(k-1) &= \mathbf{G}_{t+k-1}\dots\mathbf{G}_{t+1}\mathbf{m}_t = \mathbf{G}_{t+k-1}\mathbf{a}_t(k-2)
\end{aligned}$$

obtemos que média da distribuição preditiva dos estados é

$$\mathbf{a}_t(k) = \mathbf{G}_{t+k}\mathbf{a}_t(k-1).$$

Analogamente,

$$\begin{aligned}
\mathbf{R}_t(k) &= V(\boldsymbol{\theta}_t|D_t) \\
&= V(\mathbf{H}_{t+k}(k)\boldsymbol{\theta}_t + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)w_{t+r}|D_t) \\
&= \mathbf{H}_{t+k}(k)V(\boldsymbol{\theta}_t|D_t)\mathbf{H}'_{t+k}(k) + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)V(w_{t+r}|D_t)\mathbf{H}'_{t+k}(k-r) \\
&= \mathbf{H}_{t+k}(k)\mathbf{C}_t\mathbf{H}'_{t+k}(k) + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)\mathbf{W}_{t+r}\mathbf{H}'_{t+r}(k-r).
\end{aligned}$$

Fazendo,

$$\begin{aligned}
\mathbf{R}_t(1) &= \mathbf{G}_{t+1}\mathbf{C}_t\mathbf{G}'_{t+1} + \mathbf{W}_{t+1} \\
\mathbf{R}_t(2) &= \mathbf{G}_{t+2}\mathbf{G}_{t+1}\mathbf{C}_t\mathbf{G}'_{t+1}\mathbf{G}'_{t+2} + \mathbf{G}_{t+2}\mathbf{W}_{t+1}\mathbf{G}'_{t+2} + \mathbf{W}_{t+2} \\
&= \mathbf{G}_{t+2}(\mathbf{G}_{t+1}\mathbf{C}_t\mathbf{G}'_{t+1} + \mathbf{W}_{t+1})\mathbf{G}'_{t+2} + \mathbf{W}_{t+2} = \mathbf{G}_{t+2}\mathbf{R}_t(1)\mathbf{G}'_{t+2} + \mathbf{W}_{t+2} \\
&\vdots \\
\mathbf{R}_t(k-1) &= \mathbf{G}_{t+k-1}\mathbf{R}_t(k-2)\mathbf{G}'_{t+k-1} + \mathbf{W}_{t+k-1},
\end{aligned}$$

de modo que  $\mathbf{R}_t(k) = \mathbf{G}_{t+k} \mathbf{R}_t(t+k-1) \mathbf{G}'_{t+k} + \mathbf{W}_{t+k}$ . Logo, fica provado o item (i).

A demonstração do item (ii) segue de forma análoga a do item (i). Isto é, levando em consideração a propriedade da distribuição Gaussiana, que afirma que combinação linear de variáveis aleatórias normais tem distribuição normal, provamos que

$$\begin{aligned} f_t(k) &= E(Y_{t+k}|D_t) \\ &= E(\mathbf{F}'_{t+k} \boldsymbol{\theta}_{t+k} + v_{t+k}|D_t) \\ &= \mathbf{F}'_{t+k} E(\boldsymbol{\theta}_{t+k}|D_t) + E(v_{t+k}|D_t) \\ &= \mathbf{F}'_{t+k} \mathbf{a}_t(k) \end{aligned}$$

e

$$\begin{aligned} Q_t(k) &= V(Y_{t+k}|D_t) \\ &= E(\mathbf{F}'_{t+k} \boldsymbol{\theta}_{t+k} + v_{t+k}|D_t) \\ &= \mathbf{F}'_{t+k} V(\boldsymbol{\theta}_{t+k}|D_t) \mathbf{F}_{t+k} + V(v_{t+k}|D_t) \\ &= \mathbf{F}'_{t+k} \mathbf{R}_t(k) \mathbf{F}_{t+k} + \mathbf{V}_{t+k}. \end{aligned}$$

## Principais medidas de qualidade de ajuste

Existem algumas medidas que nos permitem quantificar a qualidade do ajuste do modelo dinâmico aos dados observados. Listamos abaixo as principais e mais usadas medidas de ajuste na literatura:

- Densidade Preditiva Observada (DPO)

$$\text{DPO} = \prod_{t=1}^N P(Y_t|D_{t-1}); \quad (2.9)$$

- Erro Absoluto Médio (MAD - *Mean Absolute Deviation*)

$$\text{MAD} = \frac{1}{h} \sum_{t=1}^h |y_t - f_t|; \quad (2.10)$$

- Erro Quadrático Médio (MSE - *Mean Square Error*)

$$\text{MSE} = \frac{1}{h} \sum_{t=1}^h (y_t - f_t)^2; \quad (2.11)$$

- Erro Percentual Médio (MAPE - *Mean Average Percentage Error*)

$$\text{MAPE} = \frac{1}{h} \sum_{t=1}^h 100 \frac{|y_t - f_t|}{y_t}; \quad (2.12)$$

- Erro Simétrico Percentual Médio (SMAPE - *Symmetric Mean Absolute Percentage Error*)

$$\text{SMAPE} = \frac{1}{h} \sum_{t=1}^h 100 \frac{|y_t - f_t|}{(y_t + f_t)/2}; \quad (2.13)$$

- Erro Percentual Mediano (MedAPE - *Median Average Percentage Error*)

$$\text{MedAPE} = \text{mediana} \left( \frac{|y_t - f_t|}{y_t} \right) \quad 1 \leq t \leq h; \quad (2.14)$$

- Erro Escalonado Médio (MASE - *Mean Average Scaled Error*).

$$\text{MASE} = \frac{1}{h} \sum_{t=1}^h 100 \frac{|y_t - f_t|}{\sum_{j=2}^h |y_j - y_{j-1}|}. \quad (2.15)$$

Em que  $h$  denota o horizonte da previsão,  $y_t$  a observação no tempo  $t$  e  $f_t$  o valor da previsão no instante  $t$ , com  $t \geq 1$ .

Essas medidas de ajuste apresentam algumas limitações, então a seleção de uma ou mais dessas métricas depende do interesse que o pesquisador quer atingir ao adotar um determinado modelo. Para mais detalhes sobre essas métricas de desempenho veja as referências Flores (1986), Armstrong e Collopy (1992), Makridakis e Hibon (2000), Hyndman e Koehler (2006) e Silva (2014).

Quando não é possível calcular a densidade preditiva um passo a frente ou a  $k$  passos a frente torna-se inviável obter as medidas descritas acima. Desse modo é pertinente utilizar algum critério de seleção de modelos que baseia-se em outras quantidades. Abaixo descrevemos a medida *Deviance Information Criterion* (DIC), uma generalização do critério de informação de Akaike proposta por Spiegelhalter et al. (2002), que utiliza no seu cálculo apenas a amostra e as estimativas dos parâmetros.

### **Deviance Information Criterion - DIC**

Considere um modelo com um vetor de observações  $\mathbf{y} = (y_1, \dots, y_n)'$  e um vetor de parâmetros  $\boldsymbol{\theta}$  com função de verossimilhança dada por  $p(\mathbf{y}|\boldsymbol{\theta})$ . A deviance do modelo é definida por  $D(\boldsymbol{\theta}) = -2 \log[p(\mathbf{y}|\boldsymbol{\theta})]$ . Desse modo, o DIC é descrito por

$$\text{DIC} = 2E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})] - D[E_{\boldsymbol{\theta}|\mathbf{y}}(\boldsymbol{\theta})],$$

em que  $E_{\boldsymbol{\theta}|y}[D(\boldsymbol{\theta})]$  e  $E_{\boldsymbol{\theta}|y}(\boldsymbol{\theta})$  são a média a *posteriori* da *deviance*  $D(\boldsymbol{\theta})$  e  $\boldsymbol{\theta}$ , respectivamente.

Nos casos de ajuste de modelos via MCMC, o DIC pode ser implementado considerando uma amostra  $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(M)}$  da distribuição a *posteriori*  $p(\boldsymbol{\theta}|y)$ , de modo que

$$DIC = 2\bar{D} - D(\bar{\boldsymbol{\theta}}),$$

com  $\bar{D} = \frac{1}{M} \sum_{j=1}^M D(\boldsymbol{\theta}^{(j)})$  e  $D(\bar{\boldsymbol{\theta}}) = D(\frac{1}{M} \sum_{j=1}^M \boldsymbol{\theta}^{(j)})$ . Para mais detalhes sobre esse critério veja Reis (2008).

### Exemplo ilustrativo

A média e variância da distribuição preditiva para horizontes mais distantes podem ser obtidas pelo comando *dlmForecast*, cujo principal argumento é o *dlmFilter*. No caso em que o interesse limita-se à distribuição preditiva um passo à frente, as estimativas de seus momentos são dadas na saída do comando *dlmFilter*.

Para ilustrar o processo de estimação dos estados e previsão das observações, a série *AirPassengers*, do *software* R, foi estudada e modelada segundo um MDL de nível local. A Figura 2.7 mostra os gráficos desta série com as estimativas do nível filtrado, suavizado e o gráfico das observações preditas a um passo à frente. Apresentamos, também, no Código 5.3 no Apêndice D.2, a programação para a realização do estudo da série *AirPassengers*.

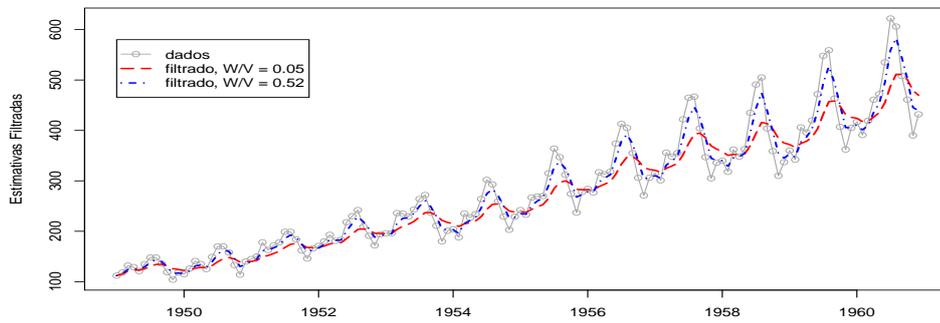
Assim como em West e Harrison (1997) e Petris et al. (2009), foram ajustados dois modelos de nível local com razões sinal-ruído distintas,  $r = W/V = 0.05$  e  $r = W/V = 0,52$ , a fim de estimar o nível real do número total de passageiros em linhas aéreas internacionais. No processo de filtragem, observa-se que para o modelo com  $r = 0,52$ , os níveis filtrados parecem acompanhar com mais precisão a série observada. Dessa forma, com base nesse modelo, foram realizados a suavização e a predição a um passo à frente.

Na estimação dos estados através da suavização, foi possível acompanhar comportamentos padrão da série, como tendência. A vantagem deste processo de estimação em relação à filtragem é que aquela considera toda as observações da série temporal em estudo, enquanto esta, considera apenas as observações até o instante que se

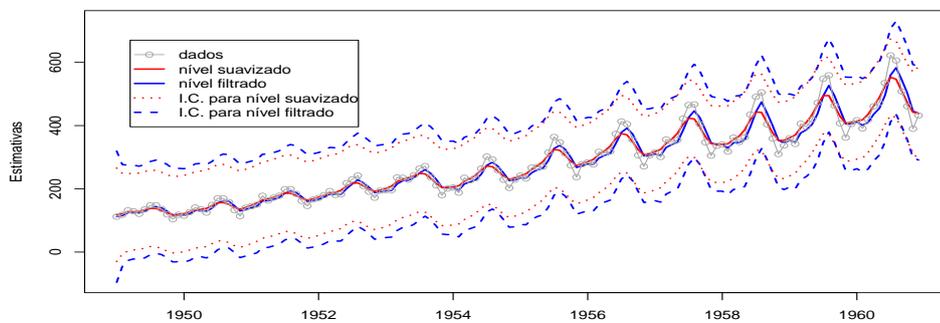
deseja estimar. Isso reflete na variabilidade das estimativas, ou seja, as estimativas suavizadas apresentam menor variância que as estimativas filtradas. Quando toda a amostra é observada, a variância das estimativas por filtragem iguala-se a das estimativas suavizadas. Tais variâncias, em termos de decomposições em valores singulares, são dadas a partir da linha 31 do Código 5.3.

Podemos visualizar melhor as diferenças entre os dois processos de estimação através da Figura 2.7b, que compara os níveis filtrados e suavizados. A figura mostra também os intervalos de credibilidade de 95% para ambas as estimativas.

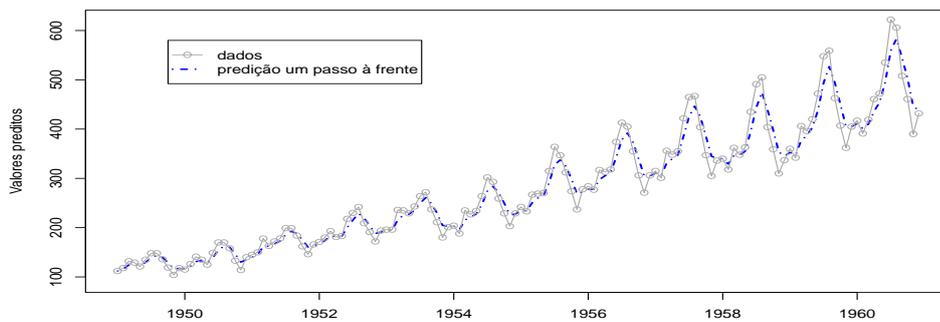
Assim como em Petris et al. (2009), no estudo da série Rio Nilo do *software* R, não foi adicionado um intervalo de credibilidade para os valores preditos  $f_t$ . Em nosso estudo, a razão desta ausência decorre da grande amplitude do intervalo de credibilidade.



(a) *Nível filtrado com razão sinal-ruído  $r = 0,5$  e  $r = 0,52$ .*



(b) *Valores suavizados considerando filtro com  $r = 0,52$ .*



(c) *Predição um passo à frente considerando filtro com  $r = 0,52$ .*

Figura 2.7: (a) Estimativas filtradas com razões sinal-ruído diferentes; (b) Estimativas Suavizadas e Filtradas com intervalo de credibilidade de 95%; (c) Predição das observações um passo à frente.

# Capítulo 3

## Especificação do Modelo

O processo inferencial em modelos dinâmicos lineares baseia-se no filtro de Kalman que, por sua vez, viabiliza uma estrutura capaz de estimar parâmetros desconhecidos por meio da função de máxima verossimilhança, além de colaborar com a previsão de valores futuros da série. No entanto, esse procedimento só é adequado quando o MDL é completamente especificado, ou seja, quando as matrizes  $\mathbf{F}_t$ ,  $\mathbf{G}_t$  e as matrizes de variância  $\mathbf{V}_t$  e  $\mathbf{W}_t$  são conhecidas. Uma forma de tornar a especificação de um modelo mais simples, é considerar que o MDL é formado por uma combinação de componentes independentes que podem ser modelados individualmente.

### 3.1 Análise da série temporal por componentes

A especificação de um modelo dinâmico para séries temporais não consiste uma tarefa simples, principalmente, quando as matrizes  $\mathbf{F}_t$ ,  $\mathbf{G}_t$  e as matrizes de variância  $\mathbf{V}_t$  e  $\mathbf{W}_t$  são desconhecidas.

Uma importante suposição em modelos clássicos para séries temporais, que também pode ser assumida na estrutura dos modelos dinâmicos, é a ideia de que a série pode ser descrita por uma decomposição (aditiva ou multiplicativa) de dois ou mais componentes, como: tendência, sazonalidade e etc.

Na perspectiva *Bayesiana*, essa estrutura pode ser visualizada como uma combinação da série temporal em componentes mais simples, onde cada um reflete uma característica distinta e é representado individualmente por um modelo dinâmico linear

que será construído através de funções (polinomiais, trigonométricas) apropriadas.

A coleção dos diferentes componentes forma um MDL único que representará o modelo geral a ser ajustado à série temporal, tal construção é denominada de *superposição* (do inglês *superposition*) por West e Harrison (1997).

Dessa forma, considere a decomposição aditiva para uma série temporal univariada  $Y_t$ , que pode ser escrita como uma soma de componentes independentes

$$Y_t = Y_{1,t} + Y_{2,t} + \dots + Y_{h,t}, \quad (3.1)$$

onde  $Y_{i,t}$  pode representar um componente de tendência, regressão e assim por diante com  $i = \{1, \dots, h\}$ . Desse modo, o  $i$ -ésimo componente é especificado por um MDL tal como segue:

$$\begin{aligned} Y_{i,t} &= \mathbf{F}'_{i,t} \boldsymbol{\theta}_{i,t} + v_{i,t}, & v_{i,t} &\sim N(0, \mathbf{V}_{i,t}), \\ \boldsymbol{\theta}_{i,t} &= \mathbf{G}_{i,t} \boldsymbol{\theta}_{i,t-1} + w_{i,t}, & w_{i,t} &\sim N(0, \mathbf{W}_{i,t}), \end{aligned}$$

onde o vetor de estados  $\boldsymbol{\theta}_{i,t}$  tem dimensão  $p_i$  e são distintos. As séries  $(Y_{i,t}, \boldsymbol{\theta}_{i,t})$  e  $(Y_{j,t}, \boldsymbol{\theta}_{j,t})$  são mutuamente independentes para todo  $i \neq j$ .

Assim, a série

$$Y_t = \sum_{i=1}^h Y_{i,t}$$

é descrita pelo DLM,

$$\begin{aligned} Y_t &= \mathbf{F}'_t \boldsymbol{\theta}_t + v_t, & v_t &\sim N(0, \mathbf{V}_t), \\ \boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + w_t, & w_t &\sim N(0, \mathbf{W}_t), \end{aligned}$$

em que

$$\boldsymbol{\theta}_t = \begin{bmatrix} \boldsymbol{\theta}_{1,t} \\ \boldsymbol{\theta}_{2,t} \\ \vdots \\ \boldsymbol{\theta}_{h,t} \end{bmatrix}, \quad \mathbf{F}_t = \begin{bmatrix} \mathbf{F}_{1,t} \\ \mathbf{F}_{2,t} \\ \vdots \\ \mathbf{F}_{h,t} \end{bmatrix}$$

as matrizes  $\mathbf{G}_t$  e  $\mathbf{W}_t$  são bloco diagonais

$$\mathbf{G}_t = \text{bloco-diag}[\mathbf{G}_{1,t}, \dots, \mathbf{G}_{h,t}], \quad \mathbf{W}_t = \text{bloco-diag}[\mathbf{W}_{1,t}, \dots, \mathbf{W}_{h,t}],$$

e  $V_t = \sum_{i=1}^h V_{i,t}$ . Todas as matrizes do MDL são definidas conhecidas.

Nesse cenário, a função de predição para a série  $Y_t$  gerada pela *superposição* de  $h$  modelos componentes é dada por

$$f_t(k) = \sum_{i=1}^h f_{it}(k).$$

Portanto, temos que a combinação linear das séries geradas pelos MDLs dos componentes independentes segue um MDL. Esse resultado é denominado de Princípio da *Superposição* (West e Harrison, 1997) e considera ainda as propriedades de aditividade associados com os modelos normais lineares. Desse modo é necessário a suposição de que os vetores  $v_{1,t}, \dots, v_{h,t}$  e  $w_{1,t}, \dots, w_{h,t}$  têm distribuição conjunta normal e que são mutuamente e internamente independentes.

No R, um comando muito útil para visualizar os elementos que compõem a série temporal é o *stl* do pacote *stats*. Esta função decompõe a série em componentes sazonais, tendência e erros através de outra função, denominada *loess*.

Para criar um MDI por *superposição* no R, o pacote *dln* disponibiliza a família de funções *dlnMod\**. Através desta classe de objetos é possível criar um modelo que incorpore, por exemplo, tendência e efeitos sazonais. Nas próximas sessões, falaremos com mais detalhes sobre essa família de funções.

### 3.1.1 Modelos Polinomiais

Uma componente muito presente em séries temporais é a tendência que comumente é definida como o desenvolvimento suave do nível médio da série ao longo do tempo.

Harvey (1989) discute essa ideia de tendência e afirma que não há uma razão fundamental para considerar essa definição, uma vez que uma descrição relacionada à previsão apresenta-se mais consistente.

Ainda segundo Harvey (1989), em termos de previsão, a tendência estimada é parte de uma série temporal que torna mais evidente os movimentos futuros desta. De modo que, é conveniente ajustar um modelo estatístico que forneça boas previsões.

Nesse sentido, podemos pensar na tendência estimada como o comportamento de  $Y_{t+k}$ , dado a informação disponível até o tempo  $t$ . Assim, uma boa alternativa para a modelagem da componente tendência são os MDLs polinomiais que são considerados

uma sub-classe dos modelos dinâmicos lineares.

Desse modo, o MDL polinomial de ordem  $p$  é definido como um MDL com matrizes  $\mathbf{F}_t = \mathbf{F}$  e  $\mathbf{G}_t = \mathbf{G}$  constantes, e a função de predição ou tendência esperada fica na forma

$$f_t(k) = E(Y_{t+k}|D_t) = a_{t,0} + a_{t,1}k + \dots + a_{t,p-1}k^{p-1}, \quad k \geq 0, \quad (3.4)$$

onde os coeficientes  $a_{t,0}, \dots, a_{t,n-1}$  são funções lineares da média a *posteriori*  $m_t = E(\boldsymbol{\theta}_t|D_t)$  e são independentes de  $k$ .

A tendência não necessariamente apresenta um desenvolvimento ascendente ou descendente, de maneira que uma determinada série sempre conterá uma tendência a não ser que as previsões a longo prazo sejam iguais a zero.

A função de predição dada em (3.4) dá indícios dos movimentos a longo prazo da série. No entanto, quando o interesse reside em previsões de curto e médio prazo, polinômios de ordem baixa apresentam boas aproximações. Portanto, neste trabalho, discutiremos sobre MDLs polinomiais de primeira e segunda ordem que, por sua vez, também são referidos como modelo de nível local e modelo de crescimento linear, respectivamente. MDLs polinomiais de ordem maior que 3 (corresponde às funções de crescimento quadrático) são raramente encontrados em aplicações estatísticas.

### Modelo de Nível Local

O modelo polinomial de primeira ordem é definido pelas seguintes equações:

**Equação das observações:**

$$Y_t = \mu_t + v_t, \quad v_t \sim N(0, V_t) \quad (3.5a)$$

**Equação do sistema:**

$$\mu_t = \mu_{t-1} + \omega_t, \quad \omega_t \sim N(0, W_t) \quad (3.5b)$$

**Informação Inicial:**

$$(\mu_0|D_0) \sim N(m_0, C_0).$$

As observações são representadas pela equação (3.5a), onde  $\mu_t$  é o nível da série no tempo  $t$  e  $v_t$  é o erro observacional. O tempo de evolução do nível é modelado segundo o passeio aleatório em (3.5b), com  $\omega_t$  sendo os erros de evolução.

As seqüências de erros de observação e evolução são mutuamente e serialmente não correlacionados. É assumido que tais erros são normalmente distribuídos com média zero e variância  $V_t$  e  $W_t$  que, admitimos conhecidas.

A informação inicial corresponde a representação probabilística da crença acerca do nível  $\mu_0$  no tempo  $t = 0$ . A média  $m_0$  é a estimativa pontual deste nível, e a variância  $C_0$  é uma medida de incerteza associada à esta informação.

O comportamento do processo  $Y_t$  é altamente influenciado pela razão sinal-ruído  $r = W_t/V_t$  que, desempenha um papel relevante na determinação de quanto o movimento da série é influenciado pelas observações e pelas variações de nível.

O nível  $\mu_t$  é usualmente considerado localmente constante, isto é, em alguns períodos para trás ou para frente de um determinado tempo  $t$ , temos que a média da série temporal, nesse intervalo, é estável. Nessas condições, temos um modelo que não prevê mudanças significativas com muita precisão, sendo mais útil para previsões a curto prazo. Quando  $W_t = 0$ , o nível é constante para todo tempo  $t$ .

A Figura 3.1 mostra duas trajetórias simuladas de  $Y_t$  e  $\mu_t$  para valores distintos da razão sinal-ruído.

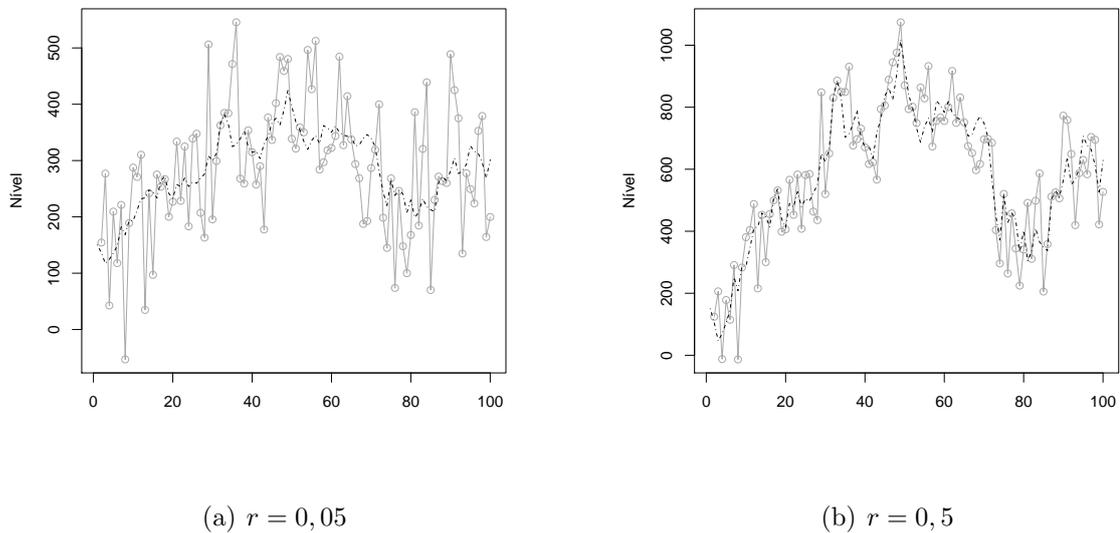


Figura 3.1: Trajetórias simuladas de acordo com um passeio aleatório para dois diferentes valores da razão sinal-ruído.

Deste modo, pode-se pensar em  $\mu_t$  como uma aproximação em séries de Taylor de

1ª ordem para uma função suave no tempo  $\mu(t)$ , de maneira que

$$\mu(t + \delta t) = \mu_t + \text{termos de ordem maior},$$

com os termos de ordem maior sendo descrito como um ruído de média zero.

Em consequência a esta estrutura, a função preditiva é contante e igual a média *a posteriori*, isto é

$$f_t(k) = E(Y_{t+k}|D_t) = E(\mu_t|D_t) = m_t, \quad (3.6)$$

para todo  $k > 0$ .

O modelo de nível local é geralmente apropriado para séries temporais que não apresentam com nitidez o elemento tendência ou variações sazonais. De maneira que, esse modelo é conveniente, somente, para aplicações a curto prazo e, sobretudo, em casos em que a variância observacional, dada por  $V_t$ , é maior que a variação do nível, medida por  $W_t$ .

O processo inferencial do modelo polinomial de primeira ordem segue similarmente ao do Filtro de Kalman apresentado no Capítulo 2. As distribuições de probabilidade necessárias para estimação, previsão e aprendizagem do sistema é dada na proposição a seguir.

**Proposição 3.1.** *Considere um MDL especificado por (3.5), segue que*

(i). *A distribuição a posteriori de  $\mu_{t-1}$  é Gaussiana com parâmetros:*

$$(\mu_{t-1}|D_{t-1}) \sim N(m_{t-1}, C_{t-1}),$$

*para alguma média  $m_{t-1}$  e variância  $C_{t-1}$ .*

(ii). *A distribuição a priori de  $\mu_t$  dado  $D_{t-1}$  é Gaussiana com parâmetros:*

$$(\mu_t|D_{t-1}) \sim N(m_{t-1}, R_t),$$

*onde  $R_t = C_{t-1} + W_t$ .*

(iii). *Distribuição preditiva de  $Y_t$  dado  $D_{t-1}$  é Gaussiana com parâmetros:*

$$(Y_t|D_{t-1}) \sim N(f_t, Q_t),$$

*onde  $f_t = m_{t-1}$  e  $Q_t = R_t + V_t$ .*

(iv). Distribuição filtrada, ou a posteriori, de  $\mu_t$  dado  $D_t$  é Gaussiana com parâmetros:

$$(\mu_t|D_t) \sim N(m_t, C_t),$$

com

$$m_t = m_{t-1} + A_t e_t \quad \text{e} \quad C_t = A_t V_t,$$

em que

$$A_t = R_t/Q_t \quad \text{e} \quad e_t = Y_t - f_t.$$

Observe que a média a posteriori é uma soma da média a priori e um erro de previsão que é ponderado pela proporção  $A_t$ . Como citado no Capítulo 2, uma forma alternativa de expressar a média  $m_t$  é

$$m_t = A_t Y_t + (1 - A_t) m_{t-1},$$

que apresenta  $m_t$  como uma média ponderada do nível  $m_{t-1}$  e a da observação  $Y_t$ . O peso  $A_t$ , denominado coeficiente adaptativo, é uma proporção que mede a variabilidade relativa contida na priori com respeito a variabilidade preditiva  $Q_t$ . Assim, quando  $A_t$  é aproximadamente zero, temos que  $R_t < V_t$ , indicando que a distribuição a priori é mais concentrada que a verossimilhança, de modo que a média a posteriori é completamente influenciada pela média a priori. Quando  $A_t$  está próximo de um, a priori é considerada pouco informativa em relação a verossimilhança.

Para um estudo mais detalhado sobre este modelo veja Capítulo 2 de West e Harrison (1997).

## Modelo de Crescimento Linear

Um modelo mais geral que o de nível local, amplamente utilizado para modelar tendências lineares é o modelo de crescimento linear ou de tendência linear local. Esse modelo inclui na equação do sistema, dada em (3.5b), um termo de inclinação que varia com o tempo e é modelado por uma passeio aleatório. Assim, temos que o modelo assume a seguinte forma:

**Equação das observações:**

$$Y_t = \mu_t + v_t, \quad v_t \sim N(0, V_t) \quad (3.7a)$$

**Equação do sistema:**

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{t,1}, & \omega_{t,1} &\sim N(0, \sigma_\mu^2) \\ \beta_t &= \beta_{t-1} + \omega_{t,2}, & \omega_{t,2} &\sim N(0, \sigma_\beta^2) \end{aligned} \quad (3.7b)$$

onde os erros  $v_t$ ,  $\omega_{t,1}$  e  $\omega_{t,2}$  são mutuamente não correlacionados. Esse é um MDL polinomial que equivale a um MDL padrão definido pelas seguintes matrizes:

$$\boldsymbol{\theta}_t = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix}, \quad \mathbf{F}_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{G}_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{W}_t = \begin{bmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix}.$$

Nestas condições, temos que  $\mu_t$  ainda é compreendido como o nível local, enquanto  $\beta_t$  é interpretado como um crescimento local, ou incremento, no nível da série.

Se  $\sigma_\mu^2$  e  $\sigma_\beta^2$  são iguais a zero, a tendência é determinística. Quando somente  $\sigma_\beta^2$  é zero, a inclinação  $\beta_t$  é fixa, de modo que o modelo em (3.7) reduz-se a um passeio aleatório. Se  $\sigma_\mu^2 = 0$ , a tendência é uma linha reta com inclinação  $\beta_0$  e intercepto  $\mu_0$  e se em nessas condições  $\sigma_\beta^2 > 0$ , a tendência é uma curva suave.

Note que a matriz de delineamento  $\mathbf{F}_t$  e a matriz de transição dos estados  $\mathbf{G}_t$  são constantes para todo  $t \in \{0, 1, \dots\}$ . Portanto, denotando  $m_{t-1} = [\hat{\mu}_{t-1}, \hat{\beta}_{t-1}]'$  e dispendo da Proposição 2.4, temos que as estimativas do vetor de estados  $\boldsymbol{\theta}_{t+k}$  e da observação  $Y_{t+k}$  dado o conjunto de informações disponíveis  $D_t$ , são, respectivamente

$$\begin{aligned} \mathbf{a}_t(k) &= \mathbf{G}_{t+k} \mathbf{G}_{t+k-1} \dots \mathbf{G}_{t+1} m_t \\ &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^k \begin{bmatrix} \hat{\mu}_t \\ \hat{\beta}_t \end{bmatrix} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\mu}_t \\ \hat{\beta}_t \end{bmatrix} \\ &= \begin{bmatrix} \hat{\mu}_t + \hat{\beta}_t k \\ \hat{\beta}_t \end{bmatrix} \end{aligned}$$

e

$$\begin{aligned} f_t(k) &= \mathbf{F}'_{t+k} \mathbf{a}_t(k) \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{\mu}_t + \hat{\beta}_t k \\ \hat{\beta}_t \end{bmatrix} \\ &= \hat{\mu}_t + \hat{\beta}_t k. \end{aligned}$$

Logo, a função predição  $f_t(k)$  é uma função linear de  $k$ , onde  $\hat{\mu}_t$  e  $\hat{\beta}_t$  são as estimativas pontuais de  $\mu_t$  e  $\beta_t$ . Conseqüentemente, tem-se que o modelo de crescimento linear é um MDL polinomial de ordem  $p = 2$ .

A função *dlnModPoly* com o argumento *order = 2* pode ser utilizado para criar um MDL com tendência linear local no R.

### 3.1.2 Modelos Sazonais

Comumente algumas séries temporais apresentam comportamentos cíclicos e periódicos que evidenciam a presença de uma componente sazonal. Esse é o caso, por exemplo, do aumento do número de vendas de passagens aéreas em determinados períodos do ano que, por sua vez, influencia diretamente no aumento do valor dessas passagens. Esse comportamento, frequentemente ocorre em períodos de alta estação, como férias e feriados nacionais, onde o fluxo de pessoas à procura de um destino para viajar é maior do que nas demais épocas do ano. Outro exemplo de série temporal com essas variações, é o número de casos de dengue no Brasil. Esta doença é transmitida por um mosquito, *Aedes aegypti*, e segundo o Ministério da Saúde, esta é uma das epidemias que têm maior impacto na saúde pública do país. A dengue é considerada uma doença com característica sazonal, de modo que, a maior incidência desta ocorre em um mesmo período, todos os anos. No Brasil, este período coincide com o verão (entre janeiro e maio), devido a maior ocorrência de chuva e ao aumento da temperatura.

Neste cenário, para séries temporais com tais características, o interesse consiste em descrever os padrões sazonais da série através de quantidades que podem ser estimadas e são utilizadas para avaliar as mudanças ao longo do tempo e para previsões a curto prazo. Desta maneira, faz-se necessário ajustar modelos adequados para séries temporais com efeitos sazonais.

Na perspectiva dos MDLs, existem duas importantes classes de modelos. A primeira abordagem é denominada de modelos sazonais de forma livre e utiliza fatores sazonais para descrever os padrões cíclicos e periódicos da série. A segunda abordagem para descrever a sazonalidade é especificada pela representação em Séries de Fourier. Ambas as formas serão discutidas nesta dissertação.

## MDLs com Efeitos Sazonais em Forma Livre

Uma forma simples de caracterizar séries temporais com sazonalidade pode ser realizada através de fatores sazonais ou indicadores para cada período do ciclo. A ideia consiste em definir um componente sazonal descrevendo efeitos ou desvios sazonais em torno de um nível *dessazonalizado* ou tendência.

Considere uma série temporal puramente sazonal, com observações trimestrais, isto é, período  $s = 4$ . A série pode ser descrita por desvios sazonais da média representados pelos coeficientes  $\alpha_i$  que, são distintos para cada trimestre, com  $i = 1, \dots, 4$ . Suponha que  $Y_t$  é uma observação referente ao primeiro trimestre,  $Y_{t+1}$  ao segundo trimestre e assim por diante. Desse modo, temos que as observações são expressas da seguinte forma

$$\begin{aligned} Y_t &= \alpha_1 + v_t \\ Y_{t+1} &= \alpha_2 + v_{t+1} \\ Y_{t+2} &= \alpha_3 + v_{t+2} \\ Y_{t+3} &= \alpha_4 + v_{t+3}. \end{aligned}$$

Cada observação carrega a informação do seu respectivo trimestre mais uma parte aleatória devida aos erros  $v'_i$ s. A partir da observação  $Y_{t+4}$  o ciclo começa a repetir o mesmo comportamento, sendo os distúrbios  $v'_i$ s, a parte aleatória que contribui para que a informação contida em  $Y_t$  não seja exatamente igual a contida em  $Y_{t+4}$ . O esquema apresentado na Figura 3.2 explica essa estrutura sazonal.

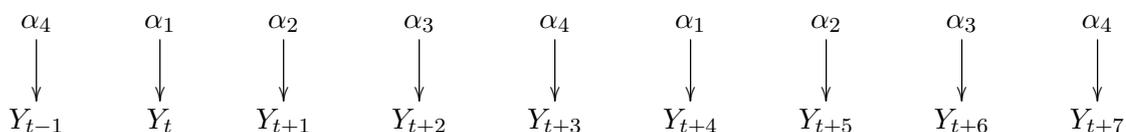


Figura 3.2: Estrutura dos efeitos sazonais e observações.

Nestas condições, o MDL puramente sazonal é definido pelo o conjunto de equações

dado em (2.8) com os seguintes vetores

$$\boldsymbol{\theta}_{t-1} = \begin{bmatrix} \alpha_4 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}, \quad \mathbf{F}_t = \mathbf{F} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Devido a evolução estocástica dos modelos dinâmicos, à medida que o tempo varia ocorre uma rotação entre os elementos do vetor de estados, de forma que no tempo  $t$  temos que

$$\boldsymbol{\theta}_t = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}.$$

A matriz de transição dos estados, agora, denominada de matriz de permutação  $\mathbf{G}_t$  é constante em relação ao tempo e é dada por

$$\mathbf{G}_t = \mathbf{G} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Portanto, temos que a equação das observações e de evolução do MDL no tempo  $t$  são

$$\begin{aligned} Y_t &= \mathbf{F}'\boldsymbol{\theta}_t + v_t = \alpha_1 + v_t \\ \boldsymbol{\theta}_t &= \mathbf{G}\boldsymbol{\theta}_{t-1} + \mathbf{w}_t = [\alpha_1, \alpha_2, \alpha_3, \alpha_4]' + \mathbf{w}_t. \end{aligned}$$

De modo mais geral, séries temporais com comportamento sazonal podem ser modeladas através de MDLs especificados pelos seguintes componentes: o vetor de estados  $\boldsymbol{\theta}_t$   $s$ -dimensional composto pelos desvios sazonais, a matriz de delineamento  $\mathbf{F} = [1, 0, \dots, 0]'$   $s$ -dimensional e uma matriz de permutação  $\mathbf{G}$  com dimensão  $(s \times s)$ .

É importante ressaltar que a matriz de permutação  $\mathbf{G}$  é cíclica, uma vez que, tem como finalidade realizar a rotação entre os efeitos sazonais do vetor de estados. Desse modo, para qualquer  $n \geq 0$ ,  $\mathbf{G}^{ns} = \mathbf{I}_s$  e  $\mathbf{G}^{k+np} = \mathbf{G}^k$  para  $k = 1, \dots, s$ .

Frequentemente, os efeitos  $\alpha'_i s$  variam com o tempo, de modo que a matriz de variância  $\mathbf{W}_t$  é diferente de zero e deve ser cuidadosamente especificada. Quando esta matriz é composta somente por zeros, o erro  $\mathbf{w}_t$  é degenerado em um vetor de zeros e o modelo é denominado de estático.

Tais efeitos sazonais estão condicionados a uma restrição de identicabilidade que possibilita a estimação dos parâmetros ou desvios sazonais e permite visualizar separadamente as componentes tendência e sazonalidade. Nesse caso, é comum que tais efeitos estejam sujeitos a restrição de soma zero, ou seja,  $\sum_{j=1}^s \alpha_j = 0$ .

Essa restrição linear conduz a uma representação mais parcimoniosa, de modo que existirão  $s - 1$  fatores sazonais livres no modelo, assim o vetor de estados terá dimensão  $s - 1$ .

Para esclarecer a relevância dessa condição de identificabilidade, considere ainda o cenário do exemplo anterior, onde agora o vetor de estados no tempo  $t - 1$  e  $t$  é dado por  $\boldsymbol{\theta}_{t-1} = [\alpha_4, \alpha_1, \alpha_2]'$  e  $\boldsymbol{\theta}_t = [\alpha_1, \alpha_2, \alpha_3]'$ , com  $\mathbf{F} = [1, 0, 0]'$ . Assumindo a restrição  $\sum_{j=1}^4 \alpha_j = 0$  e o modelo estático para a evolução de  $\boldsymbol{\theta}_{t-1}$  para  $\boldsymbol{\theta}_t$ , temos que

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix}.$$

com  $\alpha_3 = -(\alpha_1 + \alpha_2 + \alpha_4)$  sendo uma combinação linear dos demais desvios sazonais. Genericamente, para um modelo sazonal de período  $s$ , pode-se considerar um vetor de estados  $(s - 1)$ -dimensional, com  $\mathbf{F} = [1, 0, \dots, 0]'$  também  $(s - 1)$ -dimensional e matriz de permutação  $\mathbf{G}$  com dimensão  $(s - 1) \times (s - 1)$

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -1 & -1 & -1 & \dots & -1 \end{bmatrix}.$$

A função de predição é então dada por

$$f_t(k) = \mathbf{F}'_{t+k} a_t(k) = \mathbf{F}' \mathbf{G}^k \mathbf{m}_t,$$

com  $\mathbf{m}_t = [\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_s]'$ . No R, o MDL com efeitos sazonais pode ser especificado através do comando `dlnModSeas` no pacote `dln`.

## Modelos Sazonais com representação em Séries de *Fourier*

Uma alternativa para captar padrões cíclicos de uma série temporal é através de combinações lineares de harmônicos, ou seja, por meio de funções periódicas. Essa forma de modelar a sazonalidade é denominada de representação em séries de *Fourier* e utiliza funções trigonométricas simples como seno e cosseno.

A modelagem dos efeitos sazonais em forma de componentes harmônicos proporciona uma economia no número de parâmetros e torna a interpretação destes mais simples. Esses benefícios aumentam a capacidade de aprendizagem acerca do sistema e a qualidade das previsões.

Desse modo, considere  $g(t)$  uma função cíclica, tal que, para algum inteiro  $s > 1$  e para todo  $t$  com  $n \geq 0$ ,  $g(t + ns) = g(t)$ . Essa função  $g(t)$  representa uma combinação linear de harmônicos e está associada ao vetor  $s$ -dimensional de fatores sazonais  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s)$ , de modo que, qualquer função periódica de tempo discreto com período  $s$ , assume, para um ciclo completo, que  $g(t) = \alpha_t$ .

Assim, cada componente harmônica de  $g(t)$  pode ser dada por

$$S_j(t) = A_j \cos\left(\frac{2\pi j}{s}t + \gamma_j\right) \quad (3.8)$$

onde  $S_j(t)$  representa uma cossenóide com período  $s$ , amplitude  $A_j$  e fase inicial  $\gamma_j$ . Fazendo  $\omega_j = 2\pi j/s$ , tem-se

$$S_j(t) = A_j \cos(\omega_j t + \gamma_j) \quad (3.9)$$

onde  $\omega_j$  representa a frequência.

Como a função cosseno varia de  $-1$  a  $1$ , o valor de  $S_j(t)$  varia de  $-A_j$  a  $A_j$ . A amplitude  $A_j$  especifica o valor máximo que  $S_j(t)$  assume e a fase determina a posição do máximo.

Desenvolvendo (3.9), segue que

$$S_j(t) = A_j \cos \gamma_j \cos(\omega_j t) - A_j \sen \gamma_j \sen(\omega_j t). \quad (3.10)$$

Definindo

$$a_j = A_j \cos \gamma_j \quad (3.11)$$

e

$$b_j = -A_j \operatorname{sen} \gamma_j, \quad (3.12)$$

obtém-se a seguinte representação do  $j$ -ésimo componente harmônico

$$S_j(t) = a_j \cos(\omega_j t) + b_j \operatorname{sen}(\omega_j t). \quad (3.13)$$

Quando o período  $s$  é conhecido, a frequência  $\omega_j$  é conhecida, de modo que na expressão (3.9) os parâmetros desconhecidos são  $A_j$  e  $\gamma_j$ . Enquanto que na equação (3.13), os parâmetros a serem estimados são  $a_j$  e  $b_j$ .

Note que, para um  $j$  fixo e  $t$  variando de  $\{1, 2, \dots, s\}$ , a soma  $\sum_{t=1}^s S_j(t) = 0$ . Essa condição é similar a restrição de identificabilidade imposta ao modelo sazonal de forma livre. A diferença é que esta é uma consequência proveniente de identidades trigonométricas básicas, isto é, em razão da ortogonalidade das funções seno e cosseno.

Os parâmetros  $A_j$  e  $\gamma_j$  podem ser obtidos a partir de  $a_j$  e  $b_j$ , considerando as seguintes relações:

- Elevando ao quadrado e somando (3.11) e (3.12), obtém-se

$$A_j^2 = a_j^2 + b_j^2 \quad \longrightarrow \quad A = \sqrt{a_j^2 + b_j^2}; \quad (3.14)$$

- Dividindo membro a membro de (3.12) por (3.11), tem-se

$$\operatorname{tg} \gamma_j = -\frac{b_j}{a_j} \quad \longrightarrow \quad \gamma_j = \operatorname{arctg} \left( -\frac{b_j}{a_j} \right). \quad (3.15)$$

Portanto, com o uso de algumas identidades trigonométricas padrões, o modelo geral para os efeitos sazonais com representação em séries de *Fourier* é dado por

$$g(t) = \alpha_t = a_0 + \sum_{j=1}^h S_j(t) \quad (3.16)$$

$$= a_0 + \sum_{j=1}^h [a_j \cos(\omega_j t) + b_j \operatorname{sen}(\omega_j t)], \quad (3.17)$$

onde  $h$  é o maior inteiro não superior a  $s/2$ . Se o período  $s$  de variações cíclicas corresponde a número par,  $h = s/2$ . Se  $s$  é ímpar,  $h = (s - 1)/2$ . As quantidades  $a_j$  e  $b_j$  são denominadas de coeficientes de *Fourier*.

Comumente, a média da série temporal é modelada separadamente da componente sazonal, de modo que é conveniente assumir  $a_0 = 0$ . Consequentemente,  $\alpha_t = \sum_{j=1}^h S_j(t)$ .

No contexto dos MDLs, a evolução da componente harmônica em um tempo  $t$  para  $t + 1$  depende do conhecimento de  $S_j(t)$  e das quantidades  $a_j$  e  $b_j$ . No entanto, geralmente esses coeficientes não são conhecidos, de modo que apenas a informação de  $S_j(t)$  não é suficiente para determinar o valor de  $S_j(t + 1)$ . Todavia, se além de  $S_j(t)$  conhecermos o seu harmônico conjugado

$$S_j^*(t) = -a_j \text{sen}(\omega_j t) + b_j \text{cos}(\omega_j t),$$

é possível obter o  $S_j(t + 1)$  e também  $S_j^*(t + 1)$ . Dessa forma, esses harmônicos conjugados podem ser visualizados como um parâmetro auxiliar que facilita o processo de evolução dos estados. Como resultado, temos que a componente harmônica e seu conjugado no tempo  $t + 1$  é dada por

$$\begin{aligned} S_j(t + 1) &= a_j \text{cos}((t + 1)\omega_j) + b_j \text{sen}((t + 1)\omega_j) \\ &= S_j(t) \text{cos} \omega_j + S_j^*(t) \text{sen} \omega_j \end{aligned} \quad (3.18)$$

e

$$\begin{aligned} S_j^*(t + 1) &= -a_j \text{sen}((t + 1)\omega_j) + b_j \text{cos}((t + 1)\omega_j) \\ &= -S_j(t) \text{sen} \omega_j + S_j^*(t) \text{cos} \omega_j. \end{aligned} \quad (3.19)$$

Consequentemente, as equações (3.18) e (3.19) podem ser descritas através da seguinte operação matricial

$$\begin{bmatrix} S_j(t + 1) \\ S_j^*(t + 1) \end{bmatrix} = \begin{bmatrix} \text{cos} \omega_j & \text{sen} \omega_j \\ -\text{sen} \omega_j & \text{cos} \omega_j \end{bmatrix} \begin{bmatrix} S_j(t) \\ S_j^*(t) \end{bmatrix}.$$

Observe que essa estrutura caracteriza a equação de evolução de um MDL, com vetor de estados  $\boldsymbol{\theta}_t = (S_j(t), S_j^*(t))'$  e com matriz de evolução

$$\mathbf{H}_j = \begin{bmatrix} \text{cos} \omega_j & \text{sen} \omega_j \\ -\text{sen} \omega_j & \text{cos} \omega_j \end{bmatrix}.$$

Por conseguinte, a matriz de delineamento  $\mathbf{F} = (1, 0)'$ , de forma que  $y_t = S_j(t)$ .

Ao utilizar o modelo sazonal na forma de *Fourier* com período  $s$  par, o último harmônico  $S_{s/2}(t) = \cos(\pi t)$  assume apenas os valores  $-1$  e  $1$ . Como o componente  $S_{s/2}(t+1) = -S_{s/2}(t)$ , tem-se que o harmônico  $S_{s/2}$  simplesmente muda de sinal à medida que o tempo evolui. Nesse sentido, para  $j = s/2$ , a inclusão do harmônico conjugado no vetor de estados torna-se desnecessária, visto que o termo  $S_{s/2}(t)$  combina toda a informação relevante para descrever a evolução dos parâmetros. Nessas condições, a estrutura do MDL é definida pelo vetor de estados  $\theta_t = S_{s/2}(t)$ , pela matriz de transferência  $H_{s/2} = [-1]$  e pela matriz das observações  $F = [1]$ .

Genericamente, o modelo dinâmico sazonal considerando a forma de *Fourier*, é composto por um vetor de estados  $(s-1)$ -dimensional, uma matriz de delineamento com dimensão  $(s-1)$  e uma matriz de evolução com dimensão  $(s-1) \times (s-1)$ , tal que:

- Para período  $s$  par

$$\theta_t = \begin{bmatrix} S_1(t) \\ S_1^*(t) \\ \vdots \\ S_{s/2-1}(t) \\ S_{s/2-1}^*(t) \\ S_{s/2}(t) \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{G} = \text{bloco-diag}(\mathbf{H}_1, \dots, \mathbf{H}_{s/2}).$$

- Para período  $s$  ímpar

$$\theta_t = \begin{bmatrix} S_1(t) \\ S_1^*(t) \\ \vdots \\ S_{(s-1)/2}(t) \\ S_{(s-1)/2}^*(t) \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{G} = \text{bloco-diag}(\mathbf{H}_1, \dots, \mathbf{H}_{(s-1)/2}).$$

É interessante observar que a modelagem de padrões sazonais na forma de *Fourier* é um MDL formado por *superposição*, onde cada componente da combinação é representado por um harmônico. Tal característica torna essa abordagem bastante flexível, no sentido de proporcionar algumas vantagens. Uma muito importante, diz respeito à uma representação mais econômica, isto é, quando pertinente, o analista

pode construir um modelo mais simples, considerando apenas os harmônicos que são significativos e eliminando os demais. Isso resulta em um modelo mais parcimonioso e, conseqüentemente produz melhores previsões.

Outro efeito proveniente dessa estrutura do MDL por *superposição* é que a função de predição a  $k$  passos à frente é igual a soma das funções de predições de cada componente harmônico e depende apenas dos elementos do vetor  $\mathbf{m}_t$  e do horizonte  $k$ .

Até o momento, consideramos um modelo dinâmico puramente sazonal, sem efeitos estocásticos devidos aos erros aleatórios  $\mathbf{w}_t$ . No entanto, é possível considerar uma evolução estocástica no modelo ao admitir que a variância do sistema assume valores diferentes de zero. Nesse caso, é conveniente que a variância de evolução  $\mathbf{W}$  assumam a forma de uma matriz bloco diagonal, condizente com a estrutura da matriz de transferência dos estados  $\mathbf{G}$ . Como resultado, a componente sazonal não será mais periódica.

O MDL sazonal na forma de *Fourier* pode ser implementado no *software* R utilizando a função *dlmModTrig*, onde o período é especificado pelo argumento  $s$ .

Para ilustrar o uso dessa metodologia, consideramos a série temporal com dados mensais *AirPassengers*, já estudada anteriormente nessa dissertação. O MDL a ser ajustado inclui um modelo sazonal completo na forma de *Fourier* com período  $s = 12$  adicionado por um modelo de tendência linear local. Assim, tem-se que o MDL é dado por

$$\begin{aligned} Y_t &= \mu_t + \alpha_t + v_t, & v_t &\sim N(0, V_t), \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{t,1}, & \omega_{t,1} &\sim N(0, \sigma_\mu^2), \\ \beta_t &= \beta_{t-1} + \omega_{t,2}, & \omega_{t,2} &\sim N(0, \sigma_\beta^2), \\ \alpha_t &= \sum_{j=1}^6 S_j(t). \end{aligned}$$

Note que esse modelo é uma composição de elementos sazonais e de tendência, de

maneira que, as propriedades de *superposição* definem a seguinte estrutura matricial

$$\boldsymbol{\theta}_t = \begin{bmatrix} \mu_t \\ \beta_t \\ S_1(t) \\ S_1^*(t) \\ \vdots \\ S_6(t) \end{bmatrix}, \quad \mathbf{F}_t = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, \quad \mathbf{G}_t = \text{bloco-d iag}(\mathbf{G}_{1,t}, \mathbf{G}_{2,t}),$$

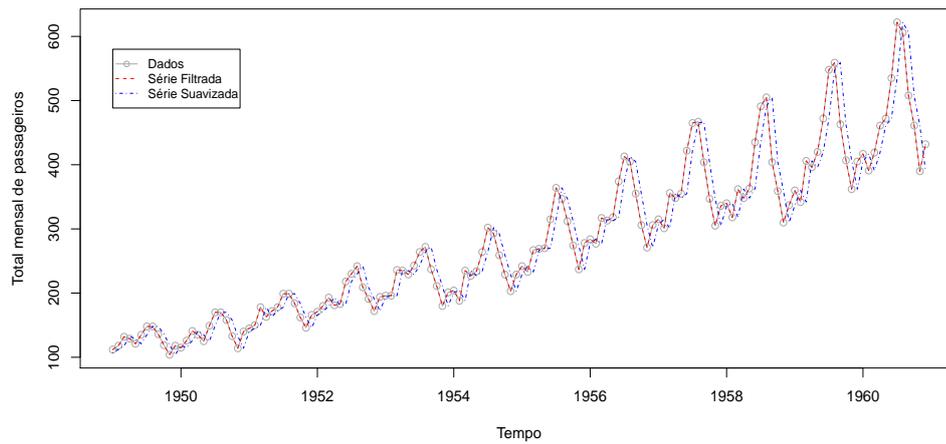
onde  $\mathbf{G}_{1,t} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$  e  $\mathbf{G}_{2,t} = \text{bloco-diag}(\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_6)$ , com  $\mathbf{G}_{1,t}$  correspondendo à matriz de transição do modelo de crescimento linear e  $\mathbf{G}_{2,t}$  representando a matriz de evolução associada ao modelo sazonal na forma de *Fourier*. A variância  $V_t$  e os elementos da matriz de variância  $\mathbf{W}_t$  foram estimados por máxima verossimilhança, usando a função *dlnMLE*.

### Exemplo de aplicação

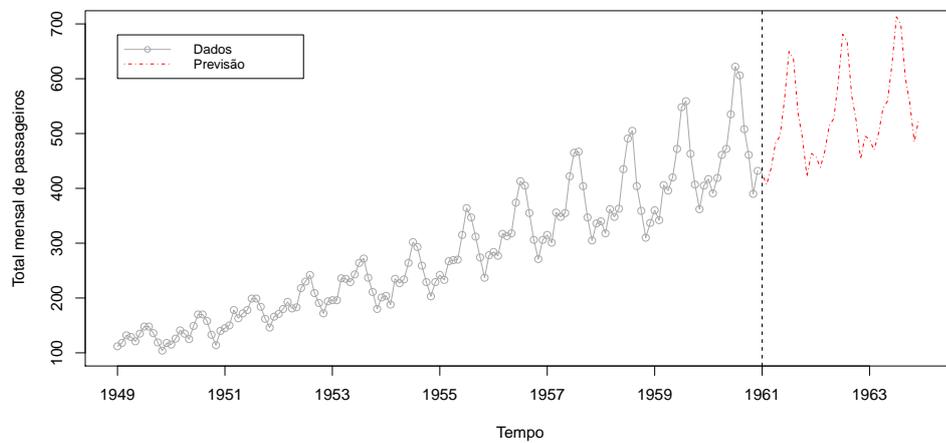
A Figura 3.3 apresenta o gráfico do ajuste desse modelo à série temporal *Air-Passengers* e também o gráfico com a previsão dos totais mensais de passageiros em companhias aéreas para 36 meses à frente. É possível notar através dessa figura, que o MDL ajustado apresenta-se adequado para modelar tais dados, visto que proporciona estimativas de filtragem e suavização bem próximas dos valores reais da série. Outro ponto importante a se considerar, é que as previsões realizadas, parece incorporar bem o comportamento sazonal e de crescimento da série.

Um modelo mais parcimonioso poderia ser utilizado considerando apenas os dois primeiros harmônicos, no entanto a economia em relação ao modelo completo não é tão expressiva, apresentando pouco ganhos no desempenho da função de predição.

O Código 5.4 no Apêndice D.3 mostra os comandos utilizados para o estudo de tal série de acordo com o modelo sazonal completo na forma de *Fourier* com crescimento linear.



(a) *Série filtrada e suavizada.*



(b) *Previsão para 36 meses à frente.*

Figura 3.3: (a) Estimativas Filtradas e Suavizadas; (b) Previsão  $k = 36$  passos à frente.

### 3.1.3 Modelos de Regressão

Foi visto, até o momento, que uma série temporal é uma *superposição* de alguns elementos, como nível, tendência e sazonalidade, que nos levam a compreender melhor o seu desenvolvimento.

No entanto, em algumas situações, a combinação desses elementos não é suficiente para o entendimento acerca do sistema, visto que, existem outros fatores que influenciam no comportamento da série temporal. Uma forma de solucionar esse impasse é incorporar ao MDL uma componente de regressão que capte o efeito que algumas

variáveis explicativas provocam na série ao longo do tempo.

Devivo a flexibilidade dos MDLs, a inclusão de variáveis independentes no modelo acontece de modo simples e pode ser pensada de forma similar a um modelo de regressão linear padrão. Nesse sentido, podemos definir o modelo de regressão dinâmica linear da seguinte forma:

**Equação das observações:**

$$Y_t = \mathbf{F}'_t \boldsymbol{\theta}_t + v_t, \quad v_t \sim N(0, V_t) \quad (3.20a)$$

**Equação do sistema:**

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \omega_t \sim N(0, \mathbf{W}_t) \quad (3.20b)$$

De acordo com a estrutura desse modelo, temos que  $Y_t$  é a série temporal observada que está sendo explicada por uma coleção de  $p$  variáveis independentes  $X_t = \{X_{t1}, \dots, X_{tp}\}$  que são assumidas conhecidas para cada tempo  $t$  ( $t = 1, \dots$ ). O vetor de regressão  $\mathbf{F}_t$  é dado por  $\mathbf{F}_t = [X_{t1}, \dots, X_{tp}]'$ , onde  $X_{t1} = 1$  e está associado à inclusão de um termo constante no modelo. A matriz de evolução  $\mathbf{G}_t$  assume uma matriz identidade de ordem  $(p \times p)$  e a matriz de variância  $\mathbf{W}_t$  uma matriz diagonal. Cada parâmetro de  $\boldsymbol{\theta}_t$  é modelado independentemente por um passeio aleatório.

A suposição de que os erros  $v_t$  e  $\mathbf{w}_t$  são mutuamente e serialmente não correlacionados é uma consequência da relação dinâmica entre  $Y_t$  e  $X_t$ . A variável explanatória  $X_t$  é não estocástica.

Denotando o vetor de parâmetros  $\boldsymbol{\theta}_t$  como  $\boldsymbol{\theta}_t = [\theta_{t1}, \dots, \theta_{tp}]'$ , a equação de observação pode ser escrita como

$$Y_t = \mu_t + v_t$$

com  $\mu_t = \mathbf{F}'_t \boldsymbol{\theta}_t = \sum_{i=1}^p \theta_{ti} X_{ti}$  sendo a resposta média. Desse modo, o MDL de regressão pode ser visualizado como uma superposição de  $p$  regressões lineares sem intercepto.

Quando  $\mathbf{W}_t = \mathbf{0}$ , o vetor de parâmetros  $\boldsymbol{\theta}_t = \boldsymbol{\theta}$  é constante ao longo do tempo, de forma que o modelo acima reduz-se a um modelo de regressão linear padrão estático. Ou seja, o MDL de regressão é uma generalização de um modelo de regressão linear

tradicional. Assim, os erros  $\mathbf{w}_t$  são os responsáveis pela estrutura dinâmica do vetor de estados.

A fim de ilustrar a utilização do MDL de regressão, em termos computacionais, foi simulada uma série temporal  $Y_t$  com  $n = 48$  observações, na forma

$$Y_t = \theta_{t1} + \theta_{t2}X_{t2} + \theta_{t3}X_{t3} + v_t.$$

O Código 5.5 no Apêndice D.3 compara os ajustes do modelo de regressão padrão com um MDL de regressão. No modelo tradicional, os parâmetros foram estimados via mínimos quadrados através da função *lm* do pacote *stats*.

A função *dmlModReg* foi utilizada para criar um modelo dinâmico cuja variância das observações foi assumida  $V = 0,978$ . Introduzimos neste modelo uma informação *a priori* acerca dos parâmetros, considerando que  $\boldsymbol{\theta}_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$ . Para estimar os coeficientes de regressão utilizamos a função *dmlFilter*. Tais coeficientes podem ser vistos como estimativas de *Bayes* admitindo uma função de perda quadrática.

Ambos os ajustes produzem resultados equivalentes no que diz respeito as estimativas dos coeficientes. Contrastes entre estes modelos são perceptíveis somente quando observamos as matrizes de variância e covariância. As estimativas da matriz de variância do MDL de regressão apresentou-se inferior em relação as estimativas obtidas com o modelo clássico. Grande parte dessa diferença provém da adição da informação *a priori* dada ao modelo dinâmico. A Figura 3.4 mostra o ajuste dos dados para ambos os modelos.

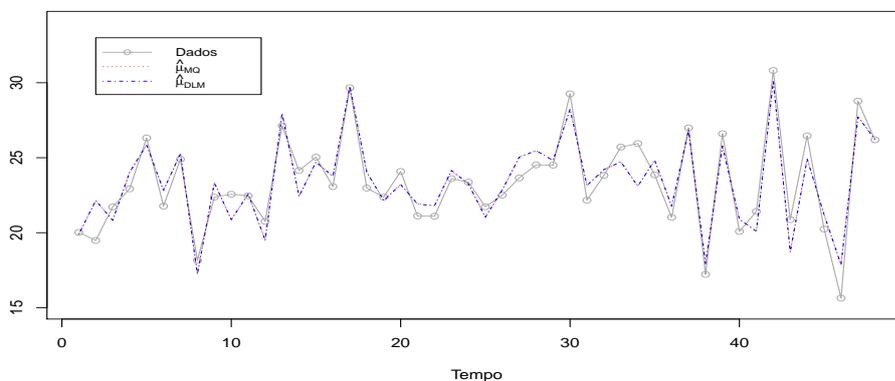


Figura 3.4: Ajustes do MDL de regressão e Modelo Padrão

## 3.2 Modelo com parâmetros desconhecidos

Toda a metodologia apresentada até o momento assume que as matrizes  $\mathbf{F}_t$ ,  $\mathbf{G}_t$ ,  $\mathbf{V}_t$  e  $\mathbf{W}_t$  são conhecidas. Essa conjectura é admitida para facilitar a compreensão sob o comportamento e propriedades do modelo. Em muitos cenários práticos, as matrizes  $\mathbf{F}_t$  e  $\mathbf{G}_t$  são conhecidas e podem ser especificadas com o auxílio de algumas estruturas para o modelo, conforme visto neste capítulo. Por outro lado, as matrizes de variância  $\mathbf{V}_t$  e  $\mathbf{W}_t$  são raramente conhecidas. Desse modo, nesta seção consideramos que tais matrizes de variância dependem de um vetor de parâmetros desconhecidos, denotado por  $\psi$ .

Na perspectiva clássica pode-se estimar  $\psi$  utilizando máxima verossimilhança, de modo que basta encontrar, algebricamente ou numericamente, os valores que maximizam a função de verossimilhança

$$p(y_1, \dots, y_n; \psi) = \prod_{t=1}^n p(y_t | y_1, \dots, y_{t-1}; \psi),$$

em que  $p(y_t | y_1, \dots, y_{t-1}; \psi)$  é a densidade condicional de  $y_t$  dado os dados até o instante  $t - 1$ , assumindo que  $\psi$  é o valor do parâmetro desconhecido.

No processo de otimização via MV é preciso ter cautela uma vez que a função de verossimilhança para o MDL pode apresentar vários máximos locais, resultando em estimativas distintas para valores iniciais diferentes. Outro inconveniente é que a superfície pode apresentá-se bastante plana, dando indícios de que o modelo não é bem identificável. Uma alternativa para contornar esse problema é simplificar o modelo, tornando-o mais parcimonioso no que diz respeito ao número de parâmetros, especialmente quando o interesse resume-se em fazer inferências e na interpretação destes parâmetros.

O pacote *dlm* fornece a função *dmlMLE* para encontrar as estimativas MV de  $\mathbf{V}_t$  e  $\mathbf{W}_t$ . Para mais detalhes veja Petris et al. (2009).

A abordagem *Bayesiana* oferece uma formulação mais consistente ao considerar que os parâmetros desconhecidos do modelo são quantidades aleatórias que seguem uma lei de probabilidade. Desse modo, baseando-se nas propriedades dos modelos de espaço de estados apresentados no Capítulo 2. Assim, para qualquer  $n \geq 1$ ,

assumimos que

$$(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n, Y_1, \dots, \boldsymbol{\theta}_n, \psi) \sim p(\boldsymbol{\theta}_0|\psi)p(\psi) \prod_{t=1}^n p(y_t|\boldsymbol{\theta}_t, \psi)p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \psi), \quad (3.21)$$

com  $p(\psi)$  descrevendo o conhecimento *a priori* sobre  $\psi$ .

Baseado no conjunto de dados  $D_t$ , inferências para o estado latente  $\boldsymbol{\theta}_s$  no tempo  $s$  e o parâmetro desconhecido  $\psi$  são realizadas calculando a distribuição conjunta *a posteriori*

$$p(\boldsymbol{\theta}_s, \psi|D_t) = p(\boldsymbol{\theta}_s|\psi, D_t)p(\psi|D_t), \quad (3.22)$$

podendo o interesse residir em  $s = t$ , representando um problema de filtragem;  $s < t$  indicando um problema de suavização; ou  $s > t$  designando um problema de predição. A densidade condicional marginal de  $\boldsymbol{\theta}_s$  é expressa a partir de

$$p(\boldsymbol{\theta}_s|D_t) = \int p(\boldsymbol{\theta}_s|\psi, D_t)p(\psi|D_t)d\psi.$$

Para modelos mais simples, quando adota-se o uso de *prioris* conjugadas, pode-se obter a forma fechada das distribuições *a posteriori* dos estados utilizando a regra de Bayes. Por outro lado, em casos cujos modelos são mais complexos, frequentemente os cálculos são analiticamente intratáveis. No entanto, existem ferramentas eficientes para aproximar a distribuição *a posteriori* de interesse, como os métodos MCMC e sequencial de Monte Carlo.

### 3.2.1 Inferência Bayesiana Conjugada para matrizes de variância desconhecidas

Como o interesse reside em estimar as matrizes de covariância  $\mathbf{V}_t$  e  $\mathbf{W}_t$ , consideraremos um caso simples onde tais matrizes são dadas por

$$\mathbf{V}_t = \sigma^2 \tilde{\mathbf{V}}_t, \quad \mathbf{W}_t = \sigma^2 \tilde{\mathbf{W}}_t, \quad \mathbf{C}_0 = \sigma^2 \tilde{\mathbf{C}}_0,$$

em que  $\sigma^2$  é um parâmetro escalar desconhecido e comum para todas as variâncias envolvidas no MDL. As matrizes  $\tilde{\mathbf{V}}_t, \tilde{\mathbf{W}}_t$ , assim como  $\mathbf{F}_t$  e  $\mathbf{G}_t$  são admitidas conhecidas. É pertinente trabalhar com o inverso de  $\sigma^2$ , ou seja,  $\varphi = 1/\sigma^2$ . Toda a incerteza

associada ao modelo está inserida no vetor de estados e no parâmetro desconhecido  $\varphi$ , de modo que o MDL assume, para qualquer  $t \geq 1$ ,

$$\begin{aligned} (Y_t | \boldsymbol{\theta}_t, \varphi) &\sim N_n(\mathbf{F}_t \boldsymbol{\theta}_t, \varphi^{-1} \tilde{\mathbf{V}}_t), \\ (\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \varphi) &\sim N_p(\mathbf{G}_t \boldsymbol{\theta}_{t-1}, \varphi^{-1} \tilde{\mathbf{W}}_t). \end{aligned}$$

Uma escolha conveniente para a distribuição a priori de  $(\boldsymbol{\theta}_0, \varphi)$  é a conjugada Normal-Gama, tal que  $(\boldsymbol{\theta}_0, \varphi) \sim NG(\mathbf{m}_0, \tilde{\mathbf{C}}_0, \gamma_0, \beta_0)$ , com

$$\varphi \sim Gama(\alpha_0, \beta_0), \quad \boldsymbol{\theta}_0 | \varphi \sim N(m_0, \varphi \tilde{\mathbf{C}}_0).$$

Então temos que o procedimento recursivo para filtragem é descrito na proposição a seguir.

**Proposição 3.2.** *Para o MDL com a estrutura descrita acima, se*

$$(\boldsymbol{\theta}_{t-1}, \varphi | D_{t-1}) \sim \mathcal{NG}(\mathbf{m}_{t-1}, \tilde{\mathbf{C}}_{t-1}, \gamma_{t-1}, \beta_{t-1})$$

onde  $t \geq 1$ , então:

(i). *A densidade a priori  $(\boldsymbol{\theta}_t, \varphi | D_{t-1})$  é Normal-Gama com parâmetros:*

$$(\boldsymbol{\theta}_t, \varphi | D_t) \sim \mathcal{NG}(\mathbf{a}_t, \tilde{\mathbf{R}}_t, \gamma_{t-1}, \beta_{t-1}),$$

onde

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1} \quad \text{e} \quad \mathbf{R}_t = \mathbf{G}_t \tilde{\mathbf{C}}_{t-1} \mathbf{G}_t' + \tilde{\mathbf{W}}_t.$$

(iii). *A densidade preditiva de  $Y_t$  dado  $D_{t-1}$  é uma  $t$ -Student, tal que:*

$$(Y_t | D_{t-1}) \sim \mathcal{T}(f_t, \tilde{Q}_t \beta_{t-1} / \gamma_{t-1}, 2\gamma_{t-1}),$$

onde

$$f_t = \mathbf{F}_t' \mathbf{a}_t \quad \text{e} \quad Q_t = \mathbf{F}_t' \tilde{\mathbf{R}}_t \mathbf{F}_t + \tilde{\mathbf{V}}_t.$$

(iv). *A densidade filtrada, ou a posteriori, de  $(\boldsymbol{\theta}_t, \varphi)$  dado  $D_t$  é Normal-Gama com parâmetros:*

$$(\boldsymbol{\theta}_t, \varphi | D_t) \sim \mathcal{NG}(\mathbf{m}_t, \tilde{\mathbf{C}}_t, \gamma_t, \beta_t),$$

com

$$\begin{aligned}\mathbf{m}_t &= \mathbf{a}_t + \tilde{\mathbf{R}}_t \mathbf{F}_t \tilde{\mathbf{Q}}_t^{-1} e_t, \\ \tilde{\mathbf{C}}_t &= \tilde{\mathbf{R}}_t - \tilde{\mathbf{R}}_t \mathbf{F}_t \tilde{\mathbf{Q}}_t^{-1} \mathbf{F}_t' \tilde{\mathbf{R}}_t, \\ \gamma_t &= \gamma_{t-1} + \frac{n}{2}, \\ \beta_t &= \beta_{t-1} + \frac{1}{2} (\mathbf{y}_t - f_t)' \tilde{\mathbf{Q}}_t (\mathbf{y}_t - f_t).\end{aligned}$$

A prova dessa proposição pode ser encontrada em West e Harrison (1997) e Petris et al. (2009). A densidade marginal filtrada dos estados é uma  $t$ -Student:

$$(\boldsymbol{\theta}|D_t) \sim \mathcal{T}(\mathbf{m}_t, \tilde{\mathbf{C}}_t \beta_t / \gamma_t, 2\gamma_t),$$

com

$$E(\boldsymbol{\theta}_t|D_t) = \mathbf{m}_t \quad \text{e} \quad \text{Var}(\boldsymbol{\theta}_t|D_t) = \frac{\beta_t}{\gamma_t - 1} \tilde{\mathbf{C}}_t.$$

Quando o parâmetro  $\sigma^2$  é conhecido, o filtro de Kalman fornece a mesma estimativa pontual  $E(\boldsymbol{\theta}_t|D_t) = \mathbf{m}_t$  e a matriz de variância  $\text{Var}(\boldsymbol{\theta}_t|D_t) = \sigma^2 \tilde{\mathbf{C}}_t$ . Por sua vez, quando  $\sigma^2$  é desconhecido, este é substituído pela esperança condicional  $\beta_t / (\gamma_t - 1)$ .

No processo de suavização, temos que

$$(\boldsymbol{\theta}_T, \varphi|D_T) \sim \mathcal{NG}(\mathbf{s}_T, \tilde{\mathbf{S}}_T, \gamma_T, \beta_T), \quad (3.23)$$

em que  $\mathbf{s}_T = \mathbf{m}_T$  e  $\tilde{\mathbf{S}}_T = \tilde{\mathbf{C}}_T$ , e

$$p(\boldsymbol{\theta}_t, \varphi|D_T) = p(\boldsymbol{\theta}_t|\varphi, D_T) p(\varphi|D_T), \quad t = 0, \dots, T. \quad (3.24)$$

Condicional a  $\varphi$ , a teoria do MDL normal é aplicável, mostrando que  $(\boldsymbol{\theta}_t, \varphi)$ , condicional a  $D_T$  tem distribuição Normal-Gama. Os parâmetros suavizados podem ser calculados de maneira similar ao caso Gaussiano. Portanto, para  $t = T - 1, \dots, 0$ , temos

$$\begin{aligned}\mathbf{s}_t &= \mathbf{m}_t + \tilde{\mathbf{C}}_t \mathbf{G}'_{t+1} \tilde{\mathbf{R}}_{t+1}^{-1} (\mathbf{s}_{t+1} - \mathbf{a}_{t+1}) \\ \tilde{\mathbf{S}}_t &= \tilde{\mathbf{C}}_t - \tilde{\mathbf{C}}_t \mathbf{G}'_{t+1} \tilde{\mathbf{R}}_{t+1}^{-1} (\tilde{\mathbf{R}}_{t+1} - \tilde{\mathbf{S}}_{t+1}) \tilde{\mathbf{R}}_{t+1}^{-1} \mathbf{G}_{t+1} \tilde{\mathbf{C}}_t.\end{aligned}$$

Então,

$$\boldsymbol{\theta}_t, \varphi|D_T \sim \mathcal{NG}(\mathbf{s}_t, \tilde{\mathbf{S}}_t, \gamma_T, \beta_T). \quad (3.25)$$

### 3.2.2 Especificação da matriz de variância $\mathbf{W}_t$

3

A matriz de variância  $\mathbf{W}_t$  controla a variação estocástica no processo de evolução do modelo sendo responsável pela estabilidade deste ao longo do tempo. Outro fator importante relacionando a  $\mathbf{W}_t$  é que esta variância adicionada na equação do sistema uma incerteza que influencia na perda de informação sob os parâmetros na transição dos tempos  $t-1$  a  $t$ . Desse modo torna-se primordial a especificação coerente da estrutura e magnitude de  $\mathbf{W}_t$  para a obtenção de bons resultados nos processos de estimação e previsão. De acordo com o filtro de Kalman dado na Proposição 2.1, a variância da distribuição a priori de  $\boldsymbol{\theta}_t$  é  $Var(\boldsymbol{\theta}_t|D_{t-1}) = \mathbf{R}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}'_t + \mathbf{W}_t$ . Se  $\mathbf{W}_t = 0$  não existe erro de evolução na equação do sistema, de modo que os estados são estáveis sem mudanças estocásticas e  $\mathbf{R}_t = \mathbf{P}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}'_t = Var(\mathbf{G}_t\boldsymbol{\theta}_{t-1}|D_{t-1})$ . Caso contrário, temos que  $\mathbf{R}_t = \mathbf{P}_t + \mathbf{W}_t$ . Nesse sentido, pode-se pensar em expressar  $\mathbf{W}_t$  como uma proporção de  $\mathbf{P}_t$ , tal que

$$\mathbf{W}_t = \frac{1-\delta}{\delta}\mathbf{P}_t$$

em que  $\delta \in (0, 1]$ . Segue que  $\mathbf{R}_t = 1/\delta\mathbf{P}_t$ , com  $1/\delta > 1$ . O parâmetro  $\delta$  é chamado de fator de desconto. Se  $\delta = 1$  então  $\mathbf{W}_t = 0$  e não há nenhuma perda de informação na evolução do estado  $\boldsymbol{\theta}_{t-1}$  a  $\boldsymbol{\theta}_t$ , de forma que  $Var(\boldsymbol{\theta}_t|D_{t-1}) = Var(\mathbf{G}_t\boldsymbol{\theta}_{t-1}|D_{t-1}) = \mathbf{P}_t$ . Quando  $\delta < 1$ , por exemplo  $\delta = 0,4$ ,  $Var(\boldsymbol{\theta}_t|D_{t-1}) = (1/0,4)Var(\mathbf{G}_t\boldsymbol{\theta}_{t-1}|D_{t-1}) = 2,5\mathbf{P}_t$ , apresentando muito incerteza. Em cenários práticos, o valor do fator de desconto é frequentemente fixo, variando entre 0,9 a 0,99, ou pode ser escolhido a partir do diagnóstico do modelo.

No caso de priori conjugada, a matriz de covariância  $\mathbf{W}_t$  pode ser especificada por fator de desconto, da seguinte forma

$$\tilde{\mathbf{W}}_t = \frac{1-\delta}{\delta}\mathbf{G}_t\tilde{\mathbf{C}}_t\mathbf{G}_t.$$

Condicional a  $\tilde{\mathbf{C}}_0$  e  $\tilde{\mathbf{V}}_t$ , o valor de  $\tilde{\mathbf{W}}_t$  pode ser calculado recursivamente para todo  $t$ .

# Capítulo 4

## Modelos Dinâmicos para a Distribuição Poisson Generalizada

Séries temporais cujas observações, a cada ponto no tempo, apresentam-se como dados de contagem são muito comuns em diversas áreas de estudo, como na saúde pública, quando deseja-se monitorar a ocorrência de câncer de mama em uma dada população através da contagem de casos ao longo do tempo, ou em geofísica, quando o interesse está em estimar o número de manchas solares na superfície do Sol. Uma distribuição que acomoda dados de contagem é a distribuição Poisson generalizada, que incorpora características importantes relacionadas à média e variância, de modo a permitir a obtenção de melhores ajustes do que o modelo Poisson padrão em razão da inclusão do parâmetro de dispersão  $\phi$ .

Grande parte dos modelos de séries temporais têm sido baseados em premissas importantes como a suposição de normalidade, que é muito restritiva para descrever muitos fenômenos na natureza, em especial, dados de contagem.

Uma opção para modelar dados não-gaussianos é realizar uma transformação nas observações para que estas se aproximem da normalidade. Entretanto, existem situações em que transformações não melhoram e nem contribuem na qualidade da previsão, além do fato de tornar difícil a interpretação dos parâmetros.

Segundo West e Harrison (1997), uma metodologia bastante útil na análise de observações ordenadas no tempo, em que a suposição de normalidade não é assumida, é a classe de Modelos Dinâmicos Lineares Generalizados (MDLG). Esta classe de mo-

delos combina os conceitos dos modelos lineares generalizados (Nelder e Wedderburn, 1972) com a estrutura dos modelos dinâmicos para a análise de séries temporais.

Apesar da relevância dessa metodologia na análise de séries temporais sob a perspectiva *Bayesiana*, o MDLG é apropriado apenas para séries cujos dados pertencem à família exponencial de Nelder e Wedderburn (1972) com parâmetro de precisão conhecido.

Buscando propor um método capaz de modelar e fazer previsões confiáveis para séries temporais adequadas a dados de contagem, desenvolvemos um modelo dinâmico *Bayesiano* para a distribuição Poisson Generalizada.

Nosso modelo foi desenvolvido na classe dos modelos dinâmicos condicionalmente Gaussianos, introduzidos por Cargnoni et al. (1997), que viabiliza o estudo de vetores individuais não-normais a partir de transformações apropriadas nos parâmetros da distribuição dos dados que, por sua vez, são modelados de acordo com os MDLs normais. Essa estrutura garante maior flexibilidade na modelagem da relação paramétrica e na evolução dos parâmetros associados aos estados ao longo do tempo, visto que, muito dos recursos dos MDLs tradicionais são empregados na formulação desse modelo. Nessas condições, é oportuno evidenciar que, após definições e transformações adequadas nos parâmetros da distribuição dos dados, as observações, a cada ponto no tempo, serão descritas conforme distribuições condicionalmente Poisson Generalizadas.

Da-Silva e Migon (2012) utilizaram a abordagem de Cargnoni et al. (1997) na formulação de um modelo dinâmico Beta Hierárquico, permitindo a formulação de um modelo para séries temporais beta multivariadas bastante flexível. O modelo dinâmico para a distribuição Poisson Generalizada (MDPG), aqui formulado, permite a incorporação dinâmica apenas no parâmetro de posição, sendo o parâmetro de forma estático e desconhecido.

As inferências para o MDPG são obtidas através do algoritmo de Monte Carlo via Cadeia de Markov (MCMC) para análise *Bayesiana*. O método combina o amostrador Gibbs, para o modelo dinâmico linear padrão, com o algoritmo de Metropolis-Hastings, para lidar com os componentes não-normais e não-lineares no esquema de simulação da distribuição *posteriori*.

## 4.1 Modelo Dinâmico para a distribuição Poisson Generalizada (MDPG)

Suponha que dispomos de uma série temporal de contagem  $y_t$ , com  $t = 1, \dots, N$ . Em cada tempo  $t$ , as observações são assumidas seguir uma distribuição condicionalmente Poisson Generalizada, conforme o modelo dado em (1.10), onde o parâmetro de posição,  $\lambda_t$ , evolui estocasticamente no tempo enquanto o parâmetro de forma,  $\phi$ , é assumido ser estático e desconhecido.

Para modelar dados segundo o modelo dinâmico Poisson Generalizado (MDPG) recorreremos à estrutura dos modelos dinâmicos condicionalmente Gaussianos, em que considera-se que após determinadas transformações nos parâmetros da distribuição dos dados, é possível o uso de propriedades do MDL normal, que facilitam a modelagem da série temporal original.

Dessa forma, faz-se necessário realizar reparametrizações na DPG que sejam apropriadas tanto computacionalmente, quanto de modo a facilitar a interpretação dos parâmetros no modelo. Outra justificativa importante para o uso de reparametrizações deve-se às restrições nos limites do parâmetro  $\phi$ , cujo suporte depende do parâmetro de locação  $\lambda_t$ . Na obtenção de reparametrizações adequadas para  $\phi$  e  $\lambda_t$ , observamos que:

- Tem-se  $a(\lambda_t, m) \leq \phi \leq 1$ , com  $a(\lambda_t, m) = \max \left\{ -1, -\frac{\lambda_t}{m} \right\}$ ;
- Tome  $a_m^t = \max \{ a(\lambda_1, m), \dots, a(\lambda_t, m) \}$  como o valor máximo entre  $a(\lambda_j, m)$ ,  $j = 1, 2, \dots, t$ , então  $a_m^t \leq \phi \leq 1$ ;
- Seja  $\lambda_t^* = \frac{1}{1+\lambda_t}$ , de modo que  $\lambda_t = \frac{1-\lambda_t^*}{\lambda_t^*}$ , com  $0 < \lambda_t^* < 1$ ;

Dessa forma, a dependência de  $\phi$  com respeito a  $\lambda_j$ ,  $j = 1, \dots, t$  é descrita em  $a_m^t$  e o parâmetro de locação  $\lambda_t$  é expresso em uma escala ( $\lambda_t^*$ ) que facilita os cálculos computacionais, por ser a estimação de  $\lambda_t^*$  mais estável do que a estimação de  $\lambda_t$ . Portanto, a distribuição Poisson Generalizada, para  $t = 1, \dots, N$ , é escrita como

$$p(y_t | \lambda_t^*, \phi) = \frac{\left( \frac{1-\lambda_t^*}{\lambda_t^*} \right) \left[ \left( \frac{1-\lambda_t^*}{\lambda_t^*} \right) + \phi y_t \right]^{y_t-1}}{y_t!} e^{\left\{ -\left( \frac{1-\lambda_t^*}{\lambda_t^*} + \phi y_t \right) \right\}}, \quad (4.1)$$

com  $0 < \lambda_t^* < 1$  e  $a_m^t \leq \phi \leq 1$ .

Observe que, agora, o limite inferior  $a_m^t$  é dado por

$$a_m^t = \max\{a(g(\lambda_1^*), m), \dots, a(g(\lambda_t^*), m)\}, \quad (4.2)$$

em que  $g(\lambda_t^*) = \frac{1-\lambda_t^*}{\lambda_t^*}$  e  $a(g(\lambda_t^*), m) = \max\left\{-1, -\frac{1-\lambda_t^*}{m\lambda_t^*}\right\}$ . Dessa forma, a esperança e variância da DPG ficam

$$E(y_t|\lambda_t^*, \phi) = \frac{1 - \lambda_t^*}{\lambda_t^*(1 - \phi)} \quad \text{e} \quad \text{Var}(y_t|\lambda_t^*, \phi) = \frac{1 - \lambda_t^*}{\lambda_t^*(1 - \phi)^3}.$$

## Modelos Dinâmicos Condicionamente Gaussianos

Os modelos dinâmicos condicionalmente Gaussianos são compostos por: (1) uma equação de observações, (2) por equações estruturais, que refletem a relação entre os parâmetros transformados e os estados latentes, (3) pela equação do sistema, que caracteriza a evolução dos parâmetros ao longo do tempo. Utilizando essa classe de modelos, o MDPG é definido por:

**Equação das Observações:**

$$p(y_t|\lambda_t^*, \phi) = \frac{\left(\frac{1-\lambda_t^*}{\lambda_t^*}\right) \left[\left(\frac{1-\lambda_t^*}{\lambda_t^*}\right) + \phi y_t\right]^{y_t-1}}{y_t!} e^{\left\{-\left(\frac{1-\lambda_t^*}{\lambda_t^*} + \phi y_t\right)\right\}},$$

com  $0 < \lambda_t^* < 1$  e  $a_m^t \leq \phi \leq 1$  em que  $a_m^t = \max\{a(g(\lambda_1^*), m), \dots, a(g(\lambda_t^*), m)\}$  e  $a(g(\lambda_t^*), m) = \max\left\{-1, -\frac{1-\lambda_t^*}{m\lambda_t^*}\right\}$ ,  $t = 1, 2, \dots, N$ .

**Equação Estrutural:**

$$\eta_t = \mathbf{F}_t \boldsymbol{\theta}_t + v_t, \quad v_t \sim N(0, V),$$

com  $\eta_t = h(\lambda_t^*) \in \mathcal{R}$ .

**Equação do Sistema:**

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}).$$

Assume-se que os erros  $v_t$  e  $\mathbf{w}_t$  são mutuamente e internamente independentes.

A transformação  $\eta_t = h(\lambda_t^*)$  descreve uma função de ligação que relaciona os estados  $\boldsymbol{\theta}_t$  aos parâmetros  $\lambda_t^*$  e  $\phi$  do modelo dinâmico Poisson generalizado. Como  $\lambda_t^* \in (0, 1)$ , uma transformação adequada é a *logit*. Dessa forma, tome

$$\eta_t = h(\lambda_t^*) = \log \left( \frac{\lambda_t^*}{1 - \lambda_t^*} \right).$$

Portanto,  $\lambda_t^* = \frac{\exp(\eta_t)}{1 + \exp(\eta_t)}$ .

O parâmetro estrutural  $\eta_t$  poderia também ser uma função do parâmetro natural  $\lambda_t$ , assumindo  $h(\lambda_t)$  como a função de ligação canônica, isto é,  $\eta_t = \log \lambda_t$ . No entanto, essa transformação apresenta-se instável, no sentido de gerar valores extremos para  $\eta_t$  que prejudicam na aplicação da metodologia, uma vez que a probabilidade  $p(y_t | \eta_t, \phi)$  assumirá quase sempre valor igual a zero.

Agora, para simplificar as análises, o modelo descrito acima será reescrito em termos do parâmetro  $\eta_t$ , considerando a função de ligação *logit*, ou seja  $\eta_t = h(\lambda_t^*) = \log \left( \frac{\lambda_t^*}{1 - \lambda_t^*} \right)$ .

### a. Equação das Observações

Para  $t = 1, \dots, N$ , seja

$$p(y_t | \eta_t, \phi) = \frac{e^{-\eta_t} (e^{-\eta_t} + \phi y_t)^{y_t - 1}}{y_t!} e^{-(e^{-\eta_t} + \phi y_t)},$$

com  $\eta_t \in \mathcal{R}$  e  $a_m^N \leq \phi \leq 1$ , em que  $a_m^N = \max\{a(g(\eta_1), m), \dots, a(g(\eta_N), m)\}$ , sendo  $g(\eta_t) = e^{-\eta_t}$  e  $a(g(\eta_t), m) = \max\left\{-1, -\frac{e^{-\eta_t}}{m}\right\}$ .

Note que o limite inferior de  $\phi$  agora considera todos os  $N$  valores do vetor de parâmetros estruturais  $\boldsymbol{\eta}$ , pois como  $\phi$  é estático, não podendo variar no tempo com  $\eta_t$ , deve satisfazer, simultaneamente, a este conjunto de restrições.

### b. Equações Estruturais

O parâmetro estrutural  $\eta_t$  é modelado individualmente como

$$\eta_t = \mathbf{F}_t \boldsymbol{\theta}_t + v_t, \quad v_t \sim N(0, V), \quad (4.3)$$

em que  $\boldsymbol{\theta}_t$  é um vetor  $p$ -dimensional dos parâmetros de estados no tempo  $t$ , que representam características específicas da série. A equação em (4.3) incorpora efeitos estocásticos individuais para o parâmetro transformado, através do erro  $v_t$  que, por

sua vez, é admitido seguir uma distribuição Normal com média 0 e variância  $V$ . A matriz de delineamento  $\mathbf{F}_t$ , com dimensão  $(1 \times p)$ , é assumida conhecida e descreve a relação entre  $\eta_t$  e o vetor de estados.

### c. Equação do Sistema

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w} \sim N(0, \mathbf{W}), \quad (4.4)$$

descreve a evolução dos parâmetros dos estados ao longo do tempo. A matriz de transferência ou de evolução dos estados  $\mathbf{G}_t$  tem dimensão  $(p \times p)$ , e  $\mathbf{w}_t$  é um vetor de erros aleatórios  $p$ -dimensional com matriz de covariância  $\mathbf{W} = \text{bloco-diag}(\mathbf{W}_1, \dots, \mathbf{W}_k)$ , sendo bloco-diagonal com  $k$  efeitos e dimensão  $(p \times p)$ . Assume-se que as sequências de erros  $v_t$  e  $\mathbf{w}_t$  são mutuamente independentes.

Quando as distribuições *a priori* para  $\boldsymbol{\theta}_0$ ,  $V$  e  $\mathbf{W}$ , no tempo  $t = 0$ , são especificadas, dispõe-se de um modelo completo, baseado na informação inicial  $D_0$ . Desta forma, assume-se que  $\boldsymbol{\theta}_0$ ,  $V$  e  $\mathbf{W}$  são mutuamente independentes, de modo que, o estado  $\boldsymbol{\theta}_0$  tem distribuição *a priori*  $N(\mathbf{m}_0, \mathbf{C}_0)$ , a variância  $V$  segue uma distribuição *a priori* Gama Inversa e, por fim, propõe-se que cada bloco da matriz  $\mathbf{W}$  tenha distribuição *a priori* Wishart Inversa.

O modelo especificado pelas equações (4.3) e (4.4) corresponde ao modelo dinâmico linear normal com variâncias desconhecidas. A diferença é que, ao invés das observações, é o parâmetro estrutural que está sendo modelado conforme um MDL padrão. O vetor  $\boldsymbol{\theta}_t$  carrega as informações dos estados, como tendência e sazonalidade, relacionados ao parâmetro  $\eta_t$  que, por sua vez, fornece as características e particularidades que servirão de suporte para a estimação e previsão da série temporal.

## 4.2 Cálculos para obtenção das distribuições condicionais completas *a posteriori*

Assim como na maioria dos modelos de séries temporais, no modelo dinâmico para a distribuição Poisson Generalizada a principal tarefa é fazer inferências sobre os estados latentes e previsão de observações futuras. Para esse fim, é necessário obter

estimativas a *posteriori* dos parâmetros do modelo com o objetivo de realizar o ciclo de inferência sequencial que consiste em: evolução, predição e atualização.

Como nosso estudo baseia-se no tratamento de séries temporais cujas observações, a cada ponto no tempo, seguem uma DPG que, por sua vez, não pertence à família exponencial, e não tem uma família conjugada conhecida, o cálculo da distribuição a *posteriori* conjunta, ou, mesmo das distribuições condicionais completas a *posteriori*, torna-se muito complexo. Uma alternativa é o uso de métodos de simulação.

Para tratar dessa dificuldade, descreveremos o algoritmo de Monte Carlo via Cadeias de Markov (MCMC) que possibilita a obtenção de amostras a partir da distribuição a *posteriori* conjunta dos estados e parâmetros do modelo.

Sejam as quantidades  $\boldsymbol{\eta}$ ,  $\mathbf{y}$  e  $\boldsymbol{\theta}$  definidas como segue:  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_N)$ ,  $\mathbf{y} = (y_1, \dots, y_N)$  e  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$ . Considere, também, por questão de facilidade computacional, as matrizes de precisão  $\Phi_0 = V^{-1}$  e  $\Phi = \text{bloco-diag}(\Phi_1, \dots, \Phi_k)$ , com  $\Phi_i = W_i^{-1}$ ,  $i = 1, \dots, k$ . Assume-se que  $\Phi_0$  segue uma distribuição *priori* Gama, de modo que,  $\Phi_0 \sim G(\nu_0, S_0)$ . Similarmente, temos que  $\Phi_1, \dots, \Phi_k$  seguem distribuição Wishart independentes, ou seja,  $\Phi_i \sim W(\nu_i, S_i)$ ,  $i = 1, \dots, k$ . Com base nessas definições, o interesse consiste em calcular e sumarizar a distribuição *posteriori* conjunta

$$\begin{aligned} p(\boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \Phi, \phi | \mathbf{y}) &\propto p(y_1, \dots, y_N | \phi, \eta_1, \dots, \eta_N) p(\phi | \eta_1, \dots, \eta_N) \\ &\times p(\eta_1, \dots, \eta_N | \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N, \Phi_0) p(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N | \Phi) \\ &\times p(\Phi_0) p(\Phi) p(\boldsymbol{\theta}_0 | \mathbf{m}_0, \mathbf{C}_0). \end{aligned}$$

Para realizar este cálculo utilizaremos o amostrador de Gibbs, que requer a especificação das distribuições condicionais completas a *posteriori*

$$\begin{aligned} p(\boldsymbol{\theta} | \boldsymbol{\eta}, \Phi_0, \Phi, \phi, \mathbf{y}) &\longleftrightarrow p(\Phi_0, \Phi | \boldsymbol{\eta}, \boldsymbol{\theta}, \phi, \mathbf{y}) \longleftrightarrow p(\boldsymbol{\eta} | \boldsymbol{\theta}, \Phi_0, \Phi, \phi, \mathbf{y}) \longleftrightarrow \\ &p(\phi | \boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \Phi, \mathbf{y}). \end{aligned} \quad (4.5)$$

Entretanto, devido a complexidade do modelo, é possível amostrar diretamente apenas a primeira e a segunda condicionais. Uma alternativa para calcular as *posteriori*  $p(\boldsymbol{\eta} | \boldsymbol{\theta}, \Phi_0, \Phi, \phi, \mathbf{y})$  e  $p(\phi | \boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \Phi, \mathbf{y})$  é incorporar nas iterações de Gibbs um passo do algoritmo de Metropolis-Hastings. O algoritmo de Metropolis-Hastings, combinado com Gibbs, permite a amostragem dessas condicionais completas através de

distribuições propostas que possibilitam a atualização iterativa dos parâmetros  $\boldsymbol{\eta}$  e  $\phi$ , com uma determinada probabilidade de aceitação.

A ordem como essas condicionais estão dispostas em (4.5) é imprescindível no esquema MCMC, posto que existe uma relação de dependência paramétrica (uma hierarquia) imposta na definição do modelo dinâmico para a distribuição Poisson Generalizada que estabelece essa sequência na simulação das condicionais completas.

#### 4.2.1 A distribuição condicional completa do vetor de estados latentes $\boldsymbol{\theta}$ , $p(\boldsymbol{\theta}|\boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi, \mathbf{y})$

O vetor de estados  $\boldsymbol{\theta}$  é condicionalmente independente de  $\mathbf{y}$ , dado  $\boldsymbol{\eta}$ . Logo, a distribuição condicional completa de  $\boldsymbol{\theta}$  é dada por

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi, \mathbf{y}) &= p(\boldsymbol{\theta}|\boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi) \\ &= p(\boldsymbol{\theta}_N|\boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi) \prod_{t=0}^{N-1} p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \dots, \boldsymbol{\theta}_N, \boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi). \end{aligned}$$

Esse resultado, combinado com a estrutura de independência condicional proveniente dos modelos de espaço de estados (veja propriedade **P.6** na Seção 2.2 do Capítulo 2), possibilitam escrever a distribuição a *posteriori* para  $\boldsymbol{\theta}$ :

$$p(\boldsymbol{\theta}|\boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi) = p(\boldsymbol{\theta}_N|\boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi) \prod_{t=0}^{N-1} p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi).$$

Aplicando o teorema de Bayes, obtém-se

$$p(\boldsymbol{\theta}|\boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi) \propto p(\boldsymbol{\theta}_N|\boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi) \prod_{t=0}^{N-1} p(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t, \boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi) p(\boldsymbol{\theta}_t|\boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi).$$

Em virtude dessa estrutura, resultados padrão dos modelos dinâmicos lineares Gaussianos podem ser aplicados para a amostragem a partir da distribuição a posteriori condicional completa de  $\boldsymbol{\theta}$  utilizando o algoritmo *Forward-Filtering, Backward-Sampling* (FFBS), desenvolvido por Carter e Kohn (1994) e Frühwirth-Schnatter (1994). Tal algoritmo fornece uma estratégia eficiente para amostrar, sequencialmente, os vetores de estados  $\boldsymbol{\theta}_N, \boldsymbol{\theta}_{N-1}, \dots, \boldsymbol{\theta}_0$ , tal como segue:

#### Amostragem de $\boldsymbol{\theta}$ utilizando o algoritmo FFBS

- a. Para  $t = 0, 1, \dots, N$ , obter a distribuição filtrada normal  $p(\boldsymbol{\theta}_t | \boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi)$  e seus momentos  $\mathbf{m}_t$  e  $\mathbf{C}_t$ , aplicando o algoritmo recursivo e sequencial de filtro de Kalman para MDLs normais;
- b. No tempo  $t = N$ , amostrar o vetor  $\boldsymbol{\theta}_N$  a partir de  $(\boldsymbol{\theta}_N | \boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi) \sim N(\mathbf{m}_N, \mathbf{C}_N)$ ;
- c. Para cada  $t = N - 1, \dots, 1, 0$ , amostrar, retrospectivamente e sequencialmente, o vetor de estados  $\boldsymbol{\theta}_t$  de  $p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, \boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi)$ , em que o estado  $\boldsymbol{\theta}_{t+1}$  já está disponível, uma vez que este já foi amostrado.

Como visto no Capítulo 2, temos que  $(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, \boldsymbol{\eta}, \Phi_0, \boldsymbol{\Phi}, \phi) \sim N(\mathbf{b}_t, \mathbf{B}_t)$ , com  $\mathbf{b}_t = \mathbf{m}_t + \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} (\boldsymbol{\theta}_{t+1} - \mathbf{G}_{t+1} \mathbf{m}_t)$  e  $\mathbf{B}_t = \mathbf{C}_t - \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1} \mathbf{G}_{t+1} \mathbf{C}'_t$ .

Resumindo, o processo de amostragem de  $\boldsymbol{\theta}$  começa utilizando-se o filtro de Kalman, para a obtenção da distribuição filtrada, iniciando no tempo  $t = 0$ , até  $t = N$ . Em seguida, calculando-se e guardando as medidas  $\mathbf{m}_t$  e  $\mathbf{C}_t$ , bem como as quantidades  $\mathbf{a}_t$  e  $\mathbf{R}_t$  de cada etapa. Em  $t = N$ , o vetor de estados final  $\boldsymbol{\theta}_N$  é amostrado, e assim, retrocedemos, sequencialmente, no tempo, calculando os elementos  $\mathbf{b}_t$  e  $\mathbf{B}_t$  de cada passo, obtendo um desenho  $\boldsymbol{\theta}_N, \boldsymbol{\theta}_{N-1}, \dots, \boldsymbol{\theta}_0$  da distribuição *posteriori* condicional completa.

#### 4.2.2 As distribuições condicionais completas das matrizes de variância $\Phi_0$ e $\boldsymbol{\Phi}$ , $p(\Phi_0 | \boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\Phi}, \phi, \mathbf{y})$ e $p(\boldsymbol{\Phi} | \boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \phi, \mathbf{y})$

Neste trabalho assume-se que a distribuição a *priori* de  $\Phi_0$  é uma Gama, ou seja,  $\Phi_0 \sim G(\nu_0, S_0)$  e que cada bloco  $\boldsymbol{\Phi}_i$  tem distribuição *priori* Wishart, tal que  $\boldsymbol{\Phi}_i \sim W(\nu_i, \mathbf{S}_i)$ . As matrizes  $\mathbf{S}_i$  são simétricas e positivas definidas de dimensão  $p_i \times p_i$ . Desse modo, as distribuições condicionais completas de  $\Phi_0$  e  $\boldsymbol{\Phi}_i$ ,  $i = 1, \dots, k$ , podem ser obtidas como seguem:

a. Distribuição condicional completa a *posteriori* de  $\Phi$

$$\begin{aligned}
p(\Phi_i | \boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \phi, \mathbf{y}) &\propto p(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N | \Phi_i) p(\Phi_i) \\
&\propto \left( \prod_{t=1}^N \prod_{l=1}^k |\Phi_l|^{1/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta}_t - \mathbf{G}_t \boldsymbol{\theta}_t)' \Phi (\boldsymbol{\theta}_t - \mathbf{G}_t \boldsymbol{\theta}_t) \right\} \right) \\
&\times |\Phi_i|^{v_i - (p_i + 1)/2} \exp \{ -tr(\mathbf{S}_i \Phi_i) \} \\
&\propto |\Phi_i|^{N/2 + v_i - (p_i + 1)/2} \exp \left\{ -tr \left( \frac{1}{2} \sum_{t=1}^N (\boldsymbol{\theta}_t - \mathbf{G}_t \boldsymbol{\theta}_t) (\boldsymbol{\theta}_t - \mathbf{G}_t \boldsymbol{\theta}_t)' \Phi \right) \right\} \\
&\times \exp \{ -tr(\mathbf{S}_i \Phi_i) \}.
\end{aligned}$$

Seja  $\mathbf{S}\mathbf{S}_t = (\boldsymbol{\theta}_t - \mathbf{G}_t \boldsymbol{\theta}_t) (\boldsymbol{\theta}_t - \mathbf{G}_t \boldsymbol{\theta}_t)'$ , considere a seguinte partição desta matriz em relação a  $\Phi$

$$\mathbf{S}\mathbf{S}_t = \begin{bmatrix} \mathbf{S}\mathbf{S}_{11,t} & \dots & \mathbf{S}\mathbf{S}_{1k,t} \\ \vdots & \ddots & \vdots \\ \mathbf{S}\mathbf{S}_{k1,t} & \dots & \mathbf{S}\mathbf{S}_{kk,t} \end{bmatrix}.$$

Desse modo,  $tr(\mathbf{S}\mathbf{S}_t \Phi) = \sum_{i=1}^k tr(\mathbf{S}\mathbf{S}_{ii,t} \Phi_i)$ , de forma que a distribuição condicional completa de  $\Phi$  é proporcional a

$$(\Phi_i | \boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \phi, \mathbf{y}) \propto |\Phi_i|^{N/2 + v_i - (p_i + 1)/2} \exp \left\{ -tr \left( \left( \frac{1}{2} \mathbf{S}\mathbf{S}_i + \mathbf{S}_i \right) \Phi_i \right) \right\},$$

com  $\mathbf{S}\mathbf{S}_i = \sum_{t=1}^N \mathbf{S}\mathbf{S}_{ii,t}$ . Logo,  $(\Phi_i | \boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \phi, \mathbf{y}) \sim W(N/2 + v_i, \frac{1}{2} \mathbf{S}\mathbf{S}_i + \mathbf{S}_i)$ .

b. Distribuição condicional completa a *posteriori* de  $\Phi_0$

$$\begin{aligned}
p(\Phi_0 | \boldsymbol{\eta}, \boldsymbol{\theta}, \Phi, \phi, \mathbf{y}) &\propto p(\eta_1, \dots, \eta_N | \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N, \Phi_0) p(\Phi_0) \\
&\propto \left( \prod_{t=1}^N p(\eta_t | \boldsymbol{\theta}_t, \Phi_0) \right) p(\Phi_0) \\
&\propto \left( \prod_{t=1}^N N(\eta_t | \mathbf{F}_t \boldsymbol{\theta}_t, \Phi_0^{-1}) \right) W(\Phi_0 | v_0, S_0) \\
&\propto \prod_{t=1}^N |\Phi_0|^{1/2} \exp \left\{ -\frac{\Phi_0}{2} (\eta_t - \mathbf{F}_t \boldsymbol{\theta}_t)^2 \right\} \\
&\times |\Phi_0|^{v_0 - 1} \exp \{ -tr(S_0 \Phi_0) \} \\
&\propto |\Phi_0|^{N/2 + v_0 - 1} \exp \left\{ -\Phi_0 \left[ \frac{1}{2} \sum_{t=1}^N (\eta_t - \mathbf{F}_t \boldsymbol{\theta}_t)^2 + S_0 \right] \right\}.
\end{aligned}$$

Fazendo  $SS_{\eta_t} = (\eta_t - \mathbf{F}_t \boldsymbol{\theta}_t)^2$ , obtemos

$$\begin{aligned} p(\Phi_0 | \boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\Phi}, \phi, \mathbf{y}) &\propto |\Phi_0|^{N/2+v_0-1} \exp \left\{ -\Phi_0 \left( \frac{1}{2} \sum_{t=1}^N SS_{\eta_t} + S_0 \right) \right\} \\ &\propto |\Phi_0|^{N/2+v_0-1} \exp \left\{ -\Phi_0 \left( \frac{1}{2} SS_{\eta_t} + S_0 \right) \right\}, \end{aligned}$$

em que  $SS_{\eta_t} = \sum_{t=1}^N SS_{\eta_t}$  e, portanto  $(\Phi_0 | \boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\Phi}, \phi, \mathbf{y}) \sim G(N/2 + v_0, \frac{1}{2} SS_{\eta_t} + S_0)$ .

### 4.2.3 Distribuição condicional completa do parâmetro estrutural $\boldsymbol{\eta}$ , $p(\boldsymbol{\eta} | \boldsymbol{\theta}, \Phi_0, \boldsymbol{\Phi}, \phi, \mathbf{y})$

A amostragem do vetor  $\boldsymbol{\eta}$  consiste na simulação a partir da distribuição

$$p(\eta_t | \boldsymbol{\theta}_t, \Phi_0, \boldsymbol{\Phi}, \phi, y_t) \propto p(y_t | \eta_t, \phi) p(\eta_t | \boldsymbol{\theta}_t, \Phi_0), \quad (4.6)$$

para  $t = 1, \dots, N$ , sendo que o primeiro termo do lado direito é a verossimilhança da distribuição Poisson Generalizada, enquanto que o segundo termo é a *priori* normal  $(\eta_t | \boldsymbol{\theta}_t, \Phi_0) \sim N(\eta_t | \mathbf{F}_t \boldsymbol{\theta}_t, \Phi_0^{-1})$ .

Uma vez que a distribuição  $p(\eta_t | \boldsymbol{\theta}_t, \Phi_0, \boldsymbol{\Phi}, \phi, y_t)$  não tem forma analítica fechada, implementaremos o algoritmo de Metropolis-Hastings (Metropolis et al. (1953) e Hastings (1970)) nas iterações de Gibbs para amostragem de tal distribuição condicional completa.

A ideia consiste em gerar valores de  $\eta_t$  para  $t = 1, \dots, N$ , a partir de uma distribuição proposta  $q(\eta_t^* | \eta_t)$  e aceitar esses valores com uma determinada probabilidade que, geralmente, varia no intervalo de 25% a 55%.

Na construção de uma distribuição geradora de candidatos  $q(\eta_t^* | \eta_t)$  para  $\eta_t$  utilizamos uma aproximação Gaussiana para a distribuição condicional completa a *posteriori* de  $\eta_t$ , isto é,

$$(\eta_t | \boldsymbol{\theta}_t, \Phi_0, \boldsymbol{\Phi}, \phi, D_t) \approx N(\hat{\eta}_t, \hat{\Sigma}_t) \quad (4.7)$$

Os valores de  $\hat{\eta}_t$  e  $\hat{\Sigma}_t$  são obtidos de acordo com McCormick et al. (2012), de modo que  $\hat{\eta}_t$  é estimado com base em um procedimento de Newton-Raphson e  $\hat{\Sigma}_t$  é aproximado através da matriz de informação observada no tempo  $t$ .

Portanto, utiliza-se

$$q(\eta_t^* | \eta_t) \stackrel{d}{=} N(\hat{\eta}_t, \hat{\Sigma}_t),$$

de modo que, para a  $j$ -ésima observação MCMC,

$$\hat{\eta}_t^{(j)} = \hat{\eta}_{t-1}^{(j-1)} - [D^2 l(\hat{\eta}_{t-1}^{(j-1)})]^{-1} D(l(\hat{\eta}_{t-1}^{(j-1)}))$$

e

$$\hat{\Sigma}_t^{(j)} = -[D^2 l(\hat{\eta}_{t-1}^{(j-1)})]^{-1},$$

onde

$$\hat{\eta}_t = \hat{\eta}_t^{(j)} = \hat{\eta}_{t-1}^{(j-1)} - [D^2(l(\hat{\eta}_{t-1}^{(j-1)}))]^{-1} D(l(\hat{\eta}_{t-1}^{(j-1)})) \quad \text{e} \quad (4.8)$$

$$\hat{\Sigma}_t = -[D^2(l(\hat{\eta}_{t-1}^{(j-1)}))]^{-1}, \quad (4.9)$$

em que  $D(l(\hat{\eta}_{t-1}^{(j-1)}))$  e  $D^2(l(\hat{\eta}_{t-1}^{(j-1)}))$  são respectivamente as derivadas de primeira e segunda ordem de  $l(\eta_t) = \log[p(y_t|\eta_t, \phi)p(\eta_t|\boldsymbol{\theta}_t, \Phi_0)]$  avaliadas no ponto  $\eta_t = \hat{\eta}_{t-1}$ .

Essa aproximação da distribuição proposta para uma normal é bastante razoável, visto que os estimadores obtidos gozam das propriedades assintóticas dos estimadores de máxima verossimilhança. Nesse caso, o passo de Metropolis-Hastings para  $\boldsymbol{\eta}$  na análise MCMC se resume nos passos:

- a. Gerar um novo valor de  $\eta_t^*$  da distribuição  $q(\eta_t^*|\eta_t^{(j-1)}) \stackrel{d}{=} N(\eta_t^*|\hat{\eta}_t, \hat{\Sigma}_t)$ .
- b. Calcular a probabilidade de aceitação  $\alpha(\eta_t^{(j-1)}, \eta_t^*) = \min\{1, R_{\eta_t}\}$ , onde

$$R_{\eta_t} = \frac{p(\eta_t^*|\cdot) q(\eta_t^{(j-1)}|\eta_t^*)}{p(\eta_t^{(j-1)}|\cdot) q(\eta_t^*|\eta_t^{(j-1)})} = \frac{p(\eta_t^*|\cdot)}{p(\eta_t^{(j-1)}|\cdot)}.$$

com  $p(\eta_t^*|\cdot) = p(y_t|\eta_t^*, \phi)p(\eta_t^*|\boldsymbol{\theta}_t, \Phi_0)$  e  $p(\eta_t^{(j-1)}|\cdot) = p(y_t|\eta_t^{(j-1)}, \phi)p(\eta_t^{(j-1)}|\boldsymbol{\theta}_t, \Phi_0)$ .

Observe que a distribuição proposta é simétrica, propiciando a simplificação em  $R_{\eta_t}$ .

- c. Gerar  $u \sim U(0, 1)$  e descrever  $\eta_t^{(j)}$  segundo o critério

$$\eta_t^{(j)} = \begin{cases} \eta_t^*, & \text{se } u < \alpha(\eta_t^{(j-1)}, \eta_t^*) \\ \eta_t^{(j-1)}, & \text{caso contrário.} \end{cases}$$

#### 4.2.4 Distribuição condicional completa do parâmetro de forma

$$\phi, p(\phi|\boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \boldsymbol{\Phi}, \mathbf{y})$$

O parâmetro  $\phi$  exerce um importante papel na DPG, visto que foi a partir da inserção deste que uma generalização da distribuição Poisson foi formulada, permitindo a

modelagem de uma ampla gama de dados. No entanto, esse parâmetro apresenta algumas restrições que devem ser consideradas cuidadosamente.

No modelo dinâmico Poisson Generalizado, assumimos que  $\phi$  é um parâmetro estático e desconhecido que, portanto, deve ser estimado. Como o suporte de  $\phi$  depende do parâmetro  $\lambda_t$  que, por sua vez, é dinâmico no tempo, definimos, no início deste capítulo, que  $a_m^N \leq \phi \leq 1$ , com  $a_m^N = \max\{a(g(\eta_1), m), \dots, a(g(\eta_N), m)\}$ , sendo  $g(\eta_t) = e^{-\eta_t}$  e  $a(g(\eta_t), m) = \max\left\{-1, -\frac{e^{-\eta_t}}{m}\right\}$ .

Esse ajuste no limite inferior do parâmetro possibilita o uso de transformações adequadas de  $\phi$  que facilitam na sua estimação. O procedimento de estimação utilizado é o algoritmo de Metropolis-Hastings dentro de Gibbs, que possibilita amostrar a distribuição condicional completa a *posteriori*  $p(\phi|\boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \boldsymbol{\Phi}, \mathbf{y})$ , a partir de uma distribuição proposta, que descrevemos a seguir.

Nesse cenário, trabalhamos com a amostragem de uma transformação de  $\phi$ . Primeiramente, transformamos  $\phi \in [a_m^N, 1]$  para uma variável  $H_1(\phi|\boldsymbol{\eta}) \in (0, 1)$ , isto é,

$$H_1(\phi|\boldsymbol{\eta}) = \frac{\phi - a_m^N}{1 - a_m^N}, \quad (4.10)$$

é uma transformação de  $\phi$  com  $H_1(\phi|\boldsymbol{\eta}) \in (0, 1)$ .

O segundo passo consiste em obter uma transformação de  $H_1(\phi|\boldsymbol{\eta})$  de  $(0, 1)$  em  $\mathcal{R}$ . Tomando a transformação *logit* de  $H_1(\phi, \boldsymbol{\eta})$ , tem-se:

$$H_2(\phi|\boldsymbol{\eta}) = \log\left(\frac{H_1(\phi|\boldsymbol{\eta})}{1 - H_1(\phi|\boldsymbol{\eta})}\right), \quad (4.11)$$

em que  $H_2(\phi|\boldsymbol{\eta}) \in \mathcal{R}$ .

É através da transformação  $H_2(\phi|\boldsymbol{\eta})$  que o algoritmo de MH será construído. Assim, considere que  $H_2(\phi|\boldsymbol{\eta})$  tem distribuição a *priori* hierárquica normal assimétrica, com parâmetro de locação  $\xi_0$ , parâmetro de escala  $\omega_0$  e parâmetro de forma (assimetria)  $\delta_0$ , ou seja,

$$H_2(\phi|\boldsymbol{\eta}) \sim \text{Normal Assimétrica}(\xi_0, \omega_0, \delta_0). \quad (4.12)$$

Admitimos que a função geradora de candidatos de  $H_2(\phi|\boldsymbol{\eta})$  também segue uma distribuição normal assimétrica, tal que,  $q(H_2^*|H_2^{(j-1)}) \stackrel{d}{=} \text{Normal Assimétrica}(\xi = H_2^{(j-1)}, \omega, \delta)$ . Portanto, o algoritmo de Metropolis-Hastings para  $H_2(\phi, \boldsymbol{\eta})$  pode ser

resumido nas seguintes etapas:

a. Gerar valores de  $H_2^*(\phi|\boldsymbol{\eta})$  de  $q(H_2^*|H_2^{(j-1)}) \stackrel{d}{=} \text{Normal Assimétrica}(H_2^{(j-1)}, \omega, \delta)$ ;

b. Obter a probabilidade de aceitação  $\alpha(H_2^{(j-1)}, H_2^*) = \min\{1, R_{H_2}\}$ , em que

$$R_{H_2} = \frac{p(H_2^*|\cdot) q(H_2^{(j-1)}|H_2^*)}{p(H_2^{(j-1)}|\cdot) q(H_2^*|H_2^{(j-1)})},$$

onde

$$\begin{aligned} p(H_2^*|\cdot) &= \left[ \prod_{t=1}^N p(y_t|\phi^*, \eta_t) \right] p(H_2^*|\xi_0, \omega_0, \delta_0) \quad \text{e} \\ p(H_2^{(j-1)}|\cdot) &= \left[ \prod_{t=1}^N p(y_t|\phi^{(j-1)}, \eta_t) \right] p(H_2^{(j-1)}|\xi_0, \omega_0, \delta_0). \end{aligned}$$

c. Gerar  $u \sim U(0, 1)$  e descrever  $H_2^{(j)}$  segundo o critério

$$H_2^{(j)}(\phi|\boldsymbol{\eta}) = \begin{cases} H_2^{(*)}(\phi|\boldsymbol{\eta}), & \text{se } u < \alpha(H_2^{(j-1)}, H_2^*), \\ H_2^{(j-1)}(\phi|\boldsymbol{\eta}), & \text{caso contrário.} \end{cases}$$

Note que, para obter a amostra na escala do verdadeiro parâmetro  $\phi$ , basta considerar as seguintes transformações inversas:

$$H_1^*(\phi|\boldsymbol{\eta}) = \frac{e^{H_2^*(\phi|\boldsymbol{\eta})}}{1 + e^{H_2^*(\phi|\boldsymbol{\eta})}} \quad \text{e} \quad H_1^{(j-1)}(\phi|\boldsymbol{\eta}) = \frac{e^{H_2^{(j-1)}(\phi|\boldsymbol{\eta})}}{1 + e^{H_2^{(j-1)}(\phi|\boldsymbol{\eta})}}.$$

Logo, temos que

$$\phi^* = a_m^N + H_1^*(\phi|\boldsymbol{\eta})[1 - a_m^N] \quad \text{e} \quad \phi^{(j-1)} = a_m^N + H_1^{(j-1)}(\phi|\boldsymbol{\eta})[1 - a_m^N].$$

A escolha de uma distribuição candidata para gerar amostras de  $p(\phi|\boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \boldsymbol{\Phi}, \mathbf{y})$  não foi uma tarefa simples. Foram necessárias muitas tentativas, simulações, manipulações algébricas e computacionais até o entendimento de que a distribuição normal assimétrica poderia apresentar bons resultados como uma distribuição proposta. Inicialmente, antes de considerar a transformação em (4.11), tentou-se utilizar uma distribuição candidata Uniforme para gerar valores de  $\phi$ , mas devido a pouca informação contida nessa distribuição, a cadeia obtida não convergiu para a distribuição de equilíbrio. Posteriormente, considerando a transformação  $H_2(\phi|\boldsymbol{\eta})$ , utilizou-se uma normal como função candidata. Os resultados foram mais satisfatórios, no entanto as

estimativas obtidas com a média e mediana dos valores amostrados estavam longe do valor real do parâmetro. Contudo, foi por meio dessa função geradora de candidatos que observamos a necessidade de utilizar uma distribuição proposta que acomodasse tanto os parâmetros de locação e escala, como também de assimetria. Esse fator determinou o uso da distribuição normal assimétrica, que revelou-se bastante apropriada no processo e amostragem de  $\phi$ .

Portanto, a ideia chave desse procedimento consiste em gerar valores de uma transformação do parâmetro  $\phi$  através da distribuição candidata normal assimétrica e, em seguida, obter tais valores na sua escala original.

É interessante destacar a relação de dependência entre os parâmetros  $\boldsymbol{\eta}$  e  $\phi$ , decorrentes da própria distribuição Poisson Generalizada. Essa dependência influencia nos resultados de convergência da cadeia, de tal forma, que em uma determinada iteração  $j$ , a amostra da distribuição a *posteriori*  $p(\boldsymbol{\eta}|\boldsymbol{\theta}, \Phi_0, \boldsymbol{\Phi}, \phi, \mathbf{y})$  afeta diretamente na simulação da amostra de  $p(\phi|\boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \boldsymbol{\Phi}, \mathbf{y})$ .

### 4.3 Predição

A obtenção de valores futuros da série a um passo à frente é muito importante no processo de validação e diagnóstico do modelo, visto que tais estimativas proporcionam o cálculo de medidas que permitem determinar se o modelo adotado representa uma boa descrição para os dados em estudo. No caso do MDPG, a distribuição preditiva a um passo a frente é dada por

$$\begin{aligned} p(y_t|D_{t-1}) &= \int \dots \int p(y_t, \phi, \eta_1, \dots, \eta_t, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_t, \Phi_0, \boldsymbol{\Phi}|D_{t-1})d\phi \dots d\boldsymbol{\Phi} \\ &= \int \dots \int p(y_t|\phi, \eta_1, \dots, \eta_t, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_t, \Phi_0, \boldsymbol{\Phi}, D_{t-1})p(\phi|\eta_1, \dots, \eta_t) \\ &\quad \times \left[ \prod_{i=1}^t p(\eta_i|\Phi_0, D_{t-1})p(\boldsymbol{\theta}_i|\boldsymbol{\Phi}, D_{t-1}) \right] p(\Phi_0|D_{t-1})p(\boldsymbol{\Phi}|D_{t-1})d\phi \dots d\boldsymbol{\Phi}. \end{aligned}$$

O cálculo dessa distribuição é bastante complexo e conseqüentemente o de seus momentos e quantis, impossibilitando a obtenção de intervalos de credibilidade para as predições com as ferramentas estatísticas que utilizamos ao longo desta dissertação. Outra limitação é a impossibilidade de calcular medidas de adequabilidade do ajuste baseadas na *densidade preditiva observada* (vide West e Harrison (1997), pág. 58).

Devido esses inconvenientes, calcularemos apenas a distribuição preditiva  $k$  passos à frente. Assim, admitindo-se que temos informações sobre a série até o tempo  $t$ , assim como as respectivas estimativas dos parâmetros no modelo,  $\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}, \hat{\Phi}_0$  e  $\hat{\Phi}$ , as previsões  $k$  passos à frente são obtidas de acordo com os procedimentos a seguir:

- a. Simular o vetor de variáveis latentes a  $k$  passos à frente,  $\hat{\boldsymbol{\theta}}_{(k)} = \boldsymbol{\theta}_{t+1}, \dots, \boldsymbol{\theta}_{t+k}$ , baseando-se no atual vetor de estados  $\boldsymbol{\theta}_t$ ;
- b. Obter valores futuros de  $\boldsymbol{\eta}$  com base no vetor de estados  $\boldsymbol{\theta}_{(k)}$ ;
- c. Amostrar as observações futuras a partir de uma função que gera variáveis com distribuição Poisson Generalizada.

Essa aproximação fornece estimativas da esperança  $E(Y_{t+k}|D_t)$  e variância  $Var(Y_{t+k}|D_t)$  da distribuição preditiva  $p(y_{t+k}|D_t)$ .

Para analisar a performance e qualidade do ajuste do MDPG à série temporal podem ser utilizadas as medidas de ajuste dadas na Seção 2.4.3 do Capítulo 2.

# Capítulo 5

## Simulações e Aplicações

Com o propósito de tornar mais compreensível e acessível a metodologia proposta no decorrer desta dissertação serão ajustados modelos dinâmicos para a distribuição Poisson Generalizada considerando séries temporais de contagem baseadas em cenários hipotéticos. Ilustraremos todos os métodos desenvolvidos, que vão desde a especificação de um MDPG até a previsão de observações futuras da série. Detalharemos também os passos necessários para a simulação de séries temporais conforme os princípios do modelo.

Destacamos que a análise de diagnóstico realizada nesta pesquisa tomou um foco secundário, haja vista que o principal objetivo deste estudo era a proposta de um modelo dinâmico flexível e versátil capaz de acomodar dados de contagem de qualquer natureza. Todavia buscando verificar a convergência da cadeia de Markov para a distribuição estacionária de interesse foram empregadas as principais técnicas descritas na literatura.

Por fim, como um meio de evidenciar a performance e o desenvolvimento da nossa metodologia, foram ajustados alguns modelos dinâmicos Poisson e comparados com os modelos dinâmicos para a distribuição Poisson Generalizada, através de análise gráfica e das principais medidas de qualidade de ajuste especificadas no Capítulo 2.

## 5.1 Especificação do MDPG

Ajustaremos um modelo dinâmico para a distribuição Poisson Generalizada no caso em que  $y_t$  é uma série temporal de contagem no tempo  $t = 1, \dots, N$ . Nesse modelo é assumido que as observações, a cada ponto no tempo, têm distribuição Poisson Generalizada e esta é especificada por dois parâmetros, o de forma  $\phi$ , que é desconhecido e estático, e o de locação,  $\lambda_t^*$ , que evolui temporalmente e estocasticamente. Recordemos que a equação das observações para o MDPG é dada por:

$$p(y_t | \eta_t, \phi) = \frac{e^{-\eta_t} (e^{-\eta_t} + \phi y_t)^{y_t - 1}}{y_t!} e^{-(e^{-\eta_t} + \phi y_t)}.$$

A aplicação de uma transformação de  $(0, 1)$  em  $\mathcal{R}$  no parâmetro dinâmico,  $\lambda_t^*$ , permite a criação de uma equação estrutural, que faz o papel de uma função de ligação, relacionando  $\lambda_t^*$  aos estados do sistema dinâmico de uma forma aditiva, isto é,

$$\eta_t = h(\lambda_t^*) = \mathbf{F}_t \boldsymbol{\theta}_t + v_t, \quad v_t \sim N(0, V). \quad (5.1)$$

Como tanto a equação estrutural, quanto a equação do sistema são assumidos seguirem modelos Gaussianos, o par de equações

$$\eta_t = h(\lambda_t^*) = \mathbf{F}_t \boldsymbol{\theta}_t + v_t, \quad v_t \sim N(0, V) \quad (5.2)$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}) \quad (5.3)$$

pode ser visto como um MDL Gaussiano tradicional. Desse modo,  $\eta_t$  tem o papel dos  $y_t$ 's (observações). Como dito anteriormente, essa estrutura é útil no processo de descrição do modelo dinâmico e na estimação dos parâmetros envolvidos.

Uma vez que  $\lambda_t^* \in (0, 1)$ , faz sentido considerar a transformação *logit* definida abaixo

$$\eta_t = \log \left( \frac{\lambda_t^*}{1 - \lambda_t^*} \right). \quad (5.4)$$

Consideramos um modelo polinomial de segunda-ordem com efeitos sazonais, isto é

$$\eta_t = \mu_t + \alpha_{t1} + v_t. \quad (5.5)$$

A equação em (5.5) define um modelo dinâmico por *superposição* composto por

- Efeitos polinomiais de segunda-ordem para  $\eta_t$  com respeito a  $\lambda_t^*$ :

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \omega_{\mu t}, \quad \omega_{\mu t} \sim N(0, \sigma_\mu^2) \quad (5.6a)$$

$$\beta_t = \beta_{t-1} + \omega_{\beta t}, \quad \omega_{\beta t} \sim N(0, \sigma_\beta^2) \quad (5.6b)$$

- Efeitos sazonais em forma livre:

$$\alpha_{tr} = \alpha_{t-1, r+1} + \omega_{tr}, \quad r = 1, \dots, s-1 \quad (5.7a)$$

$$\alpha_{t, s} = \alpha_{t-1, 1} + \omega_{t, s}. \quad (5.7b)$$

As mudanças ao longo do tempo na série temporal serão modeladas pelo parâmetro transformado  $\eta_t$ , através de  $\mu_t$ , representando o nível subjacente,  $\beta_t$  como um crescimento local,  $\alpha_{t1}$  representando o efeito sazonal com período de tamanho  $s$  no instante  $t$ , de modo que  $\alpha_{ts} = -\sum_{r=1}^{s-1} \alpha_{tr}$ . A quantidade  $v_t$  representa os desvios estocásticos da série  $\eta_t$  e os  $\omega$ 's são os desvios individuais associados aos efeitos do modelo.

De maneira mais genérica, podemos escrever esse modelo como

$$\eta_t = \mathbf{F}_t \boldsymbol{\theta}_t + v_t, \quad v_t \sim N(0, V), \quad (5.8)$$

onde o termo  $\mathbf{F}_t \boldsymbol{\theta}_t$  representa o preditor linear do parâmetro estrutural  $\eta_t$  para  $t = 1, \dots, N$ . O vetor de estados  $\boldsymbol{\theta}_t$  é então dado por

$$\boldsymbol{\theta}_t = (\mu_t, \beta_t, \alpha_{t1}, \dots, \alpha_{t, s-1}), \quad (5.9)$$

com  $2 + (s-1) = (s+1)$  parâmetros. A matriz de evolução  $\mathbf{G}_t = \mathbf{G}$  é considerada constante ao longo do tempo e tem dimensão  $(s+1) \times (s+1)$

$$\mathbf{G}_t = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & -1 & -1 & -1 & \cdots & -1 \end{bmatrix}.$$

A matriz de variância de evolução é descrita por  $\mathbf{W} = \text{bloco-diag}(\mathbf{W}_1, \mathbf{W}_2)$ , em que  $\mathbf{W}_1$  corresponde a uma matriz de variância  $(2 \times 2)$  relacionada aos parâmetros

$(\mu_t, \beta_t)'$  e  $\mathbf{W}_2$  é uma matriz de variância com dimensão  $(s - 1) \times (s - 1)$  associada aos parâmetros  $(\alpha_{t1}, \dots, \alpha_{t,s-1})'$ .

A matriz de delineamento  $\mathbf{F}_t = \mathbf{F}$  é assumida constante com dimensão  $(1 \times (s+1))$ , e é dada por

$$\mathbf{F}_t = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

## 5.2 Ajuste dos dados simulados

Com o propósito de ilustrar a performance dos modelos dinâmicos para a distribuição Poisson Generalizada, foram construídos dois cenários hipotéticos de séries temporais de contagem com base na estrutura do MDPG dada na Seção (5.1) deste capítulo.

Em cada cenário, uma série temporal com 72 observações foi simulada sob as mesmas condições, isto é, foi assumido uma estrutura comum para as séries de cada cenário, de modo a considerar a presença de crescimento localmente linear e comportamento cíclico trimestral, ou seja, com período sazonal  $s = 4$ . No entanto, características peculiares a cada série foram incorporadas com o interesse de avaliar a flexibilidade e a qualidade do ajuste do MDPG em diferentes panoramas. Nesse contexto, explicitaremos nesta seção uma estrutura geral para a simulação de dados conforme os componentes do modelo determinados acima. Em seguida, será apresentado os procedimentos necessários para a amostragem das distribuições condicionais completas *a posteriori* via o método MCMC. Por fim, exporemos o ajuste do modelo dinâmico para a distribuição Poisson Generalizada para ambos os cenários hipotéticos e faremos uma comparação com o ajuste obtido pelo modelo dinâmico Poisson.

## Simulação dos dados

Conforme os efeitos incorporados a ambos os conjunto de dados, temos que as matrizes de delineamento e transição de estados são dadas por

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{G} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & -1 & -1 \end{bmatrix}. \quad (5.10)$$

As variâncias  $V$  e  $\mathbf{W}$  são definidas comuns para os dois cenários, de maneira que,

$$V = 10^{-3} \quad \text{e} \quad \mathbf{W} = \text{bloco-diag}(\mathbf{W}_1, \mathbf{W}_2), \quad (5.11)$$

onde  $\mathbf{W}_1 = \text{diag}(10^{-5}, 10^{-6})$  e  $\mathbf{W}_2 = 10^{-6}\mathbf{I}_{(3 \times 3)}$ .

As variâncias  $V$  e  $\mathbf{W}$  foram construídas para serem suficientemente informativas a fim de prevenir mudanças estruturais bruscas devido à presença de valores discrepantes na amostra.

No processo de simulação, é fundamental descrever a relação entre  $\boldsymbol{\eta}$  e  $\phi$  que se estabelece por meio do limite inferior do parâmetro  $\phi$  que, por sua vez, depende de todos os valores do vetor  $\boldsymbol{\eta}$ . Outro aspecto relevante associado ao suporte de  $\phi$  é que à medida que este aproxima-se do seu limite superior, a amostra simulada apresenta mais inconformidades e muitos valores destoantes.

Para gerar dados com base nessa estrutura do MDPG, criamos uma rotina no *software* R denominada *sim.mdp*. Essa função tem como argumentos os momentos  $\mathbf{m}_0$  e  $\mathbf{C}_0$ , das matrizes de delineamento e evolução dos estados em (5.10) e das variâncias dada em (5.11). Outros argumentos da função *sim.mdp* são: o número de observações  $\mathbf{N}$  da série temporal, o valor máximo  $\mathbf{mx}$  que amostra poderá assumir e o valor de  $\mathbf{m}$  que é o maior inteiro positivo, maior ou igual a quatro, para que  $e^{-\eta_t} + \phi y_t > 0$  quando  $\phi < 0$ . Convém ressaltar a relevância de definir corretamente o valor de  $\mathbf{m}$ , pois este influencia bastante no comportamento da amostra gerada. Com a entrada dessas quantidades na rotina *sim.mdp* obtém-se a amostra simulada, o vetor de estados  $\boldsymbol{\theta}$ , o vetor de parâmetros estruturais  $\boldsymbol{\eta}$ , os valores de  $\lambda_t$  para  $t = 1, \dots, N$ , o valor de  $\phi$  e o próprio modelo que foi especificado anteriormente. O Código 5.6 no Apêndice D.4 apresenta os detalhes dessa função.

Uma vez que os dados simulados foram gerados, passamos a fase da estimação dos parâmetros de modo a verificar se conseguimos, através dos nossos procedimentos inferenciais, recuperar, ou seja, estimar adequadamente os valores verdadeiros dos parâmetros utilizados na geração dos dados sintéticos.

## Estimação e Predição

Os esforços na modelagem de séries temporais têm como principal finalidade a estimação e predição de observações futuras da série. Nos modelos dinâmicos para a distribuição Poisson Generalizada, o propósito é o mesmo e, para se alcançar esse fim, utilizamos a abordagem MCMC simultaneamente com outros desenvolvimentos como o algoritmo FFBS (Carter e Kohn (1994) e Frühwirth-Schnatter (1994)) que foram apresentados no Capítulo 4.

Ressalta-se que a escolha adequada dos valores iniciais dos parâmetros do modelo é um fator preponderante na convergência da cadeia de Markov. Para tanto, faz-se necessário especificar os valores iniciais  $\phi^{(0)}$ ,  $\lambda_t^{(0)}$ ,  $\eta_t^{(0)}$ ,  $\mathbf{m}_0$ ,  $\mathbf{C}_0$ ,  $V^{(0)}$  e  $\mathbf{W}^{(0)}$  no processo MCMC.

Os valores iniciais de  $\phi$  e  $\lambda_t$  são obtidos como a seguir:

Seja

$$\phi^{(0)} = 1 - \sqrt{\frac{\bar{\mathbf{y}}}{Var(\mathbf{y})}}, \quad (5.12)$$

que é o estimador pelo método dos momentos de  $\phi$  com base em uma amostra aleatória.

Na especificação de  $\lambda_t^{(0)}$ , utilizou-se propriedades de média e variância de  $(y_t|\phi, \eta_t)$ .

Tem-se que

$$E(y_t|\phi, \eta_t) = \frac{\exp(-\eta_t)}{1 - \phi} \quad \text{e} \quad Var(y_t|\phi, \eta_t) = \frac{\exp(-\eta_t)}{(1 - \phi)^3}.$$

Como

$$E(y_t|\phi, \eta_t) + Var(y_t|\phi, \eta_t) = \lambda_t \left( \frac{1}{1 - \phi} + \frac{1}{(1 - \phi)^3} \right), \quad (5.13)$$

igualando-se o lado esquerdo de (5.13) a  $y_t + Var(y)$  e, substituindo a estimativa de  $\phi$ , obtida em (5.12), no lado direito de (5.13), obtém-se

$$\lambda_t^{(0)} = \frac{y_t + Var(\mathbf{y})}{f(\phi^{(0)})},$$

com  $f(\phi^{(0)}) = \frac{1}{1-\phi^{(0)}} + \frac{1}{(1-\phi^{(0)})^3}$ .

Conseqüentemente, através de  $\lambda_t^{(0)}$  podemos obter as estimativas  $\eta_t^{(0)}$  por meio da transformação biunívoca

$$\eta_t = -\log(\lambda_t).$$

A especificação dos valores iniciais  $\mathbf{m}_0, \mathbf{C}_0, V^{(0)}$  e  $\mathbf{W}^{(0)}$  depende muito do conhecimento *a priori* que o analista tem dos dados, de modo que não existe uma fórmula para a definição dessas quantidades. Em geral, quando não se tem nenhuma familiaridade com os dados, muitas tentativas são realizadas até se obter valores iniciais que levem a bons resultados.

Cabe enfatizar que a matriz de variância do sistema é amostrada por blocos, ou seja, amostramos cada uma das sub-matrizes  $\mathbf{W}_i$ ,  $i = 1$  e  $2$ , de modo a considerar os efeitos incorporados ao modelo e assegurar que esses componentes sejam tratados como independentes.

Na primeira amostra MCMC de  $\boldsymbol{\eta}$ , a distribuição  $(\boldsymbol{\eta}|\boldsymbol{\theta}, \Phi_0, \boldsymbol{\Phi}, \phi, \mathbf{y})$  será amostrada a partir da distribuição candidata normal cujos parâmetros,  $\hat{\eta}_t$  e  $\hat{\Sigma}_t$ , dados em (4.8) e (4.9), serão baseados nas estimativas iniciais de  $\eta_t^{(0)}, \phi^{(0)}$  e  $V^{(0)}$ .

De modo análogo, para amostrar a distribuição  $p(\phi|\boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \boldsymbol{\Phi}, \mathbf{y})$  é necessário simular amostras da transformação  $H_2(\phi|\boldsymbol{\eta})$ , por meio de um função proposta normal assimétrica com parâmetros de locação  $\xi$ , de escala  $\omega$  e de forma  $\delta$ . Na primeira iteração, o parâmetro de posição  $\xi$  é derivado dos valores iniciais  $\eta_t^{(0)}$  e  $\phi^{(0)}$ , enquanto os demais são definidos a partir da compreensão prévia que o pesquisador tem do sistema.

A previsão de observações futuras da série temporal é realizada após a conclusão do método MCMC, seguindo os passos dados na Seção 4.3 do Capítulo 4.

Na tentativa de tornar mais claro todo o processo de estimação, construímos uma rotina no *software* R que desenvolve todo o processo de estimação do MDPG a partir da especificação das quantidades acima mencionadas. O Código 5.7 no Apêndice D.4 mostra as etapas para do procedimento de estimação via MCMC.

## Análise dos dados simulados

Nesta seção apresentamos os resultados provenientes do ajuste do MDPG para a série temporal desenhada em cada um de dois cenários. Paralelamente, essas séries foram ajustadas conforme um modelo dinâmico Poisson, como meio de comparar e destacar a performance do MDPG. A estrutura do modelo dinâmico Poisson é idêntica a do modelo dinâmico para a distribuição Poisson Generalizada, a diferença é que no primeiro, o parâmetro  $\phi$  é igual a zero e não é preciso especificar as quantidades  $\mathbf{k}$  e  $\mathbf{m}$  da função *sim.mdpq*.

Cabe ressaltar que no processo de estimação para ambos os modelos, utilizando o método MCMC, foram geradas 200.000 cadeias de Markov, donde 50.000 destas foram descartadas formando uma amostra de aquecimento (*burn-in*). Como é da natureza do próprio procedimento MCMC gerar amostras correlacionadas, um meio de reduzir a autocorrelação foi considerar apenas 1 em cada 30 amostras MCMC geradas, dessa forma, totalizando 5.000 amostras observadas.

### Cenário 1

A série temporal de contagem apresentada na Figura 5.1 foi simulada através da função *sim.mdpq* considerando as especificações do MDPG dadas anteriormente e admitindo  $\mathbf{mx} = 15$  e  $m = 10$ .

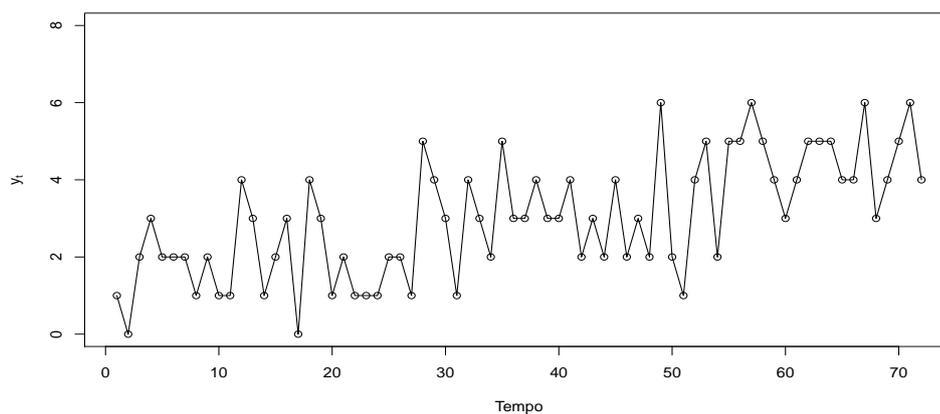


Figura 5.1: Série temporal com dados de contagem e  $m = 10$ .

A determinação do valor de  $\mathbf{m}$  é muito importante para a geração de séries tempo-

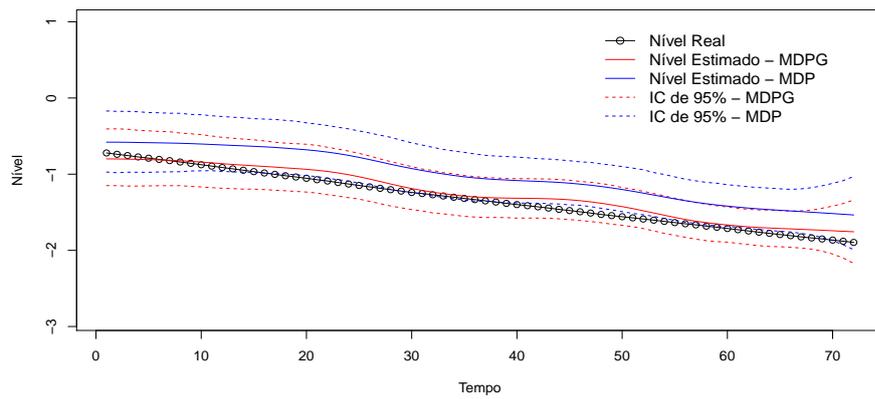
rais sob a estrutura do MDPG, visto que este é um fator que exerce bastante influência sob o parâmetro de precisão  $\phi$ . Nesse cenário, considerando todas as configurações já delineadas previamente, a série simulada retornou  $\phi = -0.314$ , indicando que a variância de cada ponto no tempo é menor que sua média, divergindo do modelo Poisson.

A Figura 5.2 mostra os valores estimados dos efeitos (nível, tendência e sazonalidade) sob o MDPG e o MDP, bem como os valores reais. Como se pode observar, de modo geral, o processo inferencial utilizado para a estimação dos parâmetros do MDPG foi capaz de estimar corretamente os valores verdadeiros. Por outro lado, observa-se que o nível da série é superestimado pelo MDP. Os efeitos de crescimento e sazonalidade são estimados igualmente bem por ambos os modelos. No entanto, o MDPG conduz a intervalos de credibilidade mais precisos.

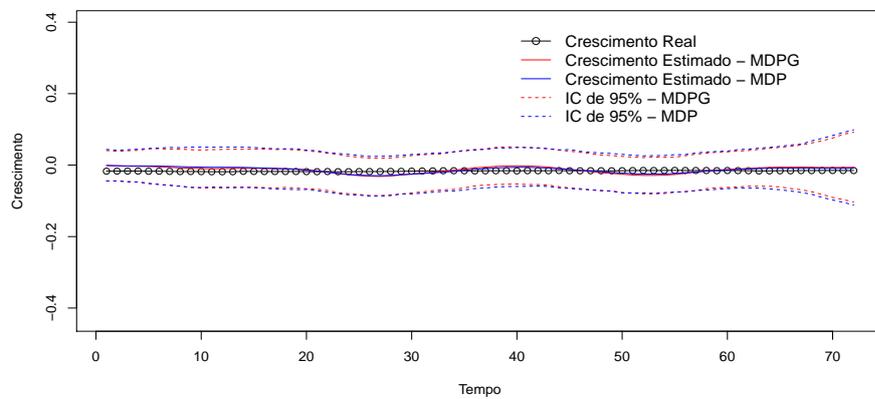
Observa-se, também, que na estimação da componente sazonal, à medida que o tempo evolui, o modelo parece incorporar um padrão cíclico mais definido tentando adaptar-se à estrutura dos dados. Repara-se também que em torno do intervalo de tempo  $t \in \{40, \dots, 60\}$ , o efeito sazonal estimado apresenta uma variabilidade maior do que o real. Esse fator pode ser oriundo da especificação inicial da sub-matriz de variância associada aos componentes cíclicos do modelo. Uma alternativa para tentar corrigir esse problema seria reduzir os valores iniciais dessa matriz. No entanto, isso requer também mudanças nos demais valores iniciais, visto que, como já mencionado, a definição inicial dos parâmetros do modelo deve ser feita cautelosamente e de maneira harmoniosa, uma vez que estes exercem forte influência na convergência da cadeia de Markov do método MCMC.

A Figura 5.3 apresenta as estimativas obtidas para  $\eta_t$  e  $\lambda_t$ . Observa-se que, semelhante ao que foi visualizado na análise gráfica dos efeitos sazonais, as estimativas  $\hat{\eta}_t$  e, conseqüentemente,  $\hat{\lambda}_t$ , também incorporam uma variabilidade maior entre o período  $t \in \{40, \dots, 60\}$ , mas nada que inviabilize o emprego dessa metodologia. Na realidade, pode-se atestar que os valores estimados estão bem próximos dos valores reais, evidenciando um bom ajuste do MDPG à série histórica em estudo. A taxa de convergência associada a este processo de amostragem via algoritmo de Metropolis-Hastings foi de 55,78%.

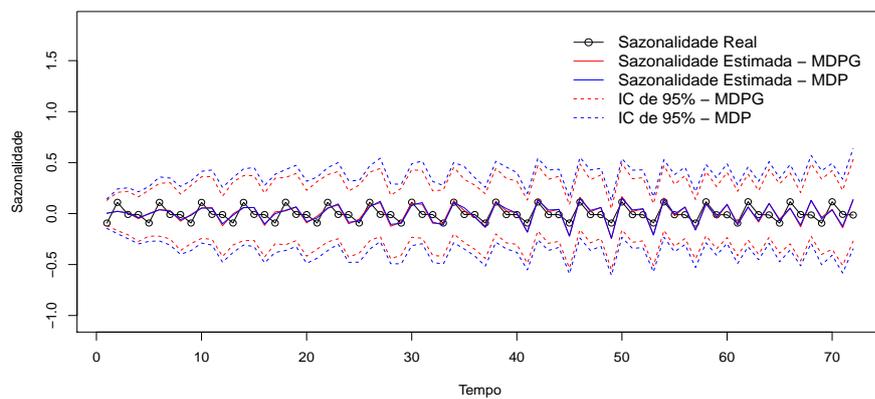
Mais uma vez é possível destacar a superioridade do ajuste MDPG em relação ao



(a)



(b)



(c)

Figura 5.2: (a) Nível Real *versus* Nível Estimado; (b) Crescimento Linear Real *versus* Crescimento Linear Estimado; (c) Efeito Sazonal Real *versus* Efeito Sazonal Estimado.

MDP. Novamente, este último modelo parece superestimar os valores dos parâmetros  $\eta_t$  e, como consequência, subestima os valores de  $\lambda_t$ .

O processo de amostragem a partir da distribuição  $(\phi|\boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \boldsymbol{\Phi}, \mathbf{y})$  é bastante complexo, em razão de  $\phi$  ser um parâmetro estático, mas cujo suporte depende do parâmetro  $\eta_t$  que, por sua vez, é dinâmico no tempo. Essa dependência de  $\lambda_t$  no suporte de  $\phi$ , mesmo nos casos de dados independentes e identicamente distribuídos gera muitos problemas, do processo de simulação à metodologia de estimação dos parâmetros. Para dados correlacionados ao longo do tempo, a situação torna-se ainda mais complicada. Portanto, é de suma relevância a especificação adequada dos valores iniciais para os parâmetros da distribuição candidata normal assimétrica no algoritmo de Metropolis-Hastings, assim como, também, é fundamental a descrição coerente das variâncias  $V$  e  $\mathbf{W}$  para uma boa e rápida convergência da cadeia de Markov. A Tabela 5.1 mostra as estatísticas resumo das amostras MCMC de  $\phi$ .

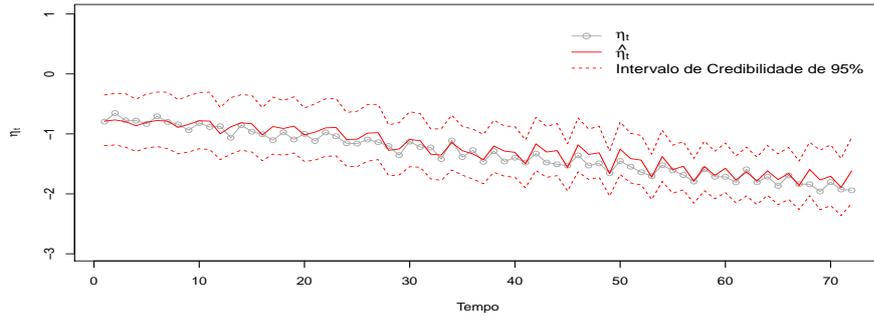
Tabela 5.1: Estatísticas resumo das amostras MCMC de  $\phi$

Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo
-0,433	-0,304	-0,271	-0,272	-0,239	-0,108

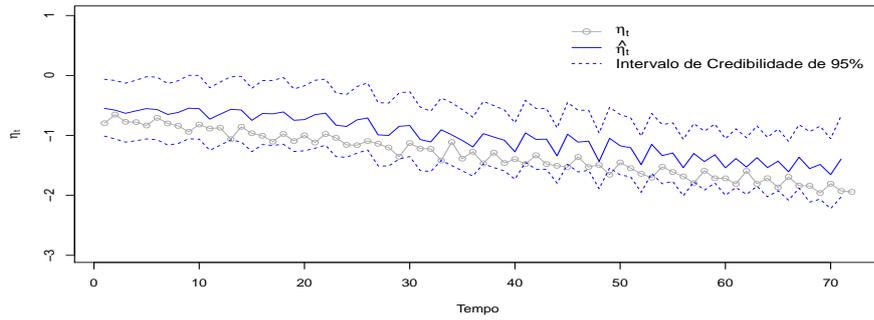
A média e a mediana *a posteriori* de  $\phi$  apresentam valores bem próximos do parâmetro real  $\phi = -0,314$ , indicando que a simulação de uma transformada de  $\phi$  através de uma distribuição candidata normal assimétrica, proporciona resultados confiáveis quando as estimativas iniciais dos hiperparâmetros são bem definidas. Tomou-se como estimativa do parâmetro  $\phi$  a média da amostra gerada,  $\hat{\phi} = -0,272$ . A taxa de aceitação dos valores gerados de  $\phi$  foi de 22,14%, representando um percentual razoável.

A Figura 5.4 ilustra o comportamento das amostras MCMC de  $\phi$ . Nos gráficos em (a) e (b), nota-se uma leve assimetria negativa na amostra derivada da distribuição proposta para gerar  $\phi$ , enquanto que nos gráficos (c) e (d) repara-se que a cadeia de Markov converge para uma distribuição estacionária de  $\phi$ .

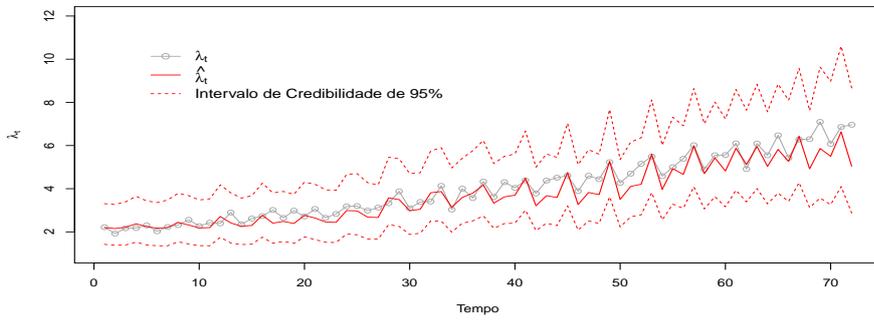
Uma vez que  $\phi$  foi estimado, pode-se obter a média estimada do modelo MDPG e, em seguida, comparar os resultados com a média estimada do modelo dinâmico Poisson, que é o próprio parâmetro  $\lambda_t$ . A Figura 5.5 ilustra as médias estimadas de ambos os modelos.



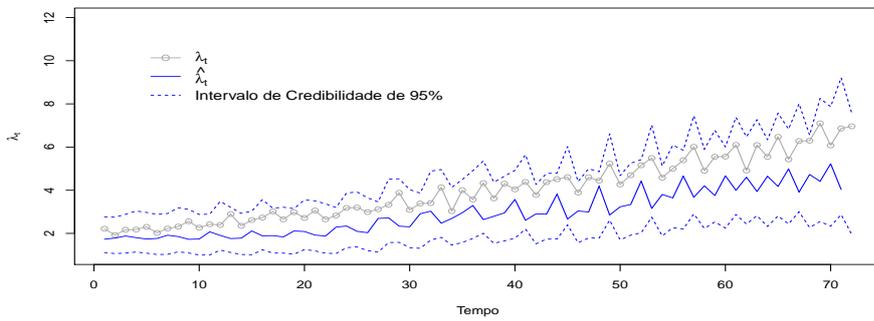
(a)



(b)



(c)



(d)

Figura 5.3: (a)  $\eta_t$  versus  $\hat{\eta}_t$  - MDPG; (b)  $\eta_t$  versus  $\hat{\eta}_t$  - MDP; (c)  $\lambda_t$  versus  $\hat{\lambda}_t$  - MDPG; (d)  $\lambda_t$  versus  $\hat{\lambda}_t$  - MDP.

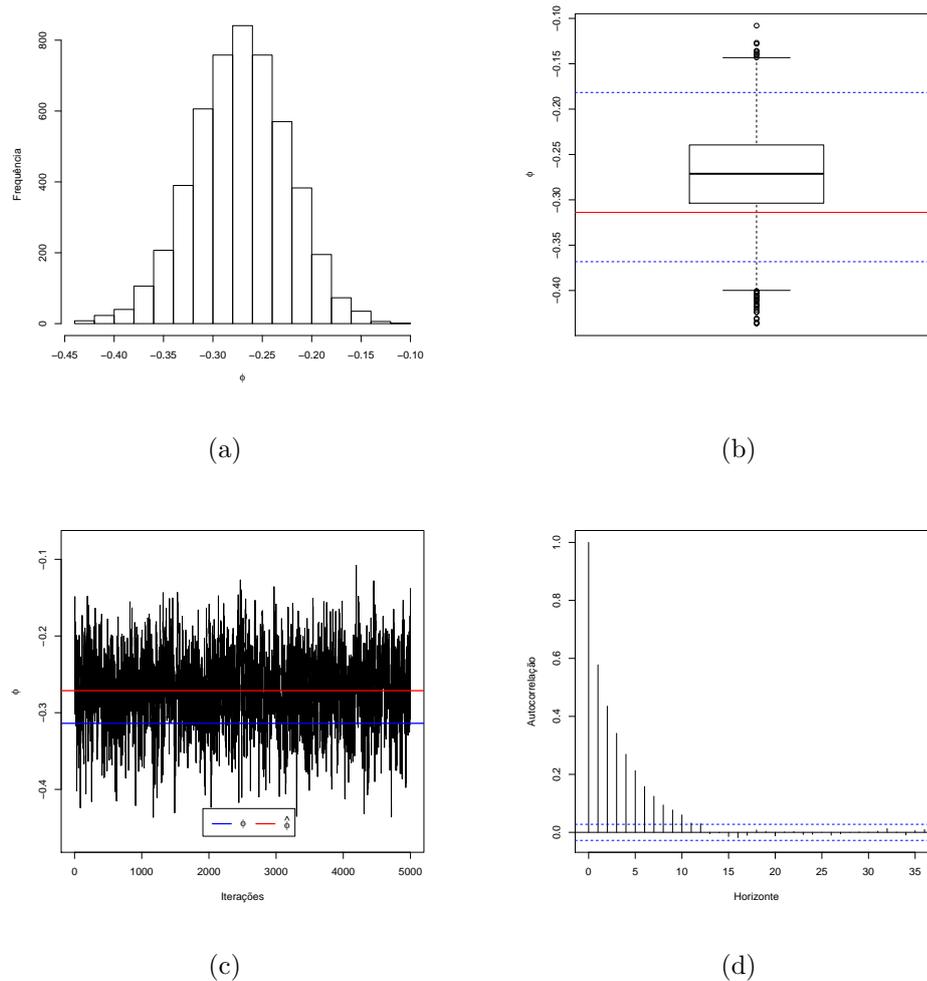


Figura 5.4: Gráficos relativos às amostras MCMC do parâmetro  $\phi$ : (a) Histograma; (b) Box-Plot, onde a linha horizontal representa o valor verdadeiro de  $\phi$  e as demais os limites do I.C. de 95%; (c) Amostras de  $\phi$  ao longo das iterações; (d) Autocorrelação da amostra de  $\phi$ .

Novamente, nota-se que o ajuste dos modelos MDPG e MDP são aproximadamente iguais e que a amplitude de ambos os intervalos de credibilidade parecem aumentar à medida que o tempo evolui. Esse alargamento deve ser decorrente da introdução de um parâmetro estrutural no modelo para a utilização de propriedades do MDL normal que, conseqüentemente, leva a uma perda de informação.

As amostras das variância  $V$  e  $\mathbf{W}$  obtidas pelo MCMC convergiram de modo razoavelmente rápido para uma distribuição de equilíbrio como mostra a Figura 5.6.

Para avaliar a qualidade do ajuste e auxiliar na seleção do modelo mais adequado à série, calculamos o DIC que forneceu os valores 248,77 e 264,32 para o MDPG e

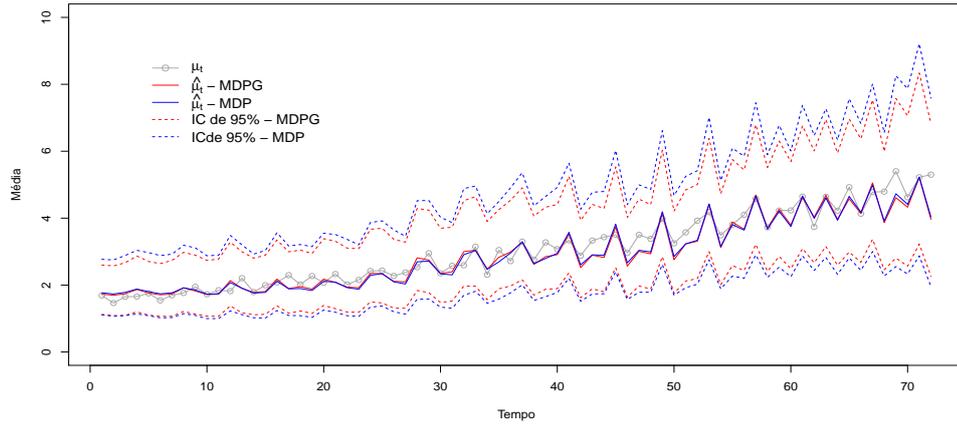


Figura 5.5: Comparação da média real com média estimada via MDPG e MDP.

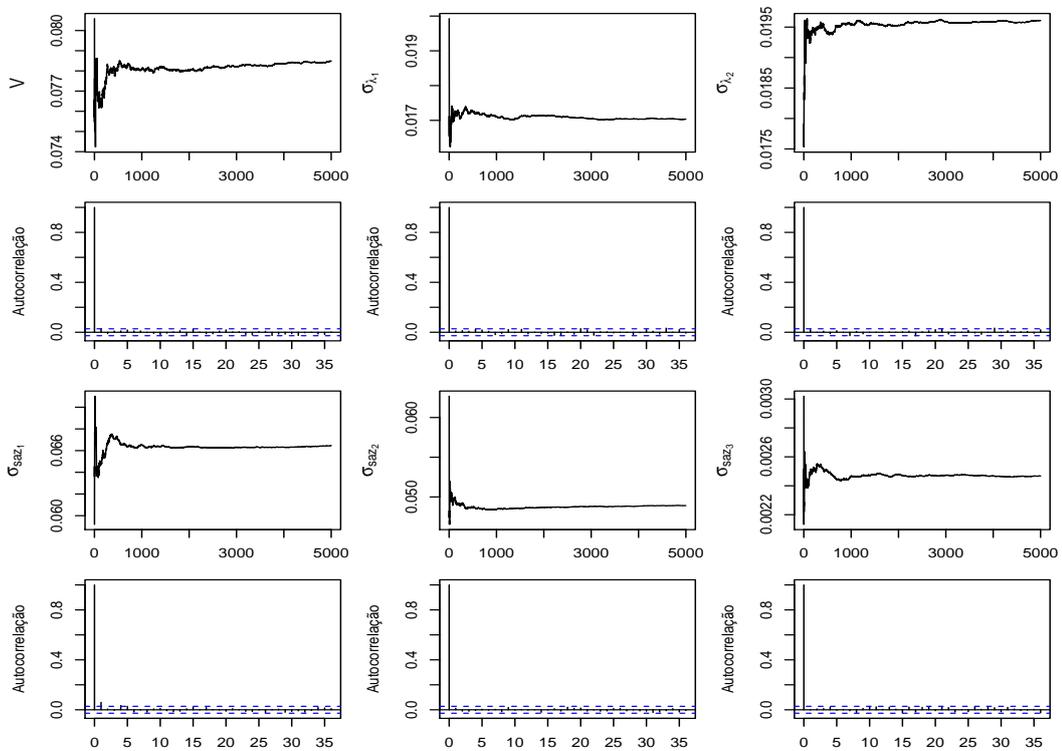


Figura 5.6: Gráficos das amostras MCMC e autocorrelações para as variância  $V$ ,  $W_1$  e  $W_2$ .

MDP, respectivamente, indicando que o modelo dinâmico para a distribuição Poisson Generalizada apresenta-se mais adequado para o ajuste dos dados.

Outro meio de analisar a adequabilidade dos modelos é realizando previsões de observações futuras considerando apenas parte da série temporal. Para tanto, toma-

mos como conhecidas as primeiras 66 observações e, com base nessa série reduzida, realizamos as previsões a  $k = 6$  passos à frente. A Figura 5.7 ilustra tais resultados e evidenciam que a performance do MDPG na previsão do comportamento futuro da série, parece ser melhor, no sentido de acompanhar com mais nitidez o desenvolvimento dos dados, do que o MDP, que, mais uma vez, tende a superestimar os valores das observações.

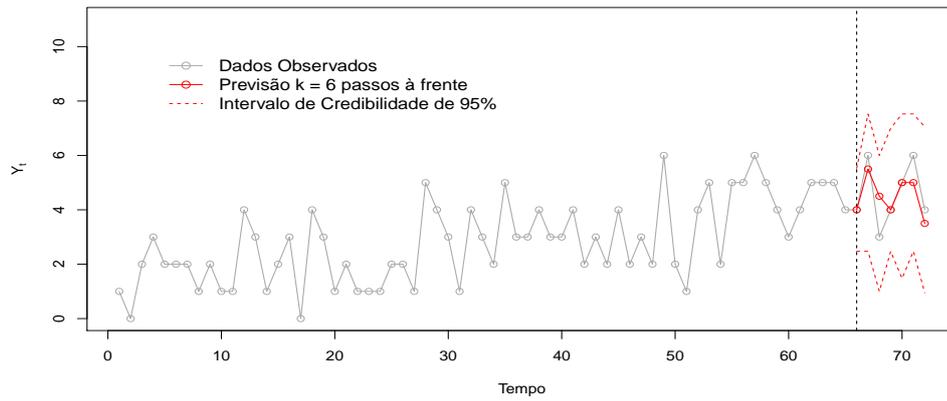
Na Tabela 5.2 as medidas obtidas reforçam nossas suspeitas sobre a superioridade do MDPG em relação ao MDP, uma vez que todas as métricas calculadas para o MDPG apresentam valores menores que as do MDP.

Tabela 5.2: Medidas de qualidade de ajuste considerando previsões  $k = 6$  passos à frente do MDPG e MDP

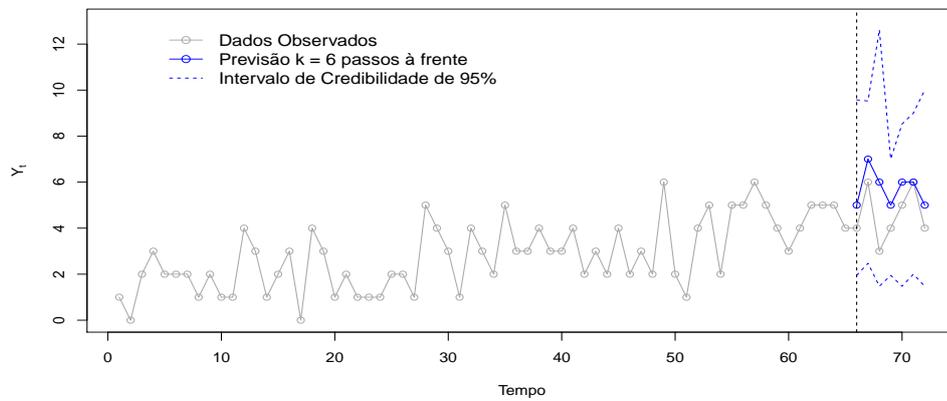
Modelo	Métricas					
	MSE	MAD	MAPE	SMAPE	MedAPE	MASE
MDPG	0,536	0,500	1,215	2,865	0,083	0,417
MDP	2,000	1,143	2,939	5,961	0,250	6.667

Assim, de acordo com a análise realizada até o momento, para o cenário 1, podemos afirmar que o MDPG é um modelo que proporciona resultados satisfatórios para tratar séries temporais de contagem, no que diz respeito à estimação dos parâmetros e previsão de observações futuras da série.

Convém destacar que como a série simulada contém poucas observações, não é apropriado fazer previsões a longo prazo, além disso, a própria especificação dos modelos dinâmicos em geral admitem apenas previsões a curto e médio prazo.



(a)



(b)

Figura 5.7: (a) Previsão  $k = 6$  passos a frente com MDPG; (b) Previsão  $k = 6$  passos a frente com MDP.

## Cenário 2

A série temporal na Figura 5.8(a) foi gerada a partir da rotina *sim.mdp* de acordo com as definições do MDPG descrito nas seções anteriores deste capítulo. No processo de simulação da série foi considerado  $\mathbf{m}x = 30$  e  $\mathbf{m} = 20$ .

Neste cenário esboçamos o estudo, de modo mais sintetizado, dessa série temporal que forneceu  $\phi = -0,095$ . A finalidade é demonstrar que quando o parâmetro  $\phi$  é próximo de zero, a modelagem via MDPG é aproximadamente equivalente a do MDP.

A Figura 5.8 também mostra as estimativas dos estados latentes da série individualmente. Através desta pode-se perceber que, apesar de apresentar uma diferença muito pequena, o nível estimado pelo MDPG aproxima-se mais do verdadeiro nível

da série em relação ao nível estimado via MDP. Isso acontece devido ao fato de o MDPG incorporar a informação sobre o parâmetro  $\phi$ , enquanto o MDP não desfruta dessa particularidade. Se no ajuste do modelo dinâmico para a distribuição Poisson Generalizada fosse considerado  $\phi = 0$ , as estimativas seriam exatamente idênticas às do MDP.

As estimativas do parâmetro estrutural  $\eta_t$ , bem como as de  $\lambda_t$ , de ambos os modelos, são ilustradas na Figura 5.9 e parecem acomodar bem o desenvolvimento dos parâmetros dinâmicos. Observa-se, também, que apesar destas mostrarem valores próximos, as estimativas procedentes do MDPG tendem a seguir, com mais precisão, as flutuações desses parâmetros, enquanto que os valores estimados via MDP apresentam um pequeno desvio em relação aos valores reais.

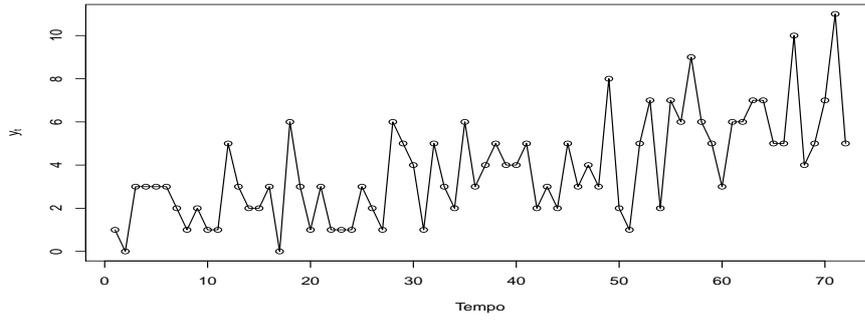
Os modelos traduzem bem o comportamento médio da série. Nota-se a partir da Figura 5.9(c) que à medida que tempo evolui, a variabilidade das estimativas de ambos os modelos aumentam. Como colocado anteriormente, esse acréscimo na variabilidade pode ser devido aos valores iniciais atribuídos às variâncias  $V$  e  $W$ .

Na estimação do parâmetro  $\phi = -0,095$ , consideramos como estimativa a mediana *a posteriori*, que forneceu  $\hat{\phi} = -0.091$ . A Figura 5.10, juntamente com a taxa de aceitação de 62,2%, produzem evidências de que a cadeia de Markov associada à amostragem de  $\phi$  converge para uma distribuição estacionária.

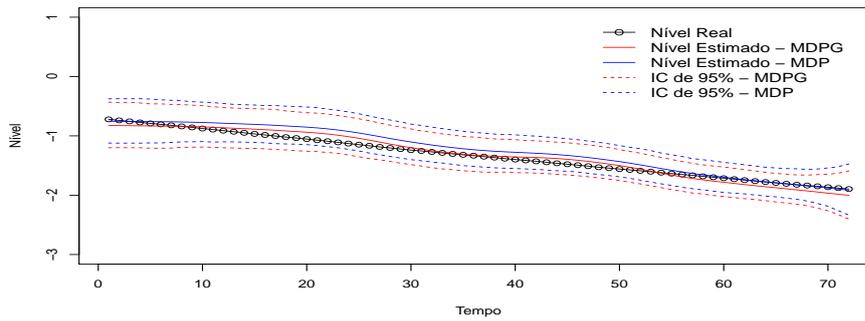
Em geral, as estimativas obtidas pelos dois modelos ajustados, além de serem próximas, apresentam boas perspectivas dos componentes da série temporal.

Como meio de comparar a performance de ambos os modelos ajustados, calculamos o DIC que forneceu os valores 291,39 e 294,44 para o MDPG e MDP, respectivamente. Apesar do MDPG apresentar DIC menor do que o do MDP, a diferença entre essas medidas é insignificante, indicando que os resultados obtidos por ambos os modelos são aproximadamente equivalentes.

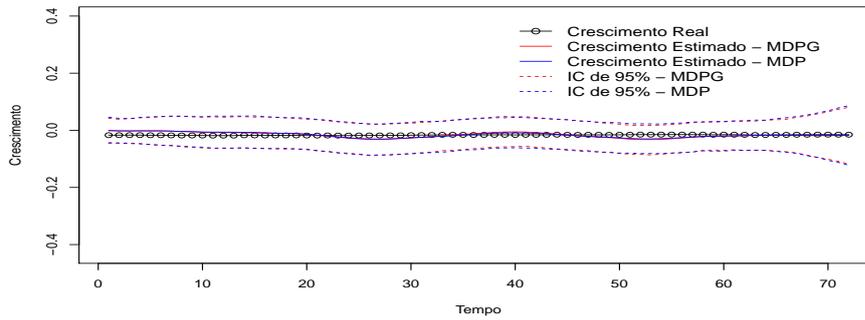
Por fim, como nosso intuito neste cenário é demonstrar que o ajuste do MDPG aproxima-se do MDP quando  $\phi$  tende a zero, apresentamos na Figura 5.11 as previsões a  $k = 6$  passos a frente, como meio de validar o MDPG como um modelo plausível para o estudo de séries temporais com dados discretos de natureza diversa.



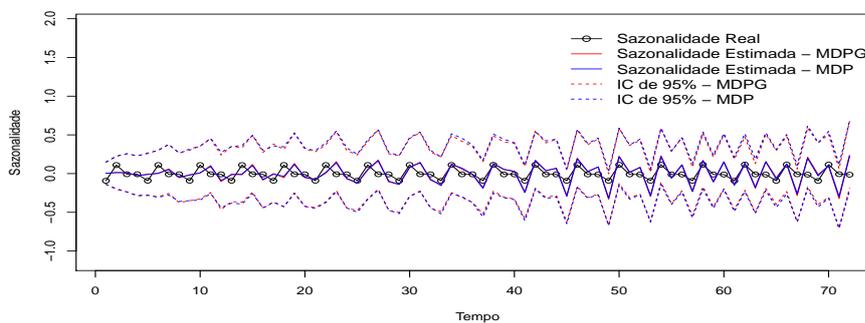
(a)



(b)

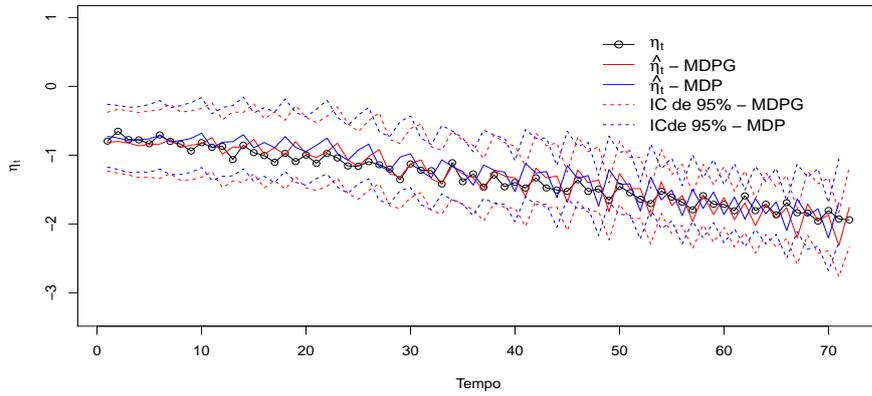


(c)

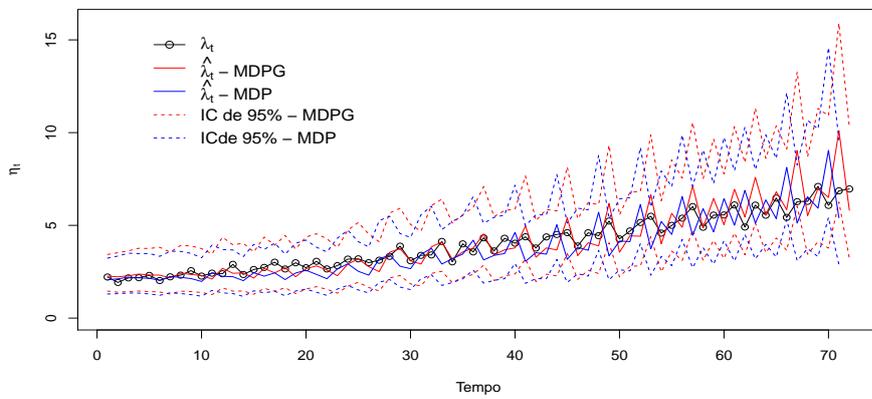


(d)

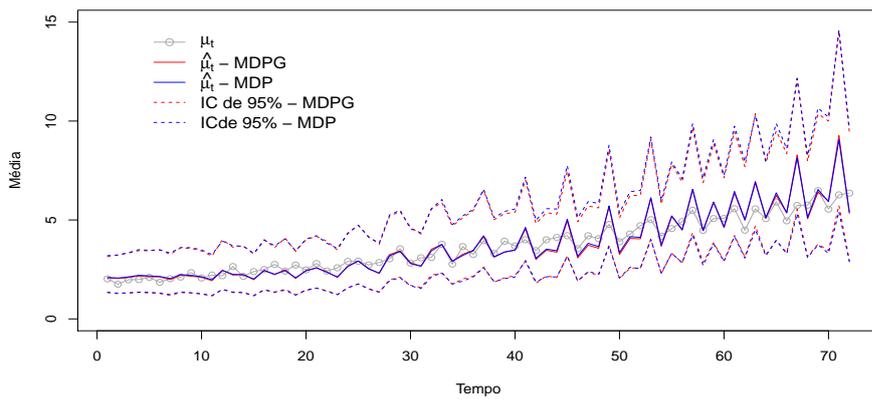
Figura 5.8: (a) Série temporal simulada do MDPG e  $m = 20$ ; (b) Nível real *versus* nível estimado; (c) Crescimento linear real *versus* Crescimento linear estimado; (d) Efeito sazonal real *versus* Efeito sazonal estimado.



(a)

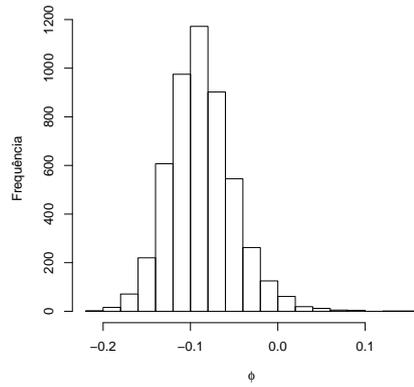


(b)

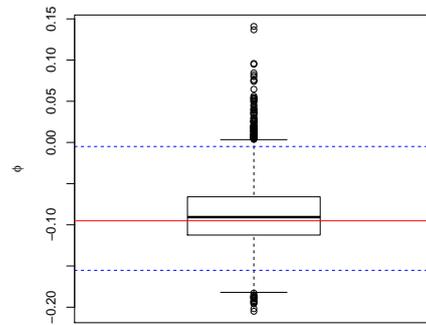


(c)

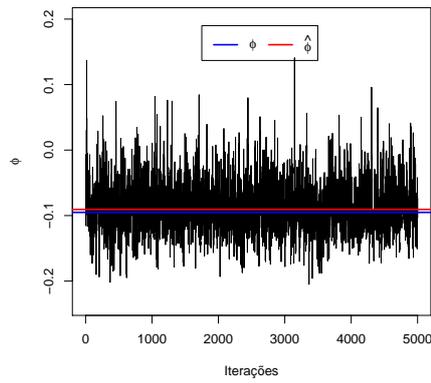
Figura 5.9: (a) Estimativas de  $\eta_t$  via MDPG e MDP; (b) Estimativas de  $\lambda_t$  via MDPG e MDP; (c) Estimativas de  $\mu_t$  via MDPG e MDP.



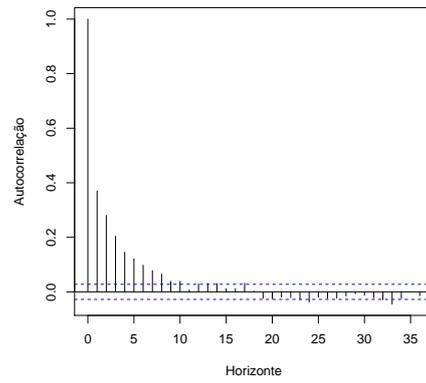
(a)



(b)

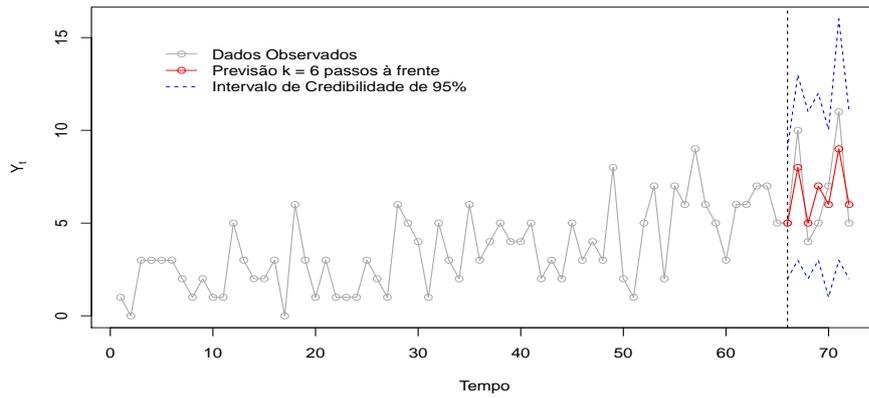


(c)

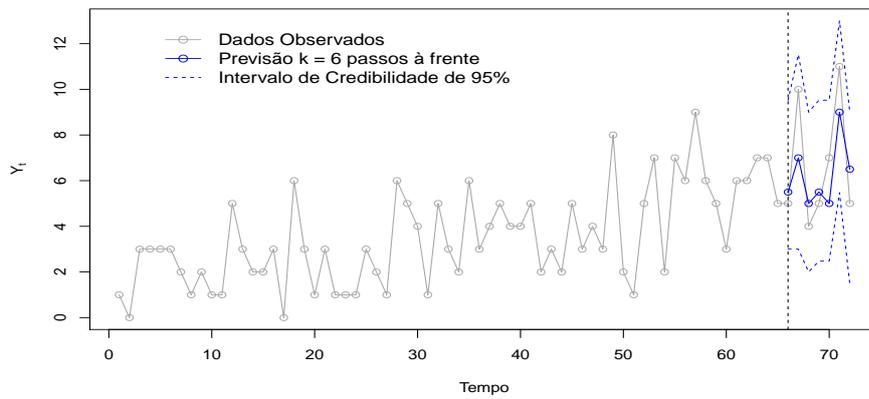


(d)

Figura 5.10: Gráficos relativos às amostras MCMC do parâmetro  $\phi$ : (a) Histograma; (b) Box-Plot, onde a linha horizontal representa o valor verdadeiro de  $\phi$  e as demais os limites do I.C. de 95%; (c) Amostras de  $\phi$  ao longo das iterações; (d) Autocorrelação da amostra de  $\phi$ .



(a)



(b)

Figura 5.11: (a) Previsão a  $k = 6$  passo a frente com o MDPG; (b) Previsão a  $k = 6$  passos a frente com o MDP.

# Conclusão

A construção de um modelo dinâmico para a distribuição Poisson Generalizada foi uma tarefa bastante árdua tendo início no estudo da distribuição Poisson Generalizada, explorando com detalhes a classe dos modelos dinâmicos lineares normais e evoluindo até o desenvolvimento do método MCMC para a realização de inferências sobre os parâmetros de interesse.

A escolha da DPG para modelar as observações da série temporal, a cada ponto no tempo, foi motivada pela ampla aplicabilidade desta na descrição de dados de natureza discreta. A partir de simulações de variáveis aleatórias DPG e da estimação dos parâmetros intrínsecos constatamos algumas restrições em relação ao espaço paramétrico de  $\phi$ , principalmente no que diz respeito a sua relação com  $\lambda$  que, por sua vez, influencia em todo o ajuste do modelo.

No intuito de apresentar resultados satisfatórios quanto à estimação de características dos dados e realizar a previsão de perspectivas futuras da série, muitas alternativas foram abordadas até que se obtivesse uma viável.

Um fator preponderante para o desenvolvimento do MDPG foi perceber a importante relação de dependência entre os parâmetros  $\lambda_t$  e  $\phi$  e também o quão relevante é a especificação do valor de  $\mathbf{m}$  na DPG, desempenhando um papel de limitador para que a probabilidade da DPG não tome valores negativos quando  $\phi$  é negativo. Uma vez que o espaço paramétrico de  $\phi$  está restrito à evolução de  $\lambda_t$  foi necessário uma reparametrização no sentido de facilitar o processo de amostragem, considerando  $\lambda_t$  dinâmico no tempo e caracterizando o parâmetro  $\phi$  como desconhecido e estático, tomando todo o espaço de  $\lambda_t$ .

Nessas configurações, o método MCMC para amostragem das distribuições à *posteriori* do modelo teve como componente chave o uso do algoritmo de Metropolis-

Hastings dentro do amostrador de Gibbs para gerar amostras do parâmetro estrutural  $\eta_t = \text{logit}(\lambda_t)$  que, posteriormente, serviu na atualização de  $\phi$ , também amostrado via MH.

Outro aspecto digno de nota é a especificação coerente dos parâmetros para a distribuição candidata de  $p(\phi|\boldsymbol{\eta}, \boldsymbol{\theta}, \Phi_0, \boldsymbol{\Phi}, \mathbf{y})$  que afeta na convergência de todas as amostras geradas para a distribuição de interesse.

Considerando o fato de que não foi possível a aplicação da metodologia desenvolvida no estudo em dados reais, o ajuste do modelo dinâmico para a distribuição Poisson Generalizada foi realizado em duas séries temporais simuladas. O primeiro cenário simulado idealizou uma série de dados de contagem composta por 72 observações, teoricamente parametrizada pelos limitantes  $f_{\mathbf{mx}} = 15$  e  $m = 10$ . Neste estudo de caso, composto por 200 mil iterações, *burn-in* de 50 mil amostras e um *gap* com defasagem 30, as estimativas dos estados latentes e dos parâmetros da distribuição obtidos através da metodologia desenvolvida mostraram-se bastante satisfatórios principalmente quando comparados com o modelo dinâmico Poisson. As diferenças entre medidas de qualidade de ajuste para o MDPG e MDP evidenciaram a superioridade de ajuste do MDPG em relação ao MDP para os dados simulados.

Para o segundo estudo de caso, objetivou-se destacar a equivalência entre as abordagens que utilizam o MDPG e MDP no caso em que o parâmetro  $\phi$  assume valores próximos a zero. De fato, foi constatado que as estimativas dos parâmetros em ambos os ajustes são bem próximas. No entanto, verifica-se ainda que o parâmetro  $\phi$  do MDPG, ainda que pequeno, é capaz de melhor captar o desenvolvimento da série, bem como o de suas características.

No que concerne a análise de convergência das cadeias do método MCMC, optou-se apenas por uma abordagem descritiva que apontou o bom ajuste do MDPG. No entanto, observou-se a necessidade da aplicação de técnicas confirmatórias na apuração do alcance da convergência.

Muitos foram os esforços computacionais realizados neste trabalho para poder respeitar todas as limitações e restrições do nosso modelo. Muitas funções foram implementadas a fim de facilitar todo o processo de construção do MDPG e estão disponíveis para consulta no Apêndice D.

As pretensões de desenvolvimentos futuros baseiam-se principalmente na expansão

do MDPG para modelar um conjunto de séries temporais de dados de contagem que desfrutam de determinadas características considerando ambos os parâmetros da DPG variando com o tempo e hierarquicamente através das séries.

# Apêndice

## A - Expansão de Lagrange

Considere a transformação inversa de Lagrange  $u = z/g(z)$  que fornece o valor de  $z$  como uma série de potência em  $u$  e  $g(z)$  uma função analítica de  $z$ . Seja  $f(z)$  outra função analítica de  $z$  que assim como  $g(z)$  é sucessivamente diferenciável com respeito a  $z$  no intervalo  $(-1, 1)$ . Então, as expansões em séries de potências são dada por:

$$z = \sum_{k=1}^{\infty} \frac{u^k}{k!} \{D^{k-1}(g(z))^k\}_{z=0}, \quad (5.14)$$

$$f(z) = \sum_{k=0}^{\infty} a_k u^k, \quad (5.15)$$

onde  $a_0 = f(0)$  e

$$a_k = \frac{1}{k!} D^{k-1} \{(g(z))^k f'(z)\}_{z=0}, \quad (5.16)$$

e

$$\frac{f(z)}{1 - z g'(z)/g(z)} = \sum_{k=0}^{\infty} b_k u^k, \quad (5.17)$$

em que  $b_0 = f(0)$  e

$$b_k = \frac{1}{k!} D^k \{(g(z))^k f(z)\}_{z=0}. \quad (5.18)$$

Observe que as expansões de Lagrange em (5.15) e (5.17) não são independentes uma vez que uma pode ser derivada da outra. Se substituirmos  $f(z)$  em (5.17) por

$$f_1(z) = (1 - z g'(z)/g(z)) f(z),$$

simplificamos (5.17) na expansão de Lagrange em (5.15). De maneira análoga, substituindo  $f(z)$  em (5.15) por

$$f_1(z) = \frac{f(z)}{(1 - zg'(z)/g(z))},$$

obtemos a expansão de Lagrange dada em (5.17).

## B - Função geradora de probabilidade

No estudo de distribuições de probabilidade discretas os cálculos para obtenção de algumas características do modelo podem ser bastante complexos e trabalhosos, por exemplo, os momentos da DPG. No entanto, essa tarefa pode ser reduzida com emprego da função geradora de probabilidade (fgp).

Considere  $X$  uma variável aleatória discreta definida sob o conjunto dos inteiros não negativos e  $P(X = x)$ , para  $x = 0, 1, 2, \dots$ , a função de massa de probabilidade associada a variável  $X$  tal que  $\sum_{x=0}^{\infty} P(X = x) = 1$ . A função geradora de probabilidade da variável aleatória  $X$  é definida por

$$g(t) = E(t^x) = \sum_{x=0}^{\infty} t^x P(X = x).$$

Note que, a fgp é um polinômio cujos coeficientes são as funções de massa de probabilidade aplicada no ponto. Assim, temos que a função geradora de probabilidade é uma série de potência e através desta pode-se obter a função de probabilidade, assim como, a função de probabilidade fornece a fgp. Ou seja, existe uma relação de unicidade entre a fgp e a função de probabilidade de uma variável aleatória  $X$ .

Substituindo  $t$  por  $e^s$ , a fgp  $g(t)$  torna-se a função geradora de momentos  $M_x(s)$ . Tomando o logaritmo natural de  $M_x(s)$ , temos uma série de potência em  $s$ , definida como função geradora de cumulantes  $k_x(s)$ . Simplificando, temos que

$$M_x(t) = g(e^t)$$

e

$$k_x(t) = \ln M_x(t) = \ln g(e^t).$$

Observe também que  $g(0) = P(X = 0)$  e que  $g(1) = \sum_{x=0}^{\infty} P(X = x) = 1$ .

A média e a variância de uma distribuição de probabilidade podem ser obtidas através da fgp calculando os dois primeiros momentos fatoriais que são obtidos tomando sucessivas diferenças de  $g(t)$  e em seguida fazendo  $t = 1$ . Portanto, temos que

$$g'(t)|_{t=1} = \left[ \sum_{x=1}^{\infty} xt^{x-1}P(X=x) \right]_{t=1} = \sum_{x=1}^{\infty} xP(X=x) = E[X];$$

e

$$g''(t)|_{t=1} = \left[ \sum_{x=2}^{\infty} x(x-1)t^{x-2}P(X=x) \right]_{t=1} = \sum_{x=2}^{\infty} x(x-1)P(X=x) = E[X(X-1)].$$

Logo, temos que  $E[X] = \mu'_{(1)} = g'(1)$  e  $E[X(X-1)] = \mu'_{(2)} = g''(1)$  de maneira que

$$Var[X] = E[X^2] - E^2[X] = E[X^2] - E[X] + E[X] - E^2[X] = \mu'_{(2)} + \mu'_{(1)} - (\mu'_{(1)})^2.$$

## C - Propriedades de Esperança Condicional

Apresentamos as provas das propriedades de esperança condicional dadas abaixo, devido a importância destas na demonstração da Proposição 2.3.

$$(i). E(\boldsymbol{\theta}_t^k | D_T) = E[E(\boldsymbol{\theta}_t^k | \boldsymbol{\theta}_{t+1}, D_T) | D_T];$$

$$(ii). V(\boldsymbol{\theta}_t | D_T) = E[Var(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_T) | D_T] + V[E(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_T) | D_T];$$

**Prova:** Para a prova de (i), considere alguns resultados dos modelos de espaço de estados e de propriedades de esperança condicional, de modo que

$$\begin{aligned} E(\boldsymbol{\theta}_t^k | D_T) &= \int \boldsymbol{\theta}_t^k \left[ \int p(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1} | D_T) d\boldsymbol{\theta}_{t+1} \right] d\boldsymbol{\theta}_t \\ &= \int \boldsymbol{\theta}_t^k \left[ \int p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_T) p(\boldsymbol{\theta}_{t+1} | D_T) d\boldsymbol{\theta}_{t+1} \right] d\boldsymbol{\theta}_t \\ &= \int p(\boldsymbol{\theta}_{t+1} | D_T) \left[ \int \boldsymbol{\theta}_t^k p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_T) d\boldsymbol{\theta}_t \right] d\boldsymbol{\theta}_{t+1} \\ &= \int p(\boldsymbol{\theta}_{t+1} | D_T) E(\boldsymbol{\theta}_t^k | \boldsymbol{\theta}_{t+1}, D_T) d\boldsymbol{\theta}_{t+1} \\ &= \int E(\boldsymbol{\theta}_t^k | \boldsymbol{\theta}_{t+1}, D_T) p(\boldsymbol{\theta}_{t+1} | D_T) d\boldsymbol{\theta}_{t+1} \\ &= E[E(\boldsymbol{\theta}_t^k | \boldsymbol{\theta}_{t+1}, D_T) | D_T]. \end{aligned}$$

A demonstração do resultado (ii), segue de modo análogo à prova de (i). Dessa forma,

$$\begin{aligned}
 V(\boldsymbol{\theta}_t|D_T) &= E[\boldsymbol{\theta}_t^2|D_T] - E^2[\boldsymbol{\theta}_t|D_T] \\
 &= E[E(\boldsymbol{\theta}_t^2|\boldsymbol{\theta}_{t+1}, D_T)|D_T] - E^2[E(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, D_T)|D_T] \\
 &= E[E(\boldsymbol{\theta}_t^2|\boldsymbol{\theta}_{t+1}, D_T)|D_T] - E[E^2(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, D_T)|D_T] \\
 &+ E[E^2(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, D_T)|D_T] - E^2[E(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, D_T)|D_T] \\
 &= E[V(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, D_T)|D_T] + V[E(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, D_T)|D_T].
 \end{aligned}$$

## D - Códigos

### D.1 - Capítulo 1

#### Estimação dos parâmetros da distribuição Poisson generalizada utilizando Método dos Momentos e Máxima Verossimilhança

Código 5.1: Estimação dos parâmetros da distribuição Poisson generalizada

---

```

># Função que retorna estimativas da PG pelo Método dos Momentos
2 > emm.dpg <- function(x){
  + phi.m <- 1-sqrt(mean(x)/var(x))
4 + lambda.m <- sqrt((mean(x))^3/var(x))
  + return(c(lambda.m, phi.m))
6 + }

8 > # Algoritmo de Newton-Raphson para calcular estimativas de MV da PG
> env.dpg <- function(x, start){
10 + eps <- 10^(-5)
  + n <- length(x)
12 + phi0 <- start
  + repeat{
14 +   grad <- sum(x*(x - 1)/(mean(x) + phi0*(x - mean(x)))) - sum(x)
  +   hess <- - sum (x*(x - 1)*(x - mean(x))/(mean(x) + phi0*(x -
16 +   mean(x)))^2)
  +   phi <- phi0 - grad/hess
18 +   if(sum( abs(phi - phi0)) < eps){break}
  +   phi0 <- phi

```

```

20 + }
    + lambda <- mean(x)*(1 - phi)
22 + return(c(lambda, phi))
    + }

```

---

## Simulação de variáveis aleatórias com distribuição Poisson generalizada pelo método da Transforma Inversa e da Ramificação

Código 5.2: Simulação de amostras aleatórias do modelo Poisson generalizado

---

```

> # Função de distribuição de probabilidade Poisson generalizada
2 > dpoisg <- function(x, lambda, phi, m){
  + n <- length(x)
4 + aux <- rep(0,n)
  +     dpg <- function(x, lambda, phi, m){
6 +         f <- (lambda*((lambda + phi*x)^(x - 1)) * exp(-lambda -
  +             phi*x))/factorial(x)
8 +     }
  +     for(i in 1:n){
10 +         if(x[i] > m & phi < 0) aux[i] <- 0
  +         else{aux[i] <- dpg(x[i], lambda, phi, m)}
12 +     }
  + return(aux)}
14
> # Função de distribuição acumulada Poisson generalizada
16 > Fdpg <- function(k, lambda, phi, m){
  + x <- c(0:k)
18 + dpg <- dpoisg(x, lambda, phi, m)
  + Fm <- cumsum(dpg)[m+1]
20 +     if(phi < 0){result <- cumsum(dpg/Fm)}
  +     else{result <- cumsum(dpg)}
22 + return(result)}

24 > # Simulação de v.a.s DPG pelo Método da Transformada Inversa
  > sim.dpg.invp <- function(n, k, lambda, phi, m){
26 + Fx <- Fdpg(k, lambda, phi, m)
  + u <- runif(n,0,1)
28 + prob <- matrix(rep(Fx,n), nc = n) - matrix(rep(u,each = k + 1),nc = n)

```

```

+ prob[prob > 0] <- 0
30 + prob[prob < 0] <- 1
+ aux <- apply(prob, 2, sum)
32 + return(aux)}

34 > # Simulação de v.a.s DPG pelo Método da Ramificação
> sim.dpg.brc <- function(n, lambda, phi){
36 + samp <- rep(0, n)
+ cont <- 1
38 + func1.brc <- function(phi, ix, iy){
+ while(iy > 0){
40 + lambda1 <- phi * iy
+ iz <- rpois(n = 1, lambda1)
42 + ix <- ix + iz
+ iy <- iz
44 + }
+ return(ix)}
46 + while(cont <= n){
+ iy <- rpois(n = 1, lambda)
48 + ix <- iy
+ if(iy > 0) ix <- func1.brc(phi, ix, iy)
50 + samp[cont] <- ix
+ cont <- cont + 1
52 + }
+ return(samp)}

```

---

## D.2 - Capítulo 2

Estimação dos estados latentes e Predição de observações futuras da série histórica *AirPassengers* via Modelos Dinâmicos Lineares Normais

---

Código 5.3: Estimação dos estados e Predição

---

```

> require(dlm)
2 > ## Reconhecendo os dados
> Air <- AirPassengers;
4 > n <- length(Air)
> is.ts(Air)

```

```

6 [1] TRUE
  >
8 > ### Filtragem
  > modelo1 <- dlmModPoly(order = 1, dV = 11200, dW = 545)
10 > AirFiltro1 <- dlmFilter(Air, modelo1)
  > modelo2 <- dlmModPoly(order = 1, dV = 11200, dW = 5805)
12 > AirFiltro2 <- dlmFilter(Air, modelo2)
  >
14 > # Estimativa mt e intervalo de credibilidade
  > attach(AirFiltro2)
16 > z <- qnorm(0.025, lower=F)
  > lim.m <- z * sqrt(unlist(dlmSvd2var(U.C, D.C)))
18 > IC.m <- as.vector(m) + lim.m %% c(-1, 1)
  > M <- cbind(m, IC.m)
20 >
  > # Gráfico
22 > plot(Air, type='o', col = c('darkgrey'))
  > lines(dropFirst(AirFiltro1$m), lty=5, col=3, lwd = 2)
24 > lines(dropFirst(AirFiltro2$m), lty=4, col=2, lwd = 2)
  >
26 > ### Suavização – Considerando Modelo 2
  > AirSuave <- dlmSmooth(AirFiltro2)
28 > attach(AirSuave)
  >
30 > # Comparando Filtragem e Suavização
  > drop(dlmSvd2var(U.C[[n+1]], D.C[[n+1]]))
32 [1] 5667.244
  > drop(dlmSvd2var(U.S[[n+1]], D.S[[n+1]]))
34 [1] 5667.244
  > drop(dlmSvd2var(U.C[[n/2+1]], D.C[[n/2+1]]))
36 [1] 5667.244
  > drop(dlmSvd2var(U.S[[n/2+1]], D.S[[n/2+1]]))
38 [1] 3793.346
  >
40 > # Estimativa de st e intervalo de credibilidade
  > lim.s <- z * sqrt(unlist(dlmSvd2var(U.S, D.S)))
42 > IC.s <- as.vector(s) + lim.s %% c(-1, 1)
  > S <- cbind(s, IC.s)

```

```

44 >
    > # Gráfico
46 > plot(dropFirst(M), plot.type='s', lty=c(2,3,3), col=c(1,3,3), lwd = 2)
    > lines(dropFirst(S[,1]), col = 118, lty = 5, lwd = 2)
48 > lines(dropFirst(S[,2]), col = 2, lty = 5, lwd = 2)
    > lines(dropFirst(S[,3]), col = 2, lty = 5, lwd = 2)
50 > lines(Air, type = 'o', col = 'darkgray')
    >
52 > ## Predição um passo à frente - Gráfico
    > attach(modelo2)
54 > plot(dropFirst(f), plot.type = 's', col = 2, lty = 2, lwd = 2)
    > lines(Air, type = 'o', col = 'darkgrey')

```

---

### D.3 - Capítulo 3

#### Ajuste do modelo dinâmico por *superposição* com crescimento linear e efeito sazonal na forma de *Fourier*

Código 5.4: Ajuste do modelo com crescimento linear e efeito sazonal na forma de *Fourier*

---

```

> ## Série Temporal
2 > Air <- AirPassengers
    >
4 > ## Modelo Geral com Tendência e Sazonalidade na forma de Fourier
    > modelo <- dlmModTrig(s = s, dV = 0.4794, dW = rep(15138.38, s-1),
6 + m0 = rnorm(s-1), C0 = diag(1e+02, s-1)) + dlmModPoly(order = 2, dV =
    7.809347e-08, dW = c(1139.5, 0.0086) )
    >
8 > ## Filtro
    > filtro <- dlmFilter(Air, modelo)
10 > mt <- dropFirst(filtro$m)[, c(1, 3, 5, 7, 9, 11, 12)]
    > y.est <- ts(apply(mt,1,sum), start = c(1949,1), end = c(1960,12),
    frequency = 12)
12 >
    > ## Suavização
14 > suave <- dlmSmooth(filtro)
    > st <- suave$s[, c(1, 3, 5, 7, 9, 11, 12)]

```

```

16 > ys.est <- ts(apply(st,1,sum), start = c(1949,1), end = c(1960,12),
    frequency = 12)
    >
18 > ## Gráficos das Estimativas Filtradas e Suavizadas
    > plot(Air, type = 'o', xlab = "Tempo", col = "darkgrey")
20 > lines(yf.est, col = "red", lty = 2)
    > lines(ys.est, col = "blue", lty = 4)
22 >
    > ## Gráfico das Previsões k =36 passos à frente
24 > plot(Air, type = 'o', xlab = 'Tempo', xlim = c(1949, 1964), col = '
    darkgrey')
    > abline(v = 1961, lty = 2)
26 > lines(pred$f, lty = 4, col = 'red')

```

---

## Comparação do modelo dinâmico de regressão com o modelo de regressão padrão

Código 5.5: Comparação do MDL de regressão com o modelo de regressão padrão

```

> ## Matriz de Covariáveis
2 > n <- 48
    > X <- matrix(runif(n = 2*n, 5, 10), ncol = 2)
4 > F <- cbind(1,X)
    >
6 > ## Simulação da série Yt
    > simu.Y <- function(F, theta){
8 + Y <- F %*% theta + rnorm(n)
    + return(Y)
10 + }
    > Y <- ts(simu.Y(F, theta = c(1, 1, 2)))
12 >
    > ## Estimação via Mínimos Quadrados
14 > mod1 <- lm(Y~F)
    > coefs1 <- coef(mod1); coefs1
16 (Intercept)          F1          F2
    0.9835077    0.9544115    2.0546769
18 > var.cov1 <- vcov(mod1); var.cov1
    (Intercept)          F1          F2

```

```

20 (Intercept)  1.76115366 -0.099862044 -0.133975391
    F1          -0.09986204  0.011166493  0.002434025
22 F2          -0.13397539  0.002434025  0.015503940
    > mul <- ts(mod1$fitted.values)
24 >
    > ## Estimação via DLM de regressão
26 > mod2 <- dlmModReg(X = F, addInt = FALSE, dV = 0.978, m0 = c(0.5, 0.7,
    1.2), C0 = diag(c(5, 2, 1)))
    > filtro <- dlmFilter(Y, mod2)
28 > coefs2 <- filtro$m[n + 1,]; coefs2
    (Intercept)          F1          F2
30          0.9580277          0.9586226          2.0537204
    > var.cov2 <- dlmSvd2var(filtro$U.C, filtro$D.C)[[ n + 1]]; var.cov2
32          [,1]          [,2]          [,3]
    [1,]  1.03523477 -0.0588828034 -0.0785543400
34 [2,] -0.05888280  0.0074457715  0.0006268648
    [3,] -0.07855434  0.0006268648  0.0099138120
36 > mu2 <- ts(F%%c(coefs2))

```

---

## D.4 - Aplicação

### Simulação de séries temporais segundo MDPG

Código 5.6: Geração de dados com base na estrutura do MDPG

---

```

> sim.mdp <- function(modelo, N, I, k, m){
2 > mod <- modelo
    > p <- length(mod$m0)
4 > Y <- matrix(0, nr = N, nc = I)
    > theta0 <- t(rmvnorm(n = 1, mean = mod$m0, sigma = mod$C0))
6 > theta <- matrix(0, nr = N, nc = p)
    > eta <- matrix(0, nr = N, nc = I)
8 > aux <- matrix(0, nr = N, nc = I)
    > theta[1, ] <- theta0
10 > eta[1, 1] <- mod$FF%% theta[1, ] + rnorm(n = 1, mean = 0, sd = sqrt(
    mod$V))
12 >
    > # Valor inicial de lambda.star

```

```

14 > aux[1, 1] <- exp(eta[1, 1])/(1 + exp(eta[1, 1]))
    >
16 > # Atualizando o vetor de estados, parâmetro estrutural e parâmetro
    transformado
    > for(j in 2:N){
18 >         theta[j, ] <- mod$GG %% theta[j - 1, ] + t(rmvnorm(n =
    >         1, mean = rep(0, p), sigma = mod$W))
20 >         eta[j, 1] <- mod$FF %% theta[j, ] + rnorm(n = 1, mean =
    >         0, sd = sqrt(mod$V))
22 >         aux[j, 1] <- exp(eta[j, 1])/(1 + exp(eta[j, 1]))
    > }
24 >
    ># Obtendo Valor de phi
26 >limi <- max(-1,-(1/m) * exp(-eta))
    >phi <- limi + (abs(limi) / 50);phi
28 >
    ># Gerando a série temporal Y
30 >     lambda <- exp(-eta)
    >     for(j in 1:N){
32 >         if(phi <= 0){
    >             Y[j] <- sim.dpg.invp(n=1, k, lambda = lambda[j],
34 >             phi = phi, m )
    >         }
36 >         else{
    >             Y[j, i] <- sim.dpg.brc(lambda = lambda[j], phi =
38 >             phi)
    >         }
40 >     }
    > return(list(Y = Y, theta = theta, eta = eta, lb.y = lambda, phi.y =
    phi, mod))
42 >}

```

---

## Estimação dos parâmetros do MDPG via MCMC

Código 5.7: Processo de estimação do MDPG via MCMC

---

```
> ### Funções dos Parâmetros da DPG #
2 > ## Phi
  > # Transformando phi para a escala (0,1)
4 > H1.phi <- function(phi, eta, m){
  + limi <- max(-1, -(1/m) * exp(- eta))
6 + H1 <- (phi - limi) / (1 - limi)
  + return(H1)}
8 >
  > # Transformando H1(phi) na escala logit
10 > H2.phi <- function(phi, eta, m){
  + aux <- H1.phi(phi, eta, m)
12 + H2 <- log(aux / (1 - aux))
  + return(H2)}
14 >
  > # Voltando à escala phi a partir da função H1:
16 > H1.to.phi <- function(H1, eta, m){
  + limi <- max(-1, -(1/m) * exp(- eta))
18 + phi <- H1*(1 - limi) + limi
  + return(phi)}
20 >
  > # Voltando à escala H1 a partir da função H2:
22 > H2.to.H1 <- function(H2){
  + H1 <- exp(H2)/(1 + exp(H2))
24 + return(H1)}
  >
26 > ### Funções dos parâmetros da distribuição candidata de Eta #
  > ## Derivada de ordem 1 da log-verossimilhança
28 > DL1.eta <- function(eta, V, theta, phi, y){
  + aux1 <- exp(-eta) - 1 - ((1 / V)^2) * (eta - FF %% theta)
30 + aux2 <- -((y - 1) * exp(-eta)) / (exp(-eta) + phi * y)
  + deriv <- aux1 + aux2
32 + return(deriv)}
  >
34 > ## Derivada de ordem 2 da log-verossimilhança
  > DL2.eta <- function(eta, V, theta, phi, y){
```

```

36 + aux1 <- exp(-eta) + (1/V)^2
    + aux2 <- (y *(y - 1) * phi * exp(-eta)) / ((exp(-eta) + phi * y)^2)
38 + deriv <- aux2 - aux1
    + return(deriv)}
40 >
    > ## Média da distribuição q
42 > eta.hat <- function(eta, V, theta, phi, y){
    + aux1 <- DL1.eta(eta, V, theta, phi, y)
44 + aux2 <- DL2.eta(eta, V, theta, phi, y)
    + aux3 <- etaj - (aux2)^(-1) * aux1
46 + return(aux3)}
    >
48 > ### Parâmetros da Skew-Normal ##
    > ## Parâmetro de Escala
50 > omega <- function(sigma, alpha){
    + delta <- alpha / sqrt(1 + alpha^2)
52 + fun <- sigma * sqrt(1 / (1 - (2 * delta^2) / pi))
    + return(fun)}
54 >
    > ## Parâmetro de locação
56 > xi <- function(mu, omega, alpha){
    + delta <- alpha / sqrt(1 + alpha^2)
58 + fun <- mu - omega * delta * sqrt(2 / pi)
    + return(fun)}
60 >
    > ## Média
62 > mu.sn <- function(xi, omega, alpha){
    + delta <- alpha / sqrt(1 + alpha^2)
64 + fun <- xi + omega * delta * sqrt(2 / pi)
    + return(fun)}
66 >
    > ## Variância
68 > sigma2.sn <- function(omega, alpha){
    + delta <- alpha / sqrt(1 + alpha^2)
70 + fun <- (omega^2) * (1 - (2 * delta^2) / pi)
    + return(fun)}
72 >
    > ### Especificação das matrizes de variância V e W #

```

```

74 > g10 <- 12; g11 <- 13; g12 <- 14
    > V0 <- (g10 - 2) * 0.0085
76 > Wlambda0 <- (g11 - 2) * diag(c(0.25^2 * 0.005, 0.15^2 * 0.025))
    > Wsazlambda0 <- (g12 - 2) * diag(c(1, 0.5, 0.5)) * 0.0055
78 >
    > #### Modelo Dinâmico Poisson Generalizado (MDPG) #
80 > ## Estruturando as matrizes G e F
    > set.seed(10101010)
82 > N <- 72; s <- 4; I <- 1; m <- 6
    >
84 > ## Modelo com tendência e sazonalidade
    > mod <- dlmModPoly(2) + dlmModSeas(s)
86 >
    > a1 <- mean(eta.init)
88 > m0 <- c(a1*c(0.855, 0.02, 0.105, 0.0055, 0.0055))
    > mod$m0 <- m0
90 >
    > C0 <- bdiag(diag(c(0.0001, 0.0025)), diag(0.0001, 3))
92 > mod$C0 <- C0
    > np <- length(m0)
94 >
    > #### Amostragem das Posteriores via MCMC #
96 > ## Definições importantes
    > MC <- 200000; TT <- length(y.sim)
98 >
    > ## Vetor de estados Theta e de parâmetros estruturais Eta
100 > gibbsTheta <- array(0, dim = c(TT + 1, np, MC - 1))
    > gibbsEta <- array(0, dim = c(TT+1, 1, MC-1))
102 > gibbsEta[, , 1] <- eta.mat <- c(0, eta.init)
    >
104 > ## Matrizes de Variância variam ao longo do tempo
    > gibbsV <- array(0, dim = c(1, 1, MC))
106 > gibbsWlambda <- array(0, dim = c(2, 2, MC))
    > gibbsWsaz <- array(0, dim = c(3, 3, MC))
108 >
    > ## Iniciou m0 com a primeira observação da série observada
110 > mod$V <- gibbsV[ , , 1] <- V0/(delta0 - 2)
    > gibbsWlambda[ , , 1] <- Wlambda0/(delta1 - 2)

```

```

112 > gibbsWsaz[ , , 1] <- Wsazlambda0/(delta2 - 2)
    > mod$W <- bdiag(gibbsWlambda[ , , 1], gibbsWsaz[ , , 1])
114 >
    > ## Vetor para atualização de Phi
116 > gibbsPhi <- array(0, dim = c(1, 1, MC))
    > phi.atual <- gibbsPhi[ , , 1] <- as.numeric(phi.init)
118 > gibbsH2.Phi <- array(0, dim = c(1, 1, MC))
    > H2.atual <- gibbsH2.Phi[ , , 1] <- H2.phi(as.numeric(phi.init), eta.init,
        m)
120 >
    > #### Especificação dos valores iniciais dos hiperparâmetros H2(Phi|Eta)
        #
122 > ## Variâncias candidatas
    > var.1 <- 0.15
124 > sd.1 <- sqrt(var.1)
    >
126 > ## Coeficiente de assimetria da Skew Normal:
    > alpha0 <- alph1 <- - 2.65
128 >
    > ## Média
130 > mu.0 <- H2.phi(phi = as.numeric(phi.init), eta = eta.init[-1], m = m)
    > w <- omega(sigma = sd.1, alpha = alpha0)
132 > xxi <- xi(mu = mu.0, omega = w, alpha = alpha0)
    >
134 > ## Redefinindo Y
    > dimy <- length(ys) + 1
136 > y.new <- matrix(c(0, ys), nr = dimy, nc = 1)
    >
138 > #### Simulação MCMC para as distribuições a posteriori do MDPG #
    > accepted1 <- matrix(0, nr = (TT+1), nc = 1)
140 > accepted2 <- 0
    > set.seed(5614)
142 > for(it in 1:(MC - 1)){
    +     eta.mat <- gibbsEta[ , , it]
144 +     eta.mat1 <- eta.mat
    +     phi.atual <- gibbsPhi[ , , it]
146 +     H2.atual <- gibbsH2.Phi[ , , it]
    +

```

```

148 + # Algoritmo FFBS
      + modFilt <- dlmFilter(eta.mat[-1], mod, simplify = TRUE)
150 + gibbsTheta[ , , it] <- theta <- dlmBSample(modFilt)
      + mod$m0 <- theta[2,]
152 +
      + # Atualização da matriz V
154 + vit <- eta.mat[-1] - theta[-1, ] %*% t(mod$FF)
      + S <- crossprod(vit) + V0
156 + gibbsV[ , , it+1] <- solve(rwishart(df = gl0 + 1 + TT, p = 1,
      + Sigma = solve(S)))
158 + mod$V <- gibbsV[ , , it + 1]
      +
160 + # Atualização das matrizes Wlambda, Wsaz
      + wt <- theta[-1, ] - ( theta[-(TT + 1), ] %*% t(mod$GG))
162 + SS.niv.lambda <- crossprod(wt)[1:2,1:2] + Wlambda0
      + SS.saz.lambda <- crossprod(wt)[3:5,3:5] + Wsazlambda0
164 + gibbsWlambda[ , , it + 1] <- solve(rwishart(df = gl1 + 1 + TT,
      + Sigma = solve(SS.niv.lambda)))
166 + gibbsWsaz[ , , it + 1] <- solve(rwishart(df = gl2 + 1 + TT,
      + Sigma = solve(SS.saz.lambda)))
168 + mod$W <- bdiag(gibbsWlambda[ , , it+1], gibbsWsaz[ , , it + 1])
      +
170 + # Atualizando ETA através do Metropolis Hastings
      + FF <- mod$FF
172 + VV <- mod$V
      + nivel <- FF %*% t(theta)
174 + for(j in 2:(TT + 1)){
      +     etaj <- eta.mat[j-1]
176 +     V1 <- VV
      +     thetaj <- theta[j, ]
178 +     phi2 <- phi.atual
      +     yy <- y.new[j, 1]
180 +
      +     # Passo 1
182 +     eta.original <- eta.mat[j]
      +
184 +     # Passo 2 - Gerar valores candidatos da distribuição
      +     # proposta q

```

```

186 +         kk <- 25
      +         mean.eta <- eta.hat(eta = etaj, V = V1, theta = thetaj,
188 +         phi = phi2, y = yy)
      +         var.eta <- -(DL2.eta(eta = etaj, V = V1, theta = thetaj,
190 +         phi = phi2, y = yy))(-1)
      +         eta.proposto <- rnorm(n = 1, mean = mean.eta, sd = kk *
192 +         sqrt(var.eta))
      +
194 +         # Passo 3 – Obtendo Parâmetros Originais
      +         lambda.orig <- exp(- eta.original)
196 +         lambda.prop <- exp(- eta.proposto)
      +
198 +         # Passo 4 – Probabilidade de aceitação
      +         denomin1 <- dpoisg(x = y.new[j, 1], lambda = lambda.orig
200 +         , phi = phi.atual, m = m) * dnorm(x = eta.original, mean
      +         = nivel[1, j], sd = sqrt(VV))
202 +         numer1 <- dpoisg(x = y.new[j, 1], lambda = lambda.prop,
      +         phi = phi.atual, m = m) * dnorm(x = eta.proposto, mean
204 +         = nivel[1, j], sd = sqrt(VV))
      +         prob.aceitacao <- min(1, (numer1/denomin1))
206 +
      +         # Passo 5 – Gerando variável  $U \sim \text{Uniforme}(0, 1)$ 
208 +         u <- runif(n = 1)
      +
210 +         # Passo 6 – Condição de aceitação
      +         if (u < prob.aceitacao){
212 +             eta.gerado <- eta.proposto
      +             accepted1[j] <- accepted1[j] + 1
214 +         }
      +         else{eta.gerado <- eta.mat[j]}
216 +         eta.mat1[j] <- eta.gerado
      +     }
218 +     gibbsEta[, , it + 1] <- eta.mat1
      +
220 +     ## Metropolis Hastings para  $H_2(\text{PHI}|\text{eta})$ :
      +     # Passo 1
222 +     H2.atual
      +     xi.atual <- xi(mu = H2.atual, omega = w, alpha = alph1)

```

```

224 +
+ # Passo 2 - Gerar valores candidatos da distribuição proposta q
226 + H2.prop <- rsn(n = 1, xi = H2.atual, omega = sd.1, alpha = alph1)
+ xi.prop <- xi(mu = H2.prop, omega = w, alpha = alph1)
228 + H.aux <- H2.to.H1(H2.prop)
+ phi.prop <- H1.to.phi(H.aux, eta.mat1[-1], m)
230 +
+ # Passo 3 - Probabilidade de aceitação
232 + lambda.new <- exp(-eta.mat1[-1])
+ aux1 <- 0
234 + aux2 <- 0
+ for(jj in 1:TT){
236 +     aux1 <- aux1 + log(dpoisg(x = ys[jj], lambda =
+     lambda.new[jj], phi = phi.atual, m = m))
238 +     aux2 <- aux2 + log(dpoisg(x = ys[jj], lambda =
+     lambda.new[jj], phi = phi.prop, m = m))
240 + }
+ r.aux1 <- exp(aux1)
242 + r.aux2 <- exp(aux2)
+ rp1 <- dsn(x = H2.prop, xi = xxi, omega = w, alpha = alph1)
244 + rp2 <- dsn(x = H2.atual, xi = xxi, omega = w, alpha = alph1)
+ R.H2 <- (r.aux2 / r.aux1) * (rp1 / rp2)
246 + r1 <- dsn(x = H2.prop, xi = xi.atual, omega = w, alpha = alph1)
+ r2 <- dsn(x = H2.atual, xi = xi.prop, omega = w, alpha = alph1)
248 + R.H2 <- R.H2 * (r1/r2)
+ prob.aceitacao <- min(1, R.H2)
250 +
+ # Passo 4 - Gerando variável U ~ Uniforme(0, 1)
252 + u <- runif(n = 1)
+
254 + # Passo 5 - Condição de aceitação
+     if (u < prob.aceitacao){
256 +         H2.gerado <- H2.prop
+         accepted2 <- accepted2 + 1
258 +     }
+     else{H2.gerado <- H2.atual}
260 +
+ gibbsH2.Phi[, , it + 1] <- H2.gerado

```

```
262 + # Transforma da escala H2 para a escala de phi
      + H.aux1 <- H2.to.H1(H2.gerado)
264 + phi.gerado <- H1.to.phi(H.aux1, eta.mat1[-1], m)
      + gibbsPhi[, , it + 1] <- phi.gerado
266 + }
```

---

# Referências Bibliográficas

- Armstrong, J. & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8:69–80.
- Cargnoni, C., Müller, P., & West, M. (1997). Bayesian forecasting of multinomial time series through conditionally gaussian dynamic models. *Journal of the American Statistical Association*, 92:640–647.
- Carter, C. K. & Kohn, R. (1994). On gibbs sampling for state-space models. *Biometrika*, 81:541–553.
- Consul, P. C. (1989). *Generalized Poisson Distributions: Applications and Properties*. Marcel Dekker.
- Consul, P. C. & Famoye, F. (1988). Maximum likelihood estimation for the generalized poisson distribution when sample mean is larger than sample variance. *Communications in Statistics - Theory and Methods*, 17:299–309.
- Consul, P. C. & Famoye, F. (2006). *Lagrangian Probability Distributions*. Birkhäuser.
- Consul, P. C. & Jain, G. C. (1973). A generalized of the poisson distribution. *Technometrics*, 15:791–799.
- Consul, P. C. & Shenton, L. R. (1972). Use of lagrange expansion for generating generalized probability distributions. *SIAM Journal of Applied Mathematics*, 23:239–248.
- Consul, P. C. & Shoukri, M. (1984). Maximum likelihood estimation for the generalized poisson distribution. *Communications in Statistics - Theory and Methods*, 13:1533–1547.

- Consul, P. C. & Shoukri, M. (1985). The generalized poisson distribution when the sample mean is larger than the sample variance. *Communications in Statistics - Theory and Methods*, 14(3):667–681.
- Consul, P. C. & Shoukri, M. (1988). Some chance mechanisms related to a generalized poisson probability model. *American Journal of Mathematical and Management Sciences*, 8:181–202.
- Cryer, J. D. & Chan, K. (2008). *Time Series Analysis with Applications in R*. Springer.
- Da-Silva, C. Q. & Migon, H. S. (2012). Hierarchical dynamic beta model. *Journal of Statistical Computation e Simulation*, 0:1–17.
- Denuit, M., Maréchal, X., Pitrebois, S., & Walhin, J. (2007). *Actuarial Modelling of Claim Counts*. John Wiley & Sons Ltd.
- Durbin, J. & Koopman, S. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.
- Famoye, F. (1997). Generalized poisson random variate generation. *American Journal of Mathematical and Management Sciences*, 17:219–237.
- Flores, B. (1986). A pragmatic view of accuracy measurement in forecasting. *International Journal Management Science*, 14(2):93–98.
- Frühwirth-Schnatter (1994). Data augmentation and dynamic exponential family regression. *Metrika*, 38:37–60.
- Gamerman, D. & Migon, H. (1993). Dynamic hierarchical models. *Journal of the Royal Statistical Society*, 55:629–642.
- Harrison, P. J. & Stevens, C. (1976). Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society, Series B*, 38:205–247.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.

- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.
- Hyndman, R. & Koehler, A. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22:679–688.
- James, B. R. (2010). *Probabilidade: Um curso em nível intermediário*. Projeto Euclides.
- Joe, H. & Zhu, R. (2005). Generalized poisson distribution: the property of mixture of poisson and comparison with negative binomial distribution. *Biometrical*, 47:219–229.
- Johnson, N., Kotz, S., & Kemp, A. (1992). *Univariate discrete distribution*. John Wiley & Sons.
- Johnson, R. A. (2007). *Applied Multivariate Statistical Analysis*. Pearson Education International.
- Lerner, B., Lone, A., & Rao, M. (1997). On generalized poisson distributions. *Probability and Mathematical Statistics*, 17(2):377–385.
- Makridakis, S. & Hibon, M. (2000). The m3-competition: results, conclusions and implications. *International Journal of Forecasting*, 16:451–476.
- McCormick, T. H., Raftery, A. E., Madigan, D., & Burd, R. S. (2012). Dynamic logistic regression and dynamic model averaging for binary classification. *Biometrics*, 68:23–30.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(410):1087–1092.
- Migon, H. S. (1984). *An approach to non-linear Bayesian forecasting problems with applications*. PhD thesis, Universidade de Warwick.
- Morettin, P. A. & Toloi, C. M. C. (1986). *Séries Temporais*. Atual Editora LTDA.

- Morettin, P. A. & Tolo, C. M. C. (2006). *Análise de Séries Temporais*. ABE - Projeto Fisher.
- Nelder, J. A. & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of Royal Statistical Society, A*, 135:370–384.
- Petris, G. (2010). An r package for dynamic linear models. *Journal of Statistical Software*, 36(12):1–16.
- Petris, G., Petrone, S., & Campagnoli, P. (2009). *Dynamic Linear Models with R*. Springer.
- Ramalho, J. J. S. (1996). Modelos de regressão para dados de contagem. Master's thesis, Universidade Técnica de Lisboa.
- Reis, E. A. (2008). Modelos dinâmicos para processos pontuais espaço-temporais. universidade federal do rio de janeiro.
- Satterthwaite, F. E. (1942). Generalized poisson distribution. *The Annals of Mathematical Statistics*, 13:410–417.
- Silva, G. S. (2014). Um modelo linear com abafador de tendência dinâmico para previsão bayesiana de séries temporais. Master's thesis, Universidade Federal do Amazonas.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B*, 64:229–235.
- West, M. & Harrison, J. P. (1997). *Bayesian Forecasting and Dynamic Models*. Springer.
- West, M., Harrison, P., & Migon, H. (1985). Dynamic generalized linear models and bayesian forecasting (with discussion). *Journal American Statistical Association*, 80:73–97.