

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

ESTUDOS TEÓRICOS NA ESTIMAÇÃO EM PROCESSOS
K-FACTOR GAMA VIA MODELO DE ESPAÇO DE ESTADOS
UTILIZANDO FILTRO DE KALMAN

Dissertação de Mestrado

VAGNER AUGUSTO BETTI

Porto Alegre, 18 de dezembro de 2012.

Dissertação submetida por Vagner Augusto Betti como requisito parcial para a obtenção do grau de Mestre em Matemática pelo Programa de Pós-Graduação em Matemática do Instituto de Matemática da Universidade Federal do Rio Grande do Sul.

Professor Orientador:

Dr. Cleber Bisognin

Banca Examinadora:

Dr. Cleber Bisognin - PPGMAT(UFRGS)

Dr. Marcio Valk - PPGMAT(UFRGS)

Dr^a. Flávia Tereza Giordani - UFSC

Dr^a. Raquel Romes Linhares - PPGMAT(UFRGS)

Data da Defesa: 18 de dezembro de 2012.

RESUMO

O presente trabalho consiste em um estudo dos processos fracionários generalizados via sistema de espaço de estados. Essa representação permitirá utilizar um procedimento recursivo para a estimação dos parâmetros desses processos chamado filtro de Kalman. Além disso, apresenta um estudo detalhado sobre o filtro de Kalman, investigando sua utilização na estimação de um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$). Este trabalho verifica que é possível representar um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) causal por um sistema de espaço de estados, o que possibilita calcular recursivamente a função de verossimilhança exata do processo, por meio do uso do filtro de Kalman, mesmo que a sua representação em espaço de estados tenha dimensão infinita.

ABSTRACT

This paper presents a study of the generalized fractional processes by using the state space. This representation will allow to use a recursive procedure for the estimation of the parameters of these processes called Kalman filter. In addition, it presents a study with details about the Kalman filter, investigating their use in the estimation of a process k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$). This work find that is possible to represent a process k -Factor GARMA ($p, \mathbf{u}, \boldsymbol{\lambda}, q$) causal in a state space system, which enable to compute recursively the exact likelihood function of the process through the use of the Kalman filter, even though their representation in state space has infinite dimension.

Conteúdo

1	Introdução	2
2	Conceitos e Propriedades	5
2.1	Probabilidade	5
2.2	Funções de uma Variável Aleatória	7
2.3	Processos Estocásticos	18
2.4	Ordens de Aproximação \mathcal{O} e o	22
3	Modelos Precusores para Séries Temporais	24
3.1	Processos ARMA(p, q)	24
3.2	Processos ARIMA(p, d, q)	26
3.3	Processos ARFIMA(p, d, q)	26
4	Processos Fracionários Generalizados	32
4.1	Polinômios Gegenbauer	32
4.2	Processos GARMa	33
4.3	Processos k -Factor GARMa	39
4.4	Estimação	46
4.4.1	ESTIMADOR FT	51
5	Modelos de Espaço de Estados	54
5.1	Representação em Espaço de Estados	57
5.2	Controlabilidade e Observabilidade	65
5.3	Filtro de Kalman	72
5.3.1	Resumo das recursões do Filtro de Kalman	81
5.3.2	Função de Verossimilhança	83
6	Conclusões e Trabalhos Futuros	87
6.1	Conclusões	87
6.2	Trabalhos Futuros	88
	Bibliografia	89

Capítulo 1

Introdução

Frequentemente, ouve-se falar nos meios de comunicação sobre valores diários de poluição nas praias gaúchas, em índices diários da bolsa de valores de São Paulo, na precipitação atmosférica mensal de Porto Alegre, na temperatura média diária das cidades. Todos esses são exemplos de séries temporais, que são conjuntos de observações ordenadas no tempo. O estudo de séries temporais tem contribuído para o desenvolvimento de diversas áreas como agricultura, economia, geofísica, meteorologia, controle de qualidade, ciências sociais, dentre outras. Por esse motivo, estudar tais séries poderá trazer muitos benefícios para nossa sociedade. Existem inúmeros objetivos que permeiam o estudo de séries temporais, dentre os quais a compreensão e descrição do mecanismo gerador da série, a identificação de periodicidades relevantes nos dados e a construção de previsões sobre valores futuros das séries. A fim de estudar séries temporais, procura-se um modelo matemático que melhor se adeque aos dados e que a partir dele seja possível analisar suas características, seu comportamento. Entre os primeiros modelos estudados para representar uma série temporal destacam-se os modelos paramétricos de ordem finita ARMA, autoregressivos e de média móvel, pois conseguem parcimoniosamente descrever uma grande variedade de séries temporais com um número finito de parâmetros.

No entanto, esses modelos não consideravam uma propriedade de certas séries temporais em que observações mesmo bastante distantes no tempo estão correlacionadas, isto é, são dependentes. Esse fenômeno é chamado de longa dependência. Inicialmente, essa propriedade foi identificada por Hurst (1951, 1957) ao estudar a série temporal do nível mensal do Rio Nilo. Mandelbrot e Wallis (1968) e McLeod e Hipel (1978) também verificaram essa propriedade em problemas de Hidrologia. A característica em questão é encontrada em diversas séries temporais como: vazão de rios, tráfego de dados em redes de computadores, sequências de DNA, séries econômicas, entre outras. A partir de então, inúmeros processos foram propostos para modelar a longa dependência como Granger e Joyeux (1980), e Hosking (1981, 1984), que introduziram o modelo fracionário autoregressivo integrado de média móvel, denotado por ARFIMA(p, d, q). Este modelo descreve o comportamento de longa memória associado a um crescimento ilimitado da função densidade espectral quando a frequência tende a zero. Recentemente, os processos de memória longa têm recebido considerável atenção entre os pesquisadores, principalmente em econometria com Cheung and Diebold (1994), e em meteorologia com Bloomfield (1992). A monografia de Beran (1994) e a revisão do artigo de Robinson (1994) fornecem dois atuais estudos de recentes desenvolvimentos de processos de longa dependência em estatística e em economia, respectivamente.

Os processos de longa dependência foram intensivamente investigados a partir da década de 80 pela qual se detectou a existência de dados reais que apresentavam comportamento periodicamente persistente, o qual não era bem modelado pelo processo ARFIMA. Baseado nisso, Gray, Zhang e Woodward (1989) propôs modelos cuja função densidade espectral não era necessariamente ilimitada na origem, permitindo apresentar ilimitação em qualquer frequência no intervalo $[0, \pi]$. Esses modelos foram denominados processos GARMA, autorregressivo de média móvel Gegenbauer, e são uma generalização dos modelos ARFIMA. Alguns anos depois, Giraitis e Leipus (1995) e Woodward, Cheng e Gray (1998) introduziram os processos k -Factor GARMA, uma generalização dos modelos GARMA, que permite a ocorrência de k ilimitações na função densidade espectral, para um número finito k de frequências no intervalo $[0, \pi]$, as quais são chamadas de frequências de Gegenbauer.

Outra classe de modelos bastante importante na análise de séries temporais muito estudada nas últimas décadas são os modelos de espaço de estados, que foram apresentados inicialmente por Kalman (1960), e Kalman e Bucy (1961), com trabalhos relacionados a controle de sistemas lineares. Akaike (1974a) foi um precursor em escrever um modelo ARMA(p, q) na representação de espaço de estados. Esses modelos, nas últimas décadas, têm sido largamente usados para modelar dados em Economia, na área médica e nas ciências dos solos, entre outras áreas da ciência. Para maiores detalhes ver Harrison e Stevens (1976); Shumway e Stoffer (1982); Harvey e Pierse (1984); Kitagawa e Gersch (1984) (área da economia), Jones (1984) (área médica), Shumway (1985) (ciências do solo). A proposta da análise de espaço de estados é inferir relevantes propriedades de séries de dados não observados (estados) a partir do conhecimento de séries de dados observados (observações). Essa representação tem uma estrutura simples que permite uma análise da série temporal, e, sob certas hipóteses, ela ganha uma importante propriedade que aparece em muitas aplicações no campo da Física, é a chamada propriedade de Markov, na qual, dado o estado presente, o futuro do sistema independe do passado. E é devido a ela que os cálculos necessários para implementar os modelos de espaço de estados podem ser colocados de forma recursiva.

O artigo Kalman (1960) descreve uma solução recursiva para a estimação dos parâmetros do modelo de espaço de estados. Em outras palavras, o chamado filtro de Kalman fornece uma solução recursiva para o problema de filtragem linear de dados discretos, e está entre as mais importantes contribuições na teoria moderna de controle e tem sido objeto de estudo de extensivas pesquisas. O avanço da era digital dos computadores nas últimas décadas contribuiu para a disseminação do filtro de Kalman devido à possibilidade da implementação de soluções recursivas para diversas aplicações em tempo real. A navegação aeroespacial, a robótica e o controle de sistemas lineares são exemplos de importantes áreas em que esse filtro é muito utilizado. O filtro de Kalman é um algoritmo de estimação recursiva que tem por objetivo estimar um estado não observado de um sistema linear de espaço de estados, baseados em observações do passado e do presente do sistema.

Neste trabalho, será utilizado o filtro de Kalman para estimar os parâmetros dos modelos k -Factor GARMA. Para essa finalidade, é necessário que esses modelos sejam representados por um sistema de espaço de estados. Deste modo, o objetivo consiste em estudar a estimação dos processos fracionários generalizados, via sistema de espaço de estados, com a utilização do filtro de Kalman. A ênfase, neste trabalho, será dada aos processos k -Factor GARMA.

Para alcançar essa meta, será estudado o que é um sistema de espaço de estados, e como representar os processos fracionários generalizados nesse sistema. Além disso, será investigado

o que é o filtro de Kalman, qual seu funcionamento, e que dados ele fornece, com o intuito de utilizar a ferramenta de estimação recursiva do filtro de Kalman nos parâmetros dos processos fracionários generalizados. Para isso, será investigado sob quais condições pode-se utilizar o filtro de Kalman em processos k -Factor GARMA, e por último, como estimar os parâmetros desse processo usando este procedimento recursivo.

No Capítulo 2, estabelecemos os conceitos iniciais sobre processos estocásticos. Maiores detalhes podem ser encontrados em Brockwell e Davis (2006), Rohatgi (1976), Magalhães (2006), Wei (2006), Morettin e Tolo (2004), Hamilton (1994) e Olive (2008).

No Capítulo 3, abordamos os modelos precursores ARMA, ARIMA e ARFIMA, com algumas de suas propriedades. Ao final deste capítulo, apresentamos uma tabela comparativa sobre o comportamento das funções de autocorrelação e das funções densidade espectral dos modelos AR, MA, ARMA, e ARFIMA.

No Capítulo 4, tratamos os processos fracionários generalizados GARMA e k -Factor GARMA e suas propriedades. O final da Seção 4.3 apresenta uma tabela comparativa sobre o comportamento das funções de autocorrelação e das funções densidade espectral dos modelos ARFIMA, Gegenbauer, GARMA e k -Factor GARMA. Na Seção 4.4, indicamos alguns estimadores para os parâmetros do processo k -Factor GARMA.

No Capítulo 5, abordamos os modelos de espaço de estados com suas representações para os modelos ARMA, ARIMA, ARFIMA, GARMA e k -Factor GARMA. Além disso, estudamos o filtro de Kalman a partir da obtenção de suas equações recursivas. A Subseção 5.3.1 apresenta um resumo dessas equações e um esquema dos ciclos da recursão de Kalman. Na Subseção 5.3.2, analisamos a função de verossimilhança para um modelo gaussiano de espaço de estados.

Capítulo 2

Conceitos e Propriedades

Neste capítulo, enunciamos diversos conceitos fundamentais necessários à compreensão de modelos em séries temporais discutidos neste trabalho. Descrevemos resultados básicos de probabilidade, tais como esperança, variância, função massa de probabilidade, função densidade de probabilidade, função de distribuição, assim como apresentamos idéias fundamentais de processos estocásticos, como funções de autocovariância, função de autocorrelação, função densidade espectral. Além disso, são citadas as definições de estacionariedade, processo ruído branco, processo linear e processo Gaussiano. Maiores detalhes sobre estes assuntos podem ser encontrados em Brockwell e Davis (2006), Hohatgi (1976), Magalhães (2006), e Hamilton (1994).

2.1 Probabilidade

Descrevemos, a seguir, certos aspectos iniciais do ponto de vista probabilístico ou estatístico.

Fixado o conjunto K denotamos $\mathcal{P}(K)$, o conjunto de todos os subconjuntos de K , isto é, o conjunto das partes de K .

Um conjunto K diz-se enumerável quando é finito ou quando existe uma bijeção $f : \mathbb{N} \rightarrow K$.

A seguir, definimos uma σ -álgebra para posteriormente definirmos uma medida de probabilidade.

Definição 2.1. Uma família \mathcal{A} de subconjuntos de um conjunto Ω é chamado de σ -álgebra sobre Ω se:

- a) O conjunto Ω pertence \mathcal{A} ;
- b) Se o conjunto A pertence \mathcal{A} , então o complemento $\Omega - A$ também pertence a \mathcal{A} ;
- c) Se $\{A_n\}_{n \in \mathbb{N}}$ é uma coleção enumerável de conjuntos em \mathcal{A} , então a união $\bigcup_{n \in \mathbb{N}} A_n$ também está em \mathcal{A} .

Segue da Definição 2.1 acima que $\mathcal{A} \subset \mathcal{P}(\Omega)$ e que $\emptyset \in \mathcal{A}$, em que \emptyset é denotado o conjunto vazio. Os conjuntos A pertencentes a σ -álgebra \mathcal{A} são denominados conjuntos \mathcal{A} -mensuráveis, ou simplesmente mensuráveis. Observa-se que se Ω for um conjunto finito ou enumerável, então a única σ -álgebra que se considera sobre Ω é $\mathcal{A} = \mathcal{P}(\Omega)$.

As Definições 2.2 e 2.6 tratam de medida de probabilidade e de variável aleatória, respectivamente. São importantes conceitos que são utilizados para definir um processo estocástico na Definição 2.30.

A seguir, definimos a medida de probabilidade.

Definição 2.2. Uma *probabilidade* \mathbb{P} sobre uma σ -álgebra \mathcal{A} de subconjuntos de Ω é uma função $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ que associa a cada A em \mathcal{A} um número real $\mathbb{P}(A) \in [0, 1]$ tal que:

- a) $\mathbb{P}(\Omega) = 1$;
- b) $\mathbb{P}(A) \geq 0$, para todo subconjunto $A \in \mathcal{A}$;
- c) Se $A_n, n \in \mathbb{N}$, é uma coleção enumerável de subconjuntos de \mathcal{A} tal que $A_n \cap A_m = \emptyset$, para $m \neq n$, então $\mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$.

Como $\Omega \cup \emptyset = \Omega$, então $\mathbb{P}(\emptyset) = 0$. O par (Ω, \mathcal{A}) é chamado de espaço mensurável. Uma vez fixada a medida probabilidade \mathbb{P} sobre a σ -álgebra \mathcal{A} , a tripla $(\Omega, \mathcal{A}, \mathbb{P})$ é chamada de espaço de probabilidade. Os subconjuntos que estão em \mathcal{A} são denominados eventos, e são apenas a eles que se atribuem probabilidade.

A seguir, é definida a probabilidade condicional entre dois eventos para, posteriormente, ser enunciado o Teorema de Bayes que será utilizado para obtenção do filtro de Kalman no Capítulo 4.

Definição 2.3. Sejam A e B dois eventos em $(\Omega, \mathcal{A}, \mathbb{P})$. Sendo $\mathbb{P}(B) > 0$, a *probabilidade condicional de A dado que ocorreu B* , é definida por:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)};$$

Caso $\mathbb{P}(B) = 0$, defini-se $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Definição 2.4. Sejam C_1, C_2, \dots, C_n eventos que estão em $(\Omega, \mathcal{A}, \mathbb{P})$, os quais formam uma partição de Ω e todos tem probabilidade positiva. Seja A um evento qualquer com $\mathbb{P}(A) > 0$. Então, para todo $j = 1, 2, \dots, n$, temos que:

$$\mathbb{P}(C_j|A) = \frac{\mathbb{P}(A|C_j)\mathbb{P}(C_j)}{\sum_{i=1}^n \mathbb{P}(A|C_i)\mathbb{P}(C_i)}.$$

Observação 2.1.

a) C_1, C_2, \dots, C_n formam uma partição de Ω se:

$$\text{i) } \bigcup_{i=1}^n C_i = \Omega;$$

ii) $C_i \cap C_j = \emptyset$, para todo $i \neq j$, em que $i, j = 1, 2, \dots, n$.

b) O Teorema de Bayes mostra uma forma de recalculas as probabilidades *à priori* $\mathbb{P}(C_j)$, $j = 1, 2, \dots, n$. As probabilidades $\mathbb{P}(C_j|A)$, $j = 1, 2, \dots, n$ são ditas probabilidades *à posteriori*, pois são resultantes a partir da utilização das probabilidades *à priori*.

Definimos, em seguida, uma função mensurável para, posteriormente, ser definida uma variável aleatória.

Definição 2.5. Considere Ω e a σ -álgebra sobre Ω denotada por \mathcal{A} , e seja Φ outro conjunto e a σ -álgebra sobre Φ denotada por \mathcal{G} . Uma função $\psi : \Omega \rightarrow \Phi$ é chamada de *mensurável* se $\psi^{-1}(G) \in \mathcal{A}$ para todo $G \in \mathcal{G}$.

Note que, se \mathcal{A} for igual a $\mathcal{P}(\Omega)$, então qualquer função $\psi : \Omega \rightarrow \Phi$ é mensurável. A seguir, é definida uma variável aleatória.

Definição 2.6. Seja $(\Omega, \mathcal{A}, \mathbb{P})$ um espaço de probabilidade e (Φ, \mathcal{G}) um espaço mensurável. Uma função $X : \Omega \rightarrow \Phi$ definida naquele espaço de probabilidade e tomando valores neste espaço mensurável é dita uma *variável aleatória*.

Geralmente se utiliza uma variável aleatória X tomando valores reais, isto é, $\Phi = \mathbb{R}$ e sua σ -álgebra \mathcal{G} como subconjuntos da reta real. Assim, tem-se que uma *variável aleatória real* X é uma função $X : \Omega \rightarrow \mathbb{R}$ tal que $X^{-1}(I) = \{\omega \in \Omega : X(\omega) \in I\} \in \mathcal{A}$. Em outras palavras, uma variável aleatória é uma função do espaço amostral Ω nos Reais em que a imagem inversa de um subconjunto qualquer dos reais I é um evento, isto é, é um subconjunto pertencente a σ -Álgebra \mathcal{A} . É devido a esse fato a garantia de que se pode calcular probabilidades com variáveis aleatórias.

Daqui em diante, toda variável aleatória é considerada variável aleatória real, salvo menção contrária.

2.2 Funções de uma Variável Aleatória

A seguir, definimos a função de distribuição acumulada de uma variável aleatória e apresentamos algumas de suas propriedades, (ver Definição 2.7). Ela é uma importante função que descreve o comportamento probabilístico de uma variável aleatória.

Definição 2.7. A *função distribuição acumulada* de qualquer variável aleatória X em $(\Omega, \mathcal{A}, \mathbb{P})$ é definida por:

$$F_X(x) = \mathbb{P}(X \in (-\infty, x]) = \mathbb{P}(X \leq x),$$

para todo $x \in \mathbb{R}$.

A seguir, são apresentadas as propriedades da função distribuição acumulada.

Proposição 2.1. Se $F_X(x)$ é a função distribuição acumulada de uma variável aleatória X no espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, então:

- a) $\lim_{x \rightarrow -\infty} F_X(x) = 0$;
- b) $\lim_{x \rightarrow +\infty} F_X(x) = 1$;
- c) $F_X(\cdot)$ é uma função não decrescente, isto é, $F_X(x) \leq F_X(y)$ sempre que $x \leq y$ para todo $x, y \in \mathbb{R}$;

d) $F_X(\cdot)$ é uma função contínua à direita.

Dependendo de quais valores uma variável aleatória possa assumir, ela poderá ser dita discreta ou contínua. A seguir, definimos variável aleatória discreta, ver Definição 2.8, e variável aleatória contínua, ver Definição 2.9. Além disso, definimos a função massa de probabilidade, ver Definição 2.8, assim como a função densidade de probabilidade, ver Definição 2.9, juntamente suas propriedades mais importantes para nosso estudo.

Definição 2.8. Uma *variável aleatória é dita ser discreta* se assumir apenas um número finito ou um número infinito enumerável de distintos valores.

A *função massa de probabilidade* de uma variável aleatória discreta X definida em $(\Omega, \mathcal{A}, \mathbb{P})$ que assume valores reais x_1, x_2, \dots é definida por:

$$f_X(x_i) = \mathbb{P}(X = x_i) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x_i\}),$$

para todo $i = 1, 2, \dots$.

Enunciamos, a seguir, as propriedades da função massa de probabilidade.

Proposição 2.2. Se $f_X(x)$ é a função massa de probabilidade de uma variável aleatória X em $(\Omega, \mathcal{A}, \mathbb{P})$ assumindo valores x_1, x_2, \dots , então:

- a) $0 \leq f_X(x_i) \leq 1$, para todo $i = 1, 2, \dots$;
- b) $\sum_{i \in \mathcal{I}} f_X(x_i) = 1$, em que $\mathcal{I} = \{1, 2, \dots\}$ é um conjunto de índices.

Em outras palavras, a soma percorre todos os valores possíveis para a variável aleatória X .

A seguir, é definida uma variável aleatória contínua. Definimos também, a função densidade de probabilidade, que é uma função que descreve o comportamento probabilístico de uma variável aleatória contínua.

Definição 2.9. Uma variável aleatória X em $(\Omega, \mathcal{A}, \mathbb{P})$, com função distribuição acumulada $F_X(\cdot)$ é dita *contínua*, se existir uma função não negativa $f(\cdot)$ tal que:

$$F_X(x) = \int_{-\infty}^x f(t)dt, \text{ para todo } x \in \mathbb{R};$$

A função $f(\cdot)$ é denominada *função densidade de probabilidade da variável aleatória contínua* X , e é denotada por $f_X(\cdot)$.

Em seguida, citamos as principais propriedades da função densidade de probabilidade.

Proposição 2.3. A *função densidade de probabilidade* da variável aleatória X definida em $(\Omega, \mathcal{A}, \mathbb{P})$ satisfaz as seguintes propriedades:

- a) $f_X(x) \geq 0$, para todo $x \in \mathbb{R}$;
- b) $\int_{-\infty}^{+\infty} f_X(x)dx = 1$.

Os próximos resultados, Teorema 2.1 e Teorema 2.2, são importantes propriedades a respeito das funções massa de probabilidade, das funções densidade de probabilidade e das funções de distribuição. Estas funções são usadas para obtenção de informações sobre a variável aleatória, como por exemplo a atribuição de probabilidades a eventos.

O teorema abaixo mostra como obter a função densidade de probabilidade de uma variável aleatória a partir de sua função de distribuição.

Teorema 2.1. Se X tem função densidade de probabilidade $f_X(\cdot)$, então $f_X(x) = \frac{d}{dx}F_X(x) \equiv F'_X(x)$ para valores de $x \in \mathbb{R}$, em que a derivada existe.

O Teorema a seguir fornece um modo de calcularmos probabilidades em intervalos.

Teorema 2.2. Seja X uma variável aleatória, definida no espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, e sejam a e b números reais tais que $a < b$, então vale:

- a) $\mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X < a)$
 - i) Se X é contínua: $F_X(x) = \mathbb{P}(X \leq x)$ e $\mathbb{P}(X < a) = \mathbb{P}(X \leq a) = F_X(a)$
 - ii) Se X é discreta: $\mathbb{P}(X < a) = \mathbb{P}(X \leq a) - \mathbb{P}(X = a)$
 - 1) Se $\mathbb{P}(X = a) \neq 0$, então $\mathbb{P}(X < a) \neq \mathbb{P}(X \leq a) = F_X(a)$
 - 2) Se $\mathbb{P}(X = a) = 0$, então $\mathbb{P}(X < a) = \mathbb{P}(X \leq a) = F_X(a)$;
- b) Se X apresenta função densidade de probabilidade $f_X(x)$, então:

$$\mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx = F_X(b) - F_X(a);$$

- c) Se X apresenta função massa de probabilidade $f_X(x)$, então:

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a),$$

mas $\mathbb{P}(a \leq X \leq b) \neq F_X(b) - F_X(a)$ se $f_X(a) > 0$.

A seguir, definimos a função de distribuição condicional de uma variável aleatória, para mais adiante definirmos a densidade condicional de um vetor aleatório, que é um conceito muito importante que se aplica na teoria de estados como veremos no Capítulo 4.

Definição 2.10. Seja X uma variável aleatória definida em $(\Omega, \mathcal{A}, \mathbb{P})$ e considere um evento $A \in \mathcal{A}$, com $\mathbb{P}(A) > 0$, a *função de distribuição condicional de X , dado que ocorreu A* , é definida por:

$$F_X(x|A) = \mathbb{P}(X \leq x|A) = \frac{\mathbb{P}([X \leq x] \cap A)}{\mathbb{P}(A)}.$$

Podemos generalizar esse conceito, tomando-se a σ -Álgebra de Borel. Deste modo, para $\mathbb{P}(A) > 0$ e qualquer $B \in \mathcal{B}(\mathbb{R})$, podemos definir a *função de distribuição condicional de X, dado que ocorreu A* pela expressão:

$$\mathbb{P}(X \in B | A) = \frac{\mathbb{P}([X \in B] \cap [A])}{\mathbb{P}(A)}.$$

Em seguida, definimos um vetor aleatório com a finalidade de definir a densidade conjunta, e a distribuição conjunta de um conjunto de variáveis aleatórias, os quais são conceitos muito usados no Estudo de Séries Temporais, como veremos no Capítulo 3.

Definição 2.11. Seja $(\Omega, \mathcal{A}, \mathbb{P})$ um espaço de probabilidade. Definimos como *vetor aleatório* a uma função \mathbf{X} de Ω em \mathbb{R}^m que seja \mathcal{A} -mensurável. Em outras palavras, a função representada por $\mathbf{X}(w) = (X_1(w), X_2(w), \dots, X_m(w))$ é tal que para todo $i = 1, 2, \dots, m$ e todo $I_i \subset \mathbb{R}$, temos $X_i^{-1}(I_i) \in \mathcal{A}$.

Note que na definição de vetor aleatório X_1, X_2, \dots, X_m são variáveis aleatórias no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$.

A seguir, definimos a função de distribuição conjunta de um vetor aleatório.

Definição 2.12. A *função de distribuição conjunta* do vetor aleatório \mathbf{X} é definida por:

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}(x_1, x_2, \dots, x_m) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_m \leq x_m),$$

para qualquer $\mathbf{x} = (x_1, x_2, \dots, x_m)$, em que $x_i \in \mathbb{R}$, para todo $i = 1, 2, \dots, m$.

Abaixo, apresentamos as propriedades da Função de distribuição Conjunta.

Proposição 2.4. Seja \mathbf{X} um vetor aleatório definido em $(\Omega, \mathcal{A}, \mathbb{P})$ então, para qualquer $\mathbf{x} \in \mathbb{R}^m$, $F_{\mathbf{X}}(\mathbf{x})$ satisfaz as seguintes propriedades:

- $F_{\mathbf{X}}(\mathbf{x})$ é não decrescente em cada uma de suas coordenadas;
- $F_{\mathbf{X}}(\mathbf{x})$ é contínua à direita em cada uma de suas coordenadas;
- Se para algum j , $x_j \rightarrow \infty$, então $F_{\mathbf{X}}(\mathbf{x}) \rightarrow 1$;
- $F_{\mathbf{X}}(\mathbf{x})$ é tal que, para todos $a_i, b_i \in \mathbb{R}$, $a_i < b_i$, $1 \leq i \leq m$, temos:

$$\mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2, \dots, a_m < X_m \leq b_m) \geq 0.$$

A seguir, definimos a função densidade conjunta de um vetor aleatório discreto.

Definição 2.13. Um vetor aleatório é dito *discreto* se suas componentes são variáveis aleatórias discretas. Sua *função massa de probabilidade conjunta* é definida por:

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(x_1, x_2, \dots, x_m) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_m = x_m);$$

A *função marginal* de X_k , $k = 1, 2, \dots, m$ é dada por:

$$f_{X_k}(x_k) = \mathbb{P}(X_k = x_k) = \sum_{\substack{x_i \\ \forall i \neq k}} f_{\mathbf{X}}(\mathbf{x}) = \sum_{\substack{x_i \\ \forall i \neq k}} \mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_m = x_m).$$

Na proposição, a seguir, enunciamos as propriedades da função massa de probabilidade conjunta.

Proposição 2.5. Seja \mathbf{X} um vetor aleatório discreto em $(\Omega, \mathcal{A}, \mathbb{P})$. Então sua função massa de probabilidade conjunta satisfaz as propriedades:

- a) $f_{\mathbf{X}}(\mathbf{x}) \geq 0$, para todo $\mathbf{x} \in \mathbb{R}^m$;
- b) $\sum_{\mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) = 1$.

A seguir, definimos a função densidade conjunta de um vetor aleatório contínuo.

Definição 2.14. Um vetor aleatório é dito contínuo se suas componentes são variáveis aleatórias contínuas. Em outras palavras, um vetor aleatório é contínuo se, dada a função de distribuição F , existe uma função $f : \mathbb{R}_m \rightarrow \mathbb{R}^+$, denominada *função densidade conjunta*, denotada (f.d.c.), e definida por:

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} f_{\mathbf{X}}(\mathbf{y}) dy_1 dy_2 \cdots dy_m.$$

A função densidade marginal é definida pela expressão:

$$f_{X_k}(x_k) = \underbrace{\int_{x_1} \cdots \int_{x_m}}_{(i \neq k)} f_{\mathbf{X}}(\mathbf{x}) \underbrace{dx_1 dx_2 \cdots dx_m}_{(i \neq k)}.$$

Note que, se desejamos a marginal conjunta de l variáveis, integramos em todas as variáveis exceto as de interesse.

Abaixo, apresentamos algumas propriedades da função densidade conjunta.

Proposição 2.6. Seja \mathbf{X} um vetor aleatório definido em $(\Omega, \mathcal{A}, \mathbb{P})$ então, para qualquer $\mathbf{x} \in \mathbb{R}^m$, $f_{\mathbf{X}}(\mathbf{x})$ satisfaz as seguintes propriedades:

- a) $f_{\mathbf{X}}(\mathbf{x}) \geq 0$, para todo $\mathbf{x} \in \mathbb{R}^m$;
- b) $\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{\mathbf{X}}(\mathbf{x}) dx_1 dx_2 \cdots dx_m = 1$.

Em seguida, definimos a função de distribuição condicional entre variáveis aleatórias.

Definição 2.15. Sejam X e Y duas variáveis aleatórias em $(\Omega, \mathcal{A}, \mathbb{P})$ e $B_1, B_2 \in \mathcal{B}(\mathbb{R})$ com $\mathbb{P}(Y \in B_2) > 0$. Definimos a *função de distribuição condicional de X dado $Y \in B_2$* , pela expressão:

$$\mathbb{P}(X \in B_1 | Y \in B_2) = \frac{\mathbb{P}([X \in B_1] \cap [Y \in B_2])}{\mathbb{P}(Y \in B_2)}.$$

Se $\mathbb{P}(Y \in B_2) = 0$, definimos $\mathbb{P}(X \in B_1 | Y \in B_2) = \mathbb{P}(X \in B_1)$.

Observação 2.2. Se Y é variável aleatória discreta e $y \in \mathbb{R}$ tal que $\mathbb{P}(Y = y) > 0$. Seja X uma variável aleatória qualquer e $B \in \mathcal{B}(\mathbb{R})$, então:

$$\mathbb{P}(X \in B|Y = y) = \frac{\mathbb{P}([X \in B] \cap [Y = y])}{\mathbb{P}(Y = y)}.$$

Temos como consequência:

$$\mathbb{P}(X \in B) = \sum_y \mathbb{P}(Y = y)\mathbb{P}(X \in B|Y = y).$$

Agora tomando $B = (-\infty, x]$, $x \in \mathbb{R}$, obtemos a *função de distribuição condicional de X , dado $Y = y$* :

$$F_{X|Y}(x|Y = y) = \frac{\mathbb{P}([X \leq x] \cap [Y = y])}{\mathbb{P}(Y = y)}.$$

Temos como consequência:

$$F_X(x) = \sum_y \mathbb{P}(Y = y)F_{X|Y}(x|Y = y).$$

Podemos escrever a função distribuição conjunta a partir da condicional:

$$F_{X,Y}(x, y) = \sum_{k:k \leq y} \mathbb{P}(Y = k)F_{X|Y}(x|Y = k).$$

Enunciamos, em seguida, uma definição para a função densidade condicional entre variáveis aleatórias.

Definição 2.16. Sejam X e Y variáveis aleatórias contínuas, podemos definir a *função densidade condicional de X dado Y* , no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, por:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)},$$

note que a densidade condicional é o quociente da densidade conjunta pela densidade marginal.

A seguir, definimos a função de distribuição condicional entre variáveis aleatórias.

Definição 2.17. Sejam X e Y variáveis aleatórias contínuas, podemos definir a *função de distribuição condicional de X dado Y* por:

$$F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(z|y)dz.$$

Observação 2.3. Como consequência das Definições 2.16 e 2.17 obtemos as seguintes expressões:

$$F_{X,Y}(x, y) = \int_{-\infty}^y f_Y(z)F_{X|Y}(x|z)dz;$$

$$F_X(x) = \int_{-\infty}^{+\infty} f_Y(y)F_{X|Y}(x|y)dy.$$

De maneira geral, a função distribuição condicional de X dado Y pode ser caracterizada pelas seguintes equações:

- a) $F_{X,Y}(x, y) = \sum_{k:k \leq y} \mathbb{P}(Y = k)F_{X|Y}(x|Y = k)$, se Y é discreta;
- b) $F_{X,Y}(x, y) = \int_{-\infty}^y f_Y(z)F_{X|Y}(x|z)dz$, se Y é contínua.

Nos resultados a seguir, definimos a Esperança matemática (valor esperado) e a variância de uma variável aleatória que representam uma medida de valor central e uma medida de dispersão da distribuição da variável, respectivamente. Posteriormente, definimos Esperança Condicional de uma variável aleatória, ver Definição 2.22, que é um conceito importante para o estudo de Sistemas de espaço de estados localizado no Capítulo 4 deste trabalho. Além disso, apresentamos a função geradora de momentos e a função característica, as quais são importantes funções para o cálculo de probabilidades e de outras quantidades relacionadas a uma variável aleatória.

Definição 2.18. Seja X uma variável aleatória discreta, com função massa de probabilidade $f_X(\cdot)$, e que assume valores reais x_1, x_2, \dots . Então o *valor esperado, ou esperança matemática, ou média* de X , denotado por $\mathbb{E}(X)$, é definido por:

$$\mathbb{E}(X) = \mu_X = \sum_{i \in \mathcal{I}} x_i f_X(x_i), \text{ em que } \mathcal{I} \text{ é um conjunto de índices.}$$

Teremos o valor esperado finito desde que $\sum_{i \in \mathcal{I}} |x_i| f_X(x_i)$ seja finito.

Seja $g(\cdot)$ uma função real qualquer e X uma variável aleatória. Então $g(X)$ é uma função de variável aleatória. E portanto, a esperança de $g(X)$, denotada por $\mathbb{E}(g(X))$, é definida por:

$$\mathbb{E}[g(X)] = \sum_{i \in \mathcal{I}} g(x_i) f_X(x_i), \text{ em que } \mathcal{I} \text{ é um conjunto de índices.}$$

Teremos o valor esperado finito desde que $\sum_{i \in \mathcal{I}} |g(x_i)| f_X(x_i)$ seja finito.

Se a soma não for absolutamente convergente, então $\mathbb{E}(X)$ e $\mathbb{E}[g(X)]$ não existem.

Definição 2.19. Seja X uma variável aleatória contínua, com função densidade de probabilidade $f_X(x)$. Então o *valor esperado, ou esperança matemática, ou média* de X , denotado por $\mathbb{E}(X)$, é dado por:

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx,$$

desde que a integral esteja bem definida.

O valor esperado será finito se $\int_{-\infty}^{+\infty} |x| f_X(x) dx$ existir.

Seja $g(\cdot)$ uma função real qualquer e X uma variável aleatória. Então a esperança de $g(X)$, denotada por $\mathbb{E}(g(X))$, é definida por:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)f_X(x)dx,$$

desde que a integral esteja bem definida.

O valor esperado será finito se $\int_{-\infty}^{+\infty} |g(x)|f_X(x)dx$ existir.

Em seguida, apresentamos algumas propriedades do valor esperado de uma variável aleatória X bastante utilizadas para a obtenção do filtro de Kalman no Capítulo 4.

Proposição 2.7. Sejam X e Y variáveis aleatórias quaisquer, definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, cujo valor esperado existe e a e b números reais. Então $\mathbb{E}(X)$ e apresenta as seguintes propriedades:

- a) $\mathbb{E}(a) = a$;
- b) $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$;
- c) Se $\mathbb{E}(X + Y)$ existir, então $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;
- d) Se $X \leq Y$, então $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

A seguir, definimos o momento de ordem k , e o momento central de ordem k de uma variável aleatória para, posteriormente, definirmos a variância de uma variável aleatória.

Definição 2.20. O *momento de ordem k* da variável aleatória X é definido por $\mathbb{E}(X^k)$, para $k = 1, 2, \dots$, desde que essa quantidade exista. Se $\mathbb{E}(X) = \mu < \infty$, definimos o *momento central de ordem k* por $\mathbb{E}[(X - \mu)^k]$, sempre que essa quantidade exista.

Definição 2.21. Se $\mathbb{E}(X^2)$ for finita, então a *variância da variável aleatória X* , é definida como o momento central de ordem 2, isto é:

$$Var(X) = \sigma^2 := \mathbb{E}[(X - E(X))^2],$$

o desvio padrão de X é definido por $DP(X) = \sigma := \sqrt{Var(X)}$.

Se $\mathbb{E}(X^2)$ não existe, então $Var(X)$ não existe.

Na proposição a seguir, apresentamos uma expressão alternativa para a variância, e outras expressões bastante úteis para cálculo das variâncias, em que assumimos a existência do valor esperado e da variância.

Proposição 2.8. Uma consequência da definição de variância de uma variável aleatória é a seguinte fórmula:

$$Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.$$

Para fins de cálculo, é importante observar que se X é uma variável aleatória discreta com espaço amostral $\Omega_X = \{x_1, x_2, \dots, x_k\}$, então:

$$\mathbb{E}(X) = \sum_{i=1}^k x_i f_X(x_i) = x_1 f_X(x_1) + x_2 f_X(x_2) + \dots + x_k f_X(x_k)$$

e

$$\mathbb{E}[g(X)] = \sum_{i=1}^k g(x_i) f_X(x_i) = g(x_1) f_X(x_1) + g(x_2) f_X(x_2) + \dots + g(x_k) f_X(x_k).$$

Em particular:

$$\mathbb{E}(X^2) = \sum_{i=1}^k x_i^2 f_X(x_i) = x_1^2 f_X(x_1) + x_2^2 f_X(x_2) + \dots + x_k^2 f_X(x_k).$$

E também:

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^k [x_i - \mathbb{E}(X)]^2 f_X(x_i) \\ &= [x_1 - \mathbb{E}(X)]^2 f_X(x_1) + [x_2 - \mathbb{E}(X)]^2 f_X(x_2) + \dots + [x_k - \mathbb{E}(X)]^2 f_X(x_k). \end{aligned}$$

Se X é variável aleatória contínua com função densidade de probabilidade $f_X(x)$, então:

$$\text{Var}(X) = \int_{-\infty}^{+\infty} [x - \mathbb{E}(X)]^2 f_X(x) dx.$$

A variância é uma importante medida de variabilidade que tem propriedades bastante utilizadas, como veremos nos cálculos para obtenção do filtro de Kalman no Capítulo 4. A seguir, enunciamos algumas propriedades da variância, em que assumimos a existência do valor esperado e da variância.

Proposição 2.9. Sejam X e Y variáveis aleatórias quaisquer, definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, e a e b números reais. Então $\text{Var}(X)$ apresenta as seguintes propriedades, dadas pela expressão:

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

Pela proposição, podemos observar que somar constantes não interfere na dispersão dos valores da variável aleatória, porém multiplicar por constantes multiplicará a variância original pelo quadrado da referida constante.

A seguir, definimos a esperança condicional de uma variável aleatória.

Definição 2.22. Sejam X e Y variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$. E supomos que $f_{X|Y}(\cdot|\cdot)$ é a função massa de probabilidade condicional de X dado que $Y = y$. Então a *esperança condicional* de X dado que $Y = y$ é dada por:

$$\mathbb{E}(X|Y = y) = \sum_{i \in \mathcal{I}} x_i f_{X|Y}(x_i|y), \text{ em que } \mathcal{I} \text{ é um conjunto de índices.}$$

A soma acima será finita desde que $\sum_{i \in \mathcal{I}} |x_i| f_{X|Y}(x_i|y)$ seja finito.

Analogamente, supomos que $f_{X|Y}(\cdot|\cdot)$ é a função densidade de probabilidade condicional de X dado que $Y = y$. Então a *esperança condicional* de X dado que $Y = y$ é dada por:

$$\mathbb{E}(X|Y = y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y) dx.$$

Essa integral será finita desde que $\int_{-\infty}^{+\infty} |x| f_{X|Y}(x|y) dx$ exista.

Observação 2.4. A esperança condicional de X dado que $Y = y$, denotada por $\mathbb{E}(X|Y = y)$, corresponde a média de valores de X , restrita aos valores de Y iguais a y .

Se não particularizarmos o valor de Y , escrevemos $\mathbb{E}(X|Y)$ que é denominada *esperança condicional de X dado Y* .

Em seguida, definimos a variância condicional de uma variável aleatória.

Definição 2.23. Sejam X e Y variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$. A *variância condicional* de X , dado que $Y = y$ é dada por:

$$Var(X|Y = y) = \mathbb{E}\{[X - \mathbb{E}(X|Y = y)]^2\} = \mathbb{E}(X^2|Y = y) - (\mathbb{E}(X|Y = y))^2,$$

desde que existam as esperanças em questão.

Em seguida, apresentamos a função geradora de momentos de uma variável aleatória a qual é uma função que facilita o cálculo de probabilidades e de outras quantidades relacionadas.

Definição 2.24. A *função geradora de momentos* de uma variável aleatória X é definida por:

$$M_X(t) = \mathbb{E}(e^{tX})$$

desde que a esperança seja finita para t em algum intervalo $-t_0 < t < t_0$; com $t_0 > 0$, isto é, para t em alguma vizinhança do zero.

Caso contrário, a função geradora de momentos não existe.

Se X é uma variável aleatória discreta, então $M_X(t) = \sum_{i \in \mathcal{I}} e^{tx_i} f_X(x_i)$;

E se X é uma variável aleatória contínua, então $M_X(t) = \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx$.

A seguir, definimos a função característica de uma variável aleatória. Ela é uma função que facilita o cálculo de probabilidades e de outras quantidades relacionadas, assim como a função geradora de momentos. No entanto, a função característica de uma variável aleatória apresenta a vantagem de sempre existir e o inconveniente de trabalharmos com uma função de valores complexos.

Definição 2.25. A função característica de uma variável aleatória X é definida por:

$$\phi_X(t) = \mathbb{E}(e^{itX}) = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)],$$

para $t \in \mathbb{R}$, e o número complexo $i = \sqrt{-1}$.

Em diversas ocasiões, estamos interessados em estudar variáveis aleatórias independentes e identicamente distribuídas devido a suas inúmeras aplicações e à simplicidade do seu uso.

Definimos, a seguir, a independência entre variáveis aleatórias.

Definição 2.26. Duas variáveis aleatórias X e Y definidas em $(\Omega, \mathcal{A}, \mathbb{P})$, são independentes se a informação sobre uma delas não modifica a probabilidade de ocorrência da outra. Isto é, para qualquer $B_1, B_2 \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(X \in B_1 | Y \in B_2) = \mathbb{P}(X \in B_1)$$

Observação 2.5. A definição acima, de Independência das variáveis aleatórias X e Y é equivalente a seguinte definição:

X, Y independentes se, e somente se,

$$\mathbb{P}(X \in B_1, Y \in B_2) = \mathbb{P}(X \in B_1)\mathbb{P}(Y \in B_2), \text{ para todo } B_1, B_2 \in \mathcal{B}(\mathbb{R}).$$

Essa definição é bastante útil e mais prática de ser usada que a primeira. Em termos da função distribuição acumulada, podemos reescrevê-la do seguinte modo:

$$X, Y \text{ independentes} \Leftrightarrow F_{X,Y}(x, y) = F_X(x)F_Y(y), \text{ para todo } (x, y) \in \mathbb{R}^2.$$

Para as variáveis aleatórias discretas, podemos escrever uma definição equivalente utilizando as funções massa de probabilidade, enquanto que para as variáveis aleatórias contínuas, podemos escrever uma definição equivalente utilizando as funções densidade de probabilidade do modo seguinte:

$$X, Y \text{ independentes} \Leftrightarrow f_{X,Y}(x, y) = f_X(x)f_Y(y), \text{ para todo } (x, y) \in \mathbb{R}^2,$$

em que $f_X(\cdot)$ e $f_Y(\cdot)$ são funções massa de probabilidade caso X e Y sejam discretas, ou $f_X(\cdot)$ e $f_Y(\cdot)$ são funções densidade de probabilidade caso X e Y sejam contínuas.

A seguir, definimos o conceito de variáveis aleatórias identicamente distribuídas.

Definição 2.27. Sejam duas variáveis aleatórias X e Y definidas em $(\Omega, \mathcal{A}, \mathbb{P})$, se $F_Y(x) = F_X(x)$ para todo $x \in \mathbb{R}$, então dizemos que X e Y são *identicamente distribuídas*.

Uma variável aleatória independente e identicamente distribuída é denotada por variável aleatória i.i.d..

A seguir, apresentamos uma proposição que mostra condições equivalentes à definição de variáveis aleatórias identicamente distribuídas.

Proposição 2.10. Sejam X e Y duas variáveis aleatórias em $(\Omega, \mathcal{A}, \mathbb{P})$. Então X e Y são identicamente distribuídas, se qualquer das seguintes condições for satisfeita:

- a) $F_Y(x) = F_X(x)$, para todo $x \in \mathbb{R}$;

- b) $f_Y(x) = f_X(x)$, para todo $x \in \mathbb{R}$;
- c) $\phi_Y(t) = \phi_X(t)$, para todo $x \in \mathbb{R}$;
- d) $M_Y(t) = M_X(t)$, para todo t em algum intervalo $-t_0 < t < t_0$; com $t_0 > 0$.

A seguir, definimos a função de verossimilhança para variáveis aleatórias quaisquer.

Definição 2.28. Seja (X_1, X_2, \dots, X_n) uma amostra de variáveis aleatórias quaisquer cujas componentes apresentam função massa de probabilidade ou função densidade de probabilidade $f_X(x_t, \boldsymbol{\theta})$, para $t = 1, \dots, n$, em que $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$. Definimos a *função de Verossimilhança* como sendo a função densidade de probabilidade conjunta das variáveis aleatórias (X_1, X_2, \dots, X_n) . Isto é:

$$L(\boldsymbol{\theta}, \mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}, \boldsymbol{\theta})$$

Note que a *função de verossimilhança* é uma função do vetor de parâmetros $\boldsymbol{\theta}$ supondo $\mathbf{x} = (x_1, \dots, x_n)$ conhecido. O *Estimador de máxima verossimilhança* será o valor $\hat{\boldsymbol{\theta}}(\mathbf{x})$ que maximiza a função $L(\boldsymbol{\theta}, \mathbf{x})$.

O *Estimador de máxima verossimilhança* para $\boldsymbol{\theta}$ é definido como $\hat{\boldsymbol{\theta}}(\mathbf{x}) = \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \mathbf{x})$.

Em seguida, definimos a função de verossimilhança para variáveis aleatórias independentes.

Definição 2.29. Seja $(X_t)_{t=1}^n$ uma amostra aleatória, isto é, variáveis aleatórias independentes e identicamente distribuídas com função massa de probabilidade ou função densidade de probabilidade $f_{X_t}(x_t, \boldsymbol{\theta})$, para $t = 1, \dots, n$, em que $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$. Definimos a *função de Verossimilhança* $L(\boldsymbol{\theta}, \mathbf{x})$ como:

$$L(\boldsymbol{\theta}, \mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}, \boldsymbol{\theta}) = \prod_{t=1}^n f_{X_t}(x_t, \boldsymbol{\theta}).$$

Para mais detalhes sobre conceitos de probabilidade, de variáveis aleatórias, de funções de variáveis aleatórias ver Rohatgi (1976) e Magalhães (2006).

2.3 Processos Estocásticos

Notação: Antes de definirmos um processo estocástico, é conveniente introduzirmos uma notação para o conjunto dos números inteiros não-negativos e para o conjunto dos números inteiros não-positivos, respectivamente, dadas por:

$$j \in \mathbb{Z}_{\geq} = \{j \in \mathbb{Z}; j \geq 0\},$$

e

$$j \in \mathbb{Z}_{\leq} = \{j \in \mathbb{Z}; j \leq 0\}.$$

Definição 2.30. Um processo estocástico é uma família de variáveis aleatórias $\{X_t\}_{t \in \mathbb{T}}$, em que todas as variáveis aleatórias estão definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, sendo $\mathbb{T} \neq \emptyset$ um conjunto de índices, e $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ a medida de probabilidade.

Em geral consideramos $\mathbb{T} = \mathbb{Z}$, porém \mathbb{T} também pode ser $\mathbb{N}, [0, +\infty)$, ou \mathbb{R} . Utilizaremos no presente trabalho $\mathbb{T} = \mathbb{Z}$, exceto quando mencionado o contrário.

As definições das funções de autocovariância e de autocorrelação, a seguir, permitem definir o grau de dependência entre as variáveis aleatórias, fornecendo, assim, informações a respeito da estrutura de dependência de um processo estocástico.

Definição 2.31. Seja $\{X_t\}_{t \in \mathbb{T}}$ um processo estocástico tal que $Var(X_t) < \infty$ para todo $t \in \mathbb{T}$. A *função de autocovariância*, é dada por:

$$\gamma_X(r, s) \equiv Cov(X_r, X_s) = \mathbb{E}[(X_r - \mathbb{E}(X_r))(X_s - \mathbb{E}(X_s))], \text{ para todo } r, s \in \mathbb{T}.$$

Definição 2.32. Seja $\{X_t\}_{t \in \mathbb{T}}$ um processo estocástico tal que $Var(X_t) < \infty$ para todo $t \in \mathbb{T}$. A *função de autocorrelação* do processo, é dada por:

$$\rho_X(r, s) = \frac{\gamma_X(r, s)}{\sqrt{Var(X_r)} \cdot \sqrt{Var(X_s)}}, \text{ para todo } r, s \in \mathbb{T}, \text{ em que } Var(X_t) = \gamma_X(t, t);$$

No uso de modelos que descrevem processos físicos é necessária a utilização de certas hipóteses para conseguirmos desejáveis propriedades probabilísticas e estatísticas. Em processos estocásticos, em geral, utiliza-se a hipótese de estacionariedade. Os conceitos abaixo são referidos a estacionariedade de processos estocásticos.

Definição 2.33. Um processo estocástico $\{X_t\}_{t \in \mathbb{Z}}$ é dito ser *fortemente estacionário* se as funções de distribuição conjunta de $(X_{t_1}, \dots, X_{t_n})$ e $(X_{t_1+h}, \dots, X_{t_n+h})$ são as iguais para todo $n \in \mathbb{N}$, $h \in \mathbb{Z}$, $t_i, t_i + h \in \mathbb{Z}$, para $i = 1, 2, \dots, n$. Isto é:

$$F_{X_{t_1}, \dots, X_{t_n}}(x_1, \dots, x_n) = F_{X_{t_1+h}, \dots, X_{t_n+h}}(x_1, x_2, \dots, x_n)$$

Definição 2.34. Um processo estocástico $\{X_t\}_{t \in \mathbb{Z}}$ é dito ser *fracamente estacionário ou estacionário* se:

- a) $\mathbb{E}(|X_t|^2) < \infty$, para todo $t \in \mathbb{Z}$;
- b) $\mathbb{E}(X_t) = \mu$, uma constante independente de t ;
- c) $\gamma_X(r, s) = \gamma_X(r + t, s + t)$, para quaisquer $r, s, t \in \mathbb{Z}$.

Intuitivamente, um processo é estacionário se, ao longo do tempo, permanecer próximo a sua média e se a escolha de uma origem não influenciar em seu comportamento.

Como a função de autocovariância de um processo estocástico estacionário, para quaisquer $r, s \in \mathbb{Z}$, satisfaz $\gamma_X(r, s) = \gamma_X(r - s, 0)$, podemos redefinir a função de autocovariância de ordem k de um processo estacionário como:

$$\gamma_X(k) \equiv \gamma_X(k, 0) = Cov(X_{t+k}, X_t),$$

para quaisquer $k, t \in \mathbb{Z}$.

Em seguida, citamos algumas propriedades da função de autocovariância de um processo estocástico estacionário.

Proposição 2.11. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico estacionário. A função de autocovariância do processo $\{X_t\}_{t \in \mathbb{Z}}$ apresenta as seguintes propriedades:

- a) $|\gamma_X(k)| \leq \gamma_X(0)$, para todo $k \in \mathbb{N}$;
- b) $\gamma_X(k) = \gamma_X(-k)$, para todo $k \in \mathbb{N}$.

A seguir, enunciamos algumas propriedades da função de autocorrelação de um processo estocástico estacionário.

Proposição 2.12. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico estacionário. A função de autocorrelação do processo $\{X_t\}_{t \in \mathbb{Z}}$ apresenta as seguintes propriedades:

- a) $\rho_X(0) = 1$;
- b) $|\rho_X(k)| \leq 1$, para todo $k \in \mathbb{N}$;
- c) $\rho_X(k) = \rho_X(-k)$, para todo $k \in \mathbb{N}$.

Definição 2.35. Sejam $\{X_t\}_{t \in \mathbb{Z}}$ e $\{Y_t\}_{t \in \mathbb{Z}}$ dois processos estocásticos estacionários. A *função de covariância cruzada entre X_r e Y_s* depende somente do índice de tempo ($s - r$) e é definida pela expressão:

$$\gamma_{XY}(r, s) \equiv Cov(X_r, Y_s) = \mathbb{E}[(X_r - \mathbb{E}(X_r))(Y_s - \mathbb{E}(Y_s))]$$

Definição 2.36. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico estacionário com função de autocovariância absolutamente convergente, isto é, $\sum |\gamma_X(k)| < \infty$ para todo $k \in \mathbb{Z}$. A *função densidade espectral de $\{X_t\}_{t \in \mathbb{Z}}$* é definida por:

$$f_X(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_X(k) e^{-iwk} = \frac{1}{2\pi} [\gamma_X(0) + 2 \sum_{k=1}^{\infty} \gamma_X(k) \cos(wk)], \quad (2.1)$$

para todo $w \in [-\pi, \pi]$.

Teorema 2.3. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico estacionário com função de autocovariância absolutamente convergente. A função densidade espectral, dada pela expressão (2.1), apresenta as seguintes propriedades:

- a) $f_X(w)$ é uma função real contínua, para todo $w \in [-\pi, \pi]$;
- b) $f_X(w) \geq 0$, para todo $w \in [-\pi, \pi]$ (Função não negativa);
- c) $f_X(w) = f_X(-w)$, para todo $w \in [-\pi, \pi]$ (Função par);
- d) $f_X(w) = f_X(w + 2\pi)$, para todo $w \in [-\pi, \pi]$ (Função periódica de período 2π).

Observação 2.6. Devido à simetria da função densidade espectral, dada pelo item c) acima, de modo geral, consideramos ao longo do texto que as frequências $w \in (0, \pi]$.

Mais detalhes sobre função densidade espectral e suas propriedades podem ser encontrados em Brockwell e Davis (1991) e Wei (1990). Neste último contém a demonstração do Teorema anterior.

A seguir, apresentamos a definição de um processo ruído branco.

Definição 2.37. O processo $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é dito *ruído branco* com média zero e variância σ_ε^2 , denotado por $\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$, se, e somente se:

- a) $\mathbb{E}(\varepsilon_t) = 0$, para todo $t \in \mathbb{Z}$;
- b) $Var(\varepsilon_t) = \mathbb{E}(\varepsilon_t^2) = \sigma_\varepsilon^2$, para todo $t \in \mathbb{Z}$;
- c) $\gamma_\varepsilon(k) = \begin{cases} \sigma_\varepsilon^2, & \text{se } k = 0; \\ 0, & \text{se } k \neq 0. \end{cases}$

Em seguida, definimos o operador de retardo ou defasagem de um processo estocástico.

Definição 2.38. Seja $\{X_t\}_{t \in \mathbb{T}}$ um processo estocástico. \mathcal{B} é dito um *operador de retardo ou defasagem* se $\mathcal{B}^j X_t = X_{t-j}$, em que $j \in \mathbb{Z}$.

Definimos, em seguida, a causalidade e invertibilidade de um processo estocástico.

Definição 2.39. Um processo estocástico $\{X_t\}_{t \in \mathbb{Z}}$ é dito *causal* se existe uma sequência $\{\psi_n\}_{n \in \mathbb{Z}_{\geq}}$ tal que:

$$\sum_{n \in \mathbb{Z}_{\geq}} |\psi_n| < \infty, \quad e \quad X_t = \sum_{n \in \mathbb{Z}_{\geq}} \psi_n \varepsilon_{t-n}, \quad t \in \mathbb{Z}.$$

Definição 2.40. Um processo estocástico $\{X_t\}_{t \in \mathbb{Z}}$ é dito *invertível* se existe uma sequência $\{\pi_n\}_{n \in \mathbb{Z}_{\geq}}$ tal que:

$$\sum_{n \in \mathbb{Z}_{\geq}} |\pi_n| < \infty, \quad e \quad \varepsilon_t = \sum_{n \in \mathbb{Z}_{\geq}} \pi_n X_{t-n}, \quad t \in \mathbb{Z}.$$

A seguir, definimos um processo linear.

Definição 2.41. Um processo $\{X_t\}_{t \in \mathbb{Z}}$ é um *processo linear* se puder ser escrito na forma:

$$X_t = \mu + \sum_{j \in \mathbb{Z}_{\geq}} \psi_j \varepsilon_{t-j} = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = \mu + \psi(\mathcal{B})\varepsilon_t, \quad (2.2)$$

em que $\psi(\mathcal{B}) = 1 + \psi_1 \mathcal{B} + \psi_2 \mathcal{B}^2 + \dots$; $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um processo ruído branco (ver Definição 2.37), e μ é a média do processo $\{X_t\}_{t \in \mathbb{Z}}$.

Observação 2.7. Um modelo de filtro linear é representado por um processo linear. Assim, a função $\psi(\mathcal{B})$ é denominada função de transferência do filtro. Ainda, caso a sequência $\{\psi_j\}_{j \geq 1}$ seja finita ou infinita convergente, o filtro é dito estável e o processo $\{X_t\}_{t \in \mathbb{Z}}$ é estacionário.

Teorema 2.4. (*Decomposição de Wold*). Qualquer processo $\{X_t\}_{t \in \mathbb{Z}}$ fracamente estacionário com média $\mu = 0$ pode ser representado, de maneira única, pela soma de um componente determinístico com um componente não determinístico, do seguinte modo:

$$X_t = \sum_{j \in \mathbb{Z}_{\geq}} \psi_j \varepsilon_{t-j} + k_t, \quad (2.3)$$

em que $\psi_0 = 1$ e $\sum_{j \in \mathbb{Z}_{\geq}} \psi_j^2 < \infty$, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um processo ruído branco, e k_t é não correlacionado com ε_{t-j} para qualquer $j \in \mathbb{Z}_{\geq}$.

O termo k_t é chamado componente determinístico do processo $\{X_t\}_{t \in \mathbb{Z}}$, enquanto $\sum_{j \in \mathbb{Z}_{\geq}} \psi_j \varepsilon_{t-j}$ é denominado componente não determinístico. Se $k_t \equiv 0$, então o processo é chamado puramente não determinístico e podemos escrever:

$$X_t = \sum_{j \in \mathbb{Z}_{\geq}} \psi_j \varepsilon_{t-j}.$$

A demonstração do Teorema da Decomposição de Wold pode ser encontrada em Wold (1938).

Observação 2.8.

- Note que todo processo linear $\{X_t\}_{t \in \mathbb{Z}}$ (ver Definição 2.41) pode ser colocado na representação Wold quando tomados $k_t = \mu$.
- A expressão $\widehat{\mathbb{E}}(X_t | X_{t-1}, X_{t-2}, \dots)$ representa a Esperança estimada do processo no tempo t condicionada ao processo nos tempos passados $t-1, t-2, \dots$.

Para mais detalhes sobre processos lineares, ver Brockwell e Davis (1991), Hamilton (1994), Wei (1990), Palma (2007).

A seguir, definimos um processo estocástico Gaussiano, importante processo frequentemente utilizado na literatura, em particular para o estudo de espaço de estados como veremos no Capítulo 4.

Definição 2.42. O processo $\{X_t\}_{t \in \mathbb{Z}}$ é um processo estocástico Gaussiano, se, e somente se, para qualquer subconjunto $\{t_1, t_2, \dots, t_n\}$ de \mathbb{Z} , isto é, para todo $n \in \mathbb{N}$ e todo $t_i \in \mathbb{Z}$, para $1 \leq i \leq n$, as variáveis aleatórias $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ tem distribuição conjunta normal.

Observação 2.9. Se ε_t é um processo ruído branco gaussiano, então um processo do tipo $X_t = \sum_{j \in \mathbb{Z}_{\geq}} \psi_j \varepsilon_{t-j}$ é um processo gaussiano, em que ψ_j é uma constante real para cada $j \in \mathbb{Z}_{\geq}$.

2.4 Ordens de Aproximação \mathcal{O} e o

Nesta seção, apresentamos as definições das ordens de aproximação \mathcal{O} e o e também algumas propriedades das mesmas. Estes conceitos serão utilizados nas Seção 4.4 do Capítulo 4.

Definição 2.43. Sejam $f, g : \mathbb{R} \rightarrow \mathbb{R}$ duas funções reais quaisquer. Então, dizemos que $f(\cdot)$ é de ordem menor ou igual a $g(\cdot)$, denotada por $f(x) = \mathcal{O}(g(x))$, quando $x \rightarrow \infty$, se e somente se, existem $A > 0$ e $x_0 \in \mathbb{R}$, tais que $|f(x)| \leq A|g(x)|$, para todo $x > x_0$. Em outras palavras, $f(x) = \mathcal{O}(g(x))$, quando $x \rightarrow \infty$, significa que $\left| \frac{f(x)}{g(x)} \right|$ é limitada para x suficientemente grande.

Definição 2.44. Sejam $f, g : \mathbb{R} \rightarrow \mathbb{R}$ duas funções reais quaisquer. Então, dizemos que $f(\cdot)$ é de ordem (estritamente) menor a $g(\cdot)$, denotada por $f(x) = o(g(x))$, quando $x \rightarrow \infty$, se e somente se, para todo $A > 0$, existe $x_0 \in \mathbb{R}$, tal que $|f(x)| < A|g(x)|$, para todo $x > x_0$. Em outras palavras, $f(x) = o(g(x))$, quando $x \rightarrow \infty$, significa que $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0$.

Definição 2.45. Sejam $f, g : \mathbb{R} \rightarrow \mathbb{R}$ duas funções reais quaisquer. Então, dizemos que $f(\cdot)$ é assintoticamente igual a $cg(\cdot)$ e denotamos por $f(x) \sim cg(x)$, quando $x \rightarrow x_0$, se e somente se, $\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = c$, em que c é uma constante finita.

A seguir, apresentamos algumas propriedades, operações elementares e relações dos símbolos apresentados nas Definições 2.43-2.45.

Propriedades:

- i) A relação da Definição 2.45 é uma relação de equivalência, isto é, ela é reflexiva, simétrica e transitiva.
- ii) Multiplicação por uma constante: $\mathcal{O}(kf(x)) = k\mathcal{O}(f(x)) = \mathcal{O}(f(x))$, para todo $k \neq 0$.
- iii) Adição de uma constante: $\mathcal{O}(k + f(x)) = k + \mathcal{O}(f(x)) = \mathcal{O}(f(x))$.
- iv) Composição: $\mathcal{O}(\mathcal{O}(f(x))) = \mathcal{O}(f(x))$.
- v) $\mathcal{O}(f(x)) + \mathcal{O}(f(x)) = \mathcal{O}(f(x))$.

As propriedades acima citadas, também são válidas para o símbolo o .

- vi) Se $f(x) \sim g(x)$, então $f(x) = g(x)(1 + o(1))$.
- vii) $f(x) = o(1)$, quando $x \rightarrow \infty$, significa que $\lim_{x \rightarrow \infty} f(x) = 0$.
- viii) $f(x) = \mathcal{O}(1)$, quando $x \rightarrow \infty$, significa que $|f(x)|$ é limitada, quando x tende ao ∞ .

Capítulo 3

Modelos Precursores para Séries Temporais

3.1 Processos ARMA(p, q)

Segundo Brockwell and Davis (2006), uma série temporal é um conjunto de observações registradas ao longo do tempo. Em nosso cotidiano são encontrados inúmeros exemplos de séries temporais, como a quantidade de chuvas mensais no Rio Grande do Sul durante o ano de 2011, os índices diários da Bolsa de São Paulo ao longo de 2012, os níveis da temperatura média diária de São José dos Ausentes durante os anos de 2000 a 2011, entre outros. Com a análise de séries temporais é possível estudar e descrever o comportamento de tais sequências de dados, bem como fazer previsões de valores futuros dessas séries. A seguir, descrevemos alguns dos principais modelos matemáticos para séries temporais pertinentes ao presente trabalho.

Com o objetivo de introduzir os processos fracionários generalizados, primeiramente abordamos os principais processos precursores que os originaram.

Os processos fracionários generalizados são uma extensão dos processos estocásticos introduzidos por G.E.P. Box e G.M. Jenkins na década de 70. Estes são definidos em termos de equações de diferenças lineares com coeficientes constantes, os quais são chamados de *processos autorregressivos de médias móveis* (ARMA). Os modelos ARMA foram um dos primeiros processos a modelar séries temporais e incluem, como casos particulares, os precursores processos AR e MA. Esses processos modelavam inúmeras séries temporais com um número finito e enxuto de parâmetros, o que foi uma revolução para a época.

Definição 3.1. O processo estocástico $\{X_t\}_{t \in \mathbb{Z}}$ é dito ser um processo *autorregressivo média móvel de ordem* (p, q) , com média μ igual zero, denotado por ARMA(p, q), se é um processo estacionário tal que:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (3.1)$$

para qualquer $t \in \mathbb{Z}$, em que $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um processo ruído branco, ϕ_i , $1 \leq i \leq p$, e θ_j , $1 \leq j \leq q$, são constantes reais.

Podemos reescrever a expressão (3.1) como:

$$\phi(\mathcal{B})(X - \mu) = \theta(\mathcal{B})\varepsilon_t, \quad t \in \mathbb{Z},$$

(3.2)

em que $\phi(\mathcal{B})$ e $\theta(\mathcal{B})$ são polinômios de ordem p e q , respectivamente, dados por:

$$\phi(z) = 1 - \phi_1 z^1 - \dots - \phi_p z^p = \sum_{i=0}^p (-\phi_i) z^i, \text{ para } p \in \mathbb{N}, \quad (3.3)$$

$$\theta(z) = 1 + \theta_1 z^1 + \dots + \theta_q z^q = \sum_{j=0}^q \theta_j z^j, \text{ para } q \in \mathbb{N}, \quad (3.4)$$

com $\theta_0 \equiv 1$; θ_m , para $1 \leq m \leq q$, são constantes reais, e com $\phi_0 \equiv -1$; ϕ_ℓ , para $1 \leq \ell \leq p$, são constantes reais.

Temos que \mathcal{B} é o operador de defasagem, isto é, $\mathcal{B}X_t = X_{t-1}$.

Observação 3.1. Se $\theta(\mathcal{B}) \equiv 1$, então o processo $\phi(\mathcal{B})X_t = \varepsilon_t$ é dito ser um processo *autor-regressivo de ordem p* , denotado por $\text{AR}(p)$. Da mesma forma se $\phi(\mathcal{B}) \equiv 1$, então o processo $X_t = \theta(\mathcal{B})\varepsilon_t$ é dito ser um processo *média móvel de ordem q* , denotado por $\text{MA}(q)$.

Assumindo que as equações $\phi(z) = 0$ e $\theta(z) = 0$ não tem raízes em comum temos que um processo $\text{ARMA}(p, q)$ é causal (ver Definição 2.39) se as raízes de $\phi(z) = 0$ estão fora do círculo unitário, e é invertível (ver Definição 2.40) se as raízes de $\theta(z) = 0$ estão fora do círculo unitário.

O Teorema a seguir nos fornece a função densidade espectral de um processo $\text{ARMA}(p, q)$ cuja demonstração pode ser encontrada em Brockwell e Davis (2006).

Teorema 3.1. Sejam $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico estacionário $\text{ARMA}(p, q)$ e $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ o processo ruído branco com média zero e variância σ_ε^2 . A *função densidade espectral* de $\{X_t\}_{t \in \mathbb{Z}}$ é dada por:

$$f_X(w) = \frac{\sigma_\varepsilon^2}{2\pi} \left| \frac{\theta(e^{-iw})}{\phi(e^{-iw})} \right|^2, \text{ para todo } w \in (0, \pi], \quad (3.6)$$

em que $\phi(\cdot)$ e $\theta(\cdot)$ são os polinômios do processo $\text{ARMA}(p, q)$.

Observação 3.2. A função de autocorrelação dos processos $\text{ARMA}(p, q)$ tem decaimento exponencial (decrece rapidamente para zero), isto é, pode ser demonstrado que:

$$|\rho_x(k)| \leq Cr^k, \text{ quando } k \rightarrow +\infty,$$

onde $C > 0$ e $r \in (0, 1)$.

A expressão acima assegura que a função de autocorrelação é geometricamente limitada.

3.2 Processos ARIMA(p, d, q)

Definimos a seguir o processo ARIMA(p, d, q), generalização dos processos ARMA(p, q), cujos os parâmetros p e q são, respectivamente, o número de termos autorregressivos e médias móveis dos polinômios, e d é o número de diferenças inteiras necessárias para tornar a série estacionária.

Definição 3.2. Um processo estocástico $\{X_t\}_{t \in \mathbb{Z}}$ é dito ser um processo *autorregressivo integrado média móvel* se a diferenciação $(1 - \mathcal{B})^d X_t$ resultar em um processo ARMA(p, q) em que $d \in \mathbb{N}$.

O modelo ARIMA(p, d, q) tem a seguinte expressão:

$$\phi(\mathcal{B})(1 - \mathcal{B})^d X_t = \theta(\mathcal{B})\varepsilon_t, \text{ para todo } t \in \mathbb{Z},$$

em que $\phi(\cdot)$ e $\theta(\cdot)$ são os polinômios dados na Definição 3.1 e $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ um processo ruído branco com média zero e variância σ_ε^2 . O parâmetro $d \in \mathbb{N}$ é chamado de parâmetro de diferenciação. O processo pode ser reescrito chamando $Y_t \equiv (1 - \mathcal{B})^d X_t$ da seguinte maneira:

$$\phi(\mathcal{B})Y_t = \theta(\mathcal{B})\varepsilon_t, \text{ para todo } t \in \mathbb{Z},$$

tal que Y_t é um processo estacionário ARMA(p, q).

Note que, se tomarmos o parâmetro de diferenciação d como zero num processo ARIMA(p, d, q), obteremos um processo ARMA(p, q).

3.3 Processos ARFIMA(p, d, q)

Existem séries temporais cujas observações, mesmo bastante distantes no tempo, apresentam grande dependência entre elas. Esse fenômeno foi identificado, primeiramente, pelo hidrólogo Harold E. Hurst em 1951, enquanto investigava a série temporal dos níveis do rio Nilo, e por este motivo ficou conhecido como "Efeito Hurst". Hoje, denominamos essa característica de longa dependência.

Para explicar o efeito Hurst, Mandelbrot (1965) e Mandelbrot e van Ness (1968) utilizaram os conceitos de integração e de diferenciação fracionária e definiram o *Movimento Browniano Fracionário*. Este é um processo estocástico estacionário a tempo contínuo, cuja função de autocorrelação decai hiperbolicamente, isto é, apresenta longa dependência. Em seguida, Mandelbrot e Wallis (1969) define o *ruído gaussiano fracionário*, que é uma versão do Movimento Browniano Fracionário a tempo discreto, mostrando que este processo também exibe o efeito de Hurst.

Posteriormente, inúmeros processos foram propostos para modelar a longa dependência. Granger e Joyeux (1980), e Hosking (1981, 1984), independentemente, introduziram o modelo fracionário autorregressivo integrado de média móvel, denotado por ARFIMA(p, d, q), o qual descreve o comportamento de longa memória associado a um crescimento ilimitado da função densidade espectral quando a frequência tende a zero. Estes modelos generalizam os modelos ARMA e ARIMA, propostos em Box e Jenkins (1970), e na atualização Box et al. (1994).

Em recentes estudos sobre séries temporais, têm-se focado nos processos estocásticos com a característica de longa dependência. A propriedade de longa dependência para um processo

estocástico $\{X_t\}_{t \in \mathbb{Z}}$ pode ser definida de diferentes maneiras, ou no domínio do tempo, ou no domínio da frequência. Antes de definirmos os processos ARFIMA(p, d, q), introduziremos uma definição para a propriedade de *longa dependência*, característica principal desses modelos.

Definição 3.3. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico estacionário. No domínio do tempo, se existe um número real $u \in (0, 1)$ tal que:

$$\rho_x(k) \sim C_1 k^{-u}, \text{ quando } k \rightarrow \infty,$$

em que $C_1 \neq 0$ e $\rho_x(\cdot)$ é a função de autocorrelação do processo, então $\{X_t\}_{t \in \mathbb{Z}}$ possui *longa dependência*. Equivalentemente, no domínio da frequência, se existe um número real $b \in (0, 1)$ tal que:

$$f_x(w) \sim C_2 |w|^{-b}, \text{ quando } w \rightarrow 0,$$

em que $C_2 > 0$ e $f_x(\cdot)$ é a função densidade espectral do processo, então $\{X_t\}_{t \in \mathbb{Z}}$ possui *longa dependência* (ver Bary, 1964).

Notação: Na Definição 3.3, a notação $f(w) \sim g(w)$, quando $w \rightarrow 0$, significa

$$\lim_{w \rightarrow 0} \frac{f(w)}{g(w)} = 1.$$

Para detalhes sobre ordens de aproximação, ver Seção 2.4, e para maiores detalhes sobre processos de longa dependência, ver Brockwell e Davis (2006), Palma (2007) e Morettin e Toloi (2004).

Observação 3.3. Em relação a definição 3.3 observamos:

- i) A relação $b = 1 - u$ é verdadeira;
- ii) No domínio do tempo, quando $u \in (0, 1)$, $\rho_x(k)$ tende a zero tão lentamente que $\sum_{k \in \mathbb{Z}_{\geq}} |\rho_x(k)|$ diverge, em que $\rho_x(\cdot)$ é a função de autocorrelação do processo;
- iii) O caso em que $u = 1$ é referido como *curta dependência* por alguns autores.

A seguir, apresentamos a propriedade de dependência intermediária.

Definição 3.4. Na Definição 3.3, se $u \in (1, 2)$ dizemos que $\{X_t\}_{t \in \mathbb{Z}}$ é um processo estocástico estacionário com *dependência intermediária*.

Observação 3.4. A Definição 3.3 para a propriedade de *longa dependência* não é a única possível. Definições alternativas podem ser encontradas na literatura e elas não são equivalentes. No contexto de processos estocásticos estacionários $\{X_t\}_{t \in \mathbb{Z}}$, com variância finita, os itens a seguir são definições comuns de *longa dependência*:

- i) $\sum_{k=-n}^n \gamma_x(k) \sim n^\alpha L_1(n)$, quando $n \rightarrow \infty$ e $0 < \alpha < 1$;
- ii) $\gamma_x(k) \sim k^{-\beta} L_2(k)$, quando $k \rightarrow \infty$ e $0 < \beta < 1$;

iii) $f_x(w) \sim w^{-\delta} L_3(|w|)$, quando $w \rightarrow 0$ e $0 < \delta < 1$.

As funções $L_1(\cdot)$, $L_2(\cdot)$ são funções de variação suave no infinito (ver Observação 3.5, ou Observação A.1 em Bisognin (2007)) enquanto que a função $L_3(\cdot)$ é de variação suave em zero. Para maiores detalhes, ver Doukhan et al. (2003).

Observação 3.5. Uma função é de *variação suave* em zero (respectivamente, infinito) se ela é limitada em um intervalo finito e se, para todo $x > 0$, $g(tx)/g(t) \rightarrow 1$, quando $t \rightarrow 0$ (respectivamente, infinito). As funções constante e logarítmica são exemplos de funções de variação suave.

Utilizando a definição de *longa dependência*, Granger e Joyeux (1980), Hosking (1981, 1984) e Geweke e Porter-Hudak (1983) apresentam os *processos autorregressivos fracionariamente integrados de média móvel* (denotados por ARFIMA(p, d, q)) como um exemplo de processos com a característica de *longa dependência*. A seguir, definimos os processos ARFIMA(p, d, q), que são uma generalização dos processos ARIMA(p, d, q) em que o parâmetro d pode assumir qualquer valor real.

Definição 3.5. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico satisfazendo a equação:

$$\phi(\mathcal{B})(1 - \mathcal{B})^d(X_t - \mu) = \theta(\mathcal{B})\varepsilon_t, \quad t \in \mathbb{Z}, \quad (3.7)$$

em que μ é a média do processo, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um processo ruído branco (ver definição 2.37), \mathcal{B} é o operador de *defasagem* ou de *retardo* (ver definição 2.38), $\phi(\cdot)$ e $\theta(\cdot)$ são os polinômios de ordem p e q , respectivamente, definidos por:

$$\phi(z) = \sum_{\ell=0}^p (-\phi_\ell) z^\ell \quad \text{e} \quad \theta(z) = \sum_{m=0}^q \theta_m z^m, \quad (3.8)$$

em que ϕ_ℓ , $1 \leq \ell \leq p$, e θ_m , $1 \leq m \leq q$, são constantes reais tais que $\phi_0 = -1$, e $\theta_0 = 1$.

Então, $\{X_t\}_{t \in \mathbb{Z}}$ é um *processo autorregressivo fracionariamente integrado de média móvel de ordem* (p, d, q), com média μ , denotado por ARFIMA(p, d, q), em que d é o *grau de diferenciação*.

Observação 3.6. Na Definição 3.5, temos que um ARFIMA(p, d, q), com $d = 0$, é um processo ARMA(p, q), chamado *processo autorregressivo de média móvel de ordem* (p, q). A função de autocorrelação dos processos ARMA(p, q) possui decaimento exponencial, isto é, decresce ligeiramente para zero. Esse fato implica que observações afastadas têm uma insignificante correlação entre elas, por isso esse modelo é dito de "curta memória" ou de curta dependência.

Teorema 3.2. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo ARFIMA(p, d, q) dado pela expressão (3.7). Então:

- a) $\{X_t\}_{t \in \mathbb{Z}}$ é estacionário se $d < \frac{1}{2}$ e todas as raízes da equação $\phi(z) = 0$ estão fora do círculo unitário;
- b) $\{X_t\}_{t \in \mathbb{Z}}$ é invertível se $d > -\frac{1}{2}$ e todas as raízes da equação $\theta(z) = 0$ estão fora do círculo unitário.

Teorema 3.3. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo ARFIMA(p, d, q) definido pela expressão (3.7) tal que $d \in (-\frac{1}{2}, \frac{1}{2})$. Suponhamos que as equações $\phi(z) = 0$ e $\theta(z) = 0$ não têm raízes em comum e que as suas raízes estão fora do círculo unitário. Então:

- a) $\{X_t\}_{t \in \mathbb{Z}}$ é estacionário e invertível, com representações autorregressiva e média móvel dadas respectivamente por:

$$\varepsilon_t = \sum_{k \in \mathbb{Z}_{\geq}} \pi_k X_{t-k} \quad \text{e} \quad X_t = \sum_{k \in \mathbb{Z}_{\geq}} \psi_k \varepsilon_{t-k},$$

em que $\{\pi_k\}_{k \in \mathbb{Z}_{\geq}}$ e $\{\psi_k\}_{k \in \mathbb{Z}_{\geq}}$ são os coeficientes de \mathcal{B}^k na expansão de:

$$\pi(\mathcal{B}) = \frac{\phi(\mathcal{B})}{\theta(\mathcal{B})}(1 - \mathcal{B})^d \quad \text{e} \quad \psi(\mathcal{B}) = \frac{\theta(\mathcal{B})}{\phi(\mathcal{B})}(1 - \mathcal{B})^{-d},$$

respectivamente.

- b) Considerando $\{X_t\}_{t \in \mathbb{Z}}$ estacionário e invertível e tomando $Y_t = (1 - \mathcal{B})^d X_t$, temos que o processo $\{Y_t\}_{t \in \mathbb{Z}}$, dado por $\phi(\mathcal{B})Y_t = \theta(\mathcal{B})\varepsilon_t$, é um processo ARMA(p, q) com função densidade espectral $f_Y(\cdot)$.

Tomando $U_t = \frac{\psi(\mathcal{B})}{\theta(\mathcal{B})}X_t$, obtemos que $\{U_t\}_{t \in \mathbb{Z}}$ é um processo ARFIMA(p, d, q), com $p = 0 = q$, e com função densidade espectral $f_U(\cdot)$.

Dadas essas considerações temos, então:

- i) A função densidade espectral do processo $\{X_t\}_{t \in \mathbb{Z}}$, denotada por $f_X(\cdot)$, é dada pelo produto das funções densidade espectral dos correspondentes processo ARMA(p, q) e processo ARFIMA(p, d, q), com $p = 0 = q$, isto é:

$$f_X(w) = f_Y(w)f_U(w) = \frac{\sigma_\varepsilon^2}{2\pi} \left| \frac{\theta(e^{-iw})}{\phi(e^{-iw})} \right|^2 \left[2\text{sen}\left(\frac{w}{2}\right) \right]^{-2d}, \quad (3.9)$$

para todo $w \in (0, \pi]$.

Note que o $\lim_{w \rightarrow 0} w^{2d} f_X(w)$ existe e é finito.

- ii) Seja $\rho_X(\cdot)$ a função de autocorrelação do processo $\{X_t\}_{t \in \mathbb{Z}}$. Então:

$$\lim_{k \rightarrow +\infty} k^{1-2d} \rho_X(k) \quad \text{existe e é finito.}$$

- iii) Seja $\gamma_X(\cdot)$ a função de autocovariância do processo $\{X_t\}_{t \in \mathbb{Z}}$. Então:

$$\gamma_X(h) \sim c_\gamma |h|^{2d-1}, \quad \text{quando } |h| \rightarrow \infty,$$

$$\text{em que } c_\gamma = \frac{\sigma^2 |\theta(1)|^2}{\pi |\phi(1)|^2} \Gamma(1 - 2d) \text{sen}(\pi d).$$

Como consequência dos Teoremas 3.2 e 3.3, e também das Definições de longa e intermediária dependência 3.3 e 3.4, enunciamos o seguinte resultado:

Corolário 3.1. Sejam $\{X_t\}_{t \in \mathbb{Z}}$ um processo ARFIMA(p, d, q) definido pela expressão (3.7). Suponhamos que $\phi(z) = 0$ e $\theta(z) = 0$ não têm raízes em comum e que as suas raízes estão fora do círculo unitário. Então:

- a) Se $d \in (0, \frac{1}{2})$, Então $\{X_t\}_{t \in \mathbb{Z}}$ é processo estacionário e invertível de longa dependência.
- b) Se $d \in (-\frac{1}{2}, 0)$, Então $\{X_t\}_{t \in \mathbb{Z}}$ é processo estacionário e invertível de intermediária dependência.
- c) Se $d = 0$, Então $\{X_t\}_{t \in \mathbb{Z}}$ é processo estacionário e invertível de curta dependência.

Note que um processo ARFIMA(p, d, q), com $d = 0$, é um processo ARMA(p, q) cuja função de autocorrelação tem decaimento exponencial.

O próximo Teorema examina a existência e unicidade da solução estacionária do processo ARFIMA(p, d, q) dado pela equação (3.7), juntamente com a causalidade e invertibilidade. Sua demonstração é encontrada em Palma (2007), pág.44.

Teorema 3.4. Considere o processo ARFIMA(p, d, q) definido pela equação (3.7). Assumimos que os polinômios $\phi(\cdot)$ e $\theta(\cdot)$ não têm raízes em comum e que $d \in (-1, \frac{1}{2})$. Então:

- a) Se as raízes de $\phi(\cdot)$ estão fora do círculo unitário $\{z : |z| = 1\}$, então existe e é única a solução estacionária de (3.7) dada por:

$$X_t = \sum_{j=-\infty}^{+\infty} \psi_j \varepsilon_{t-j}, \quad \text{em que } \psi(z) = (1-z)^{-d} \frac{\theta(z)}{\phi(z)}.$$

- b) Se as raízes de $\phi(\cdot)$ estão fora do disco unitário fechado $\{z : |z| \leq 1\}$, então a solução $\{X_t\}_{t \in \mathbb{Z}}$ é causal.
- c) Se as raízes de $\theta(\cdot)$ estão fora do disco unitário fechado $\{z : |z| \leq 1\}$, então a solução $\{X_t\}_{t \in \mathbb{Z}}$ é invertível.

A seguir, apresentamos um quadro com os principais modelos para séries temporais pertinentes a este trabalho. O quadro menciona os mais importantes comportamentos desses processos no domínio do tempo (caracterizados pela função de autocorrelação, $\rho_X(\cdot)$) e no domínio da frequência (caracterizados pela função densidade espectral, $f_X(\cdot)$).

Modelos	Domínio do Tempo	Domínio da Frequência
AR(p)	$\rho_x(\cdot)$ decai exponencialmente (rapidamente) e/ou , senoidalmente amortecida, e é infinita em extensão;	$f_x(\cdot)$ é limitada e periódica;
MA(q)	$\rho_x(\cdot)$ é finita, no sentido de apresentar um corte após o "lag" q ;	$f_x(\cdot)$ é limitada e periódica;
ARMA(p, q)	$\rho_x(\cdot)$ decai exponencialmente e/ou senoidalmente amortecida após o "lag" $q - p$ e é infinita em extensão;	$f_x(\cdot)$ possui decaimento exponencial, e é geometricamente limitada;
ARFIMA(p, d, q)	$\rho_x(\cdot)$ decai hiperbolicamente (suavemente) para zero, quando $d \in (0, \frac{1}{2})$;	$f_x(\cdot)$ é ilimitada (tende ao infinito) quando as frequências se aproximam de zero.

Maiores detalhes sobre os processos ARFIMA(p, d, q) podem ser encontrados em Fox e Taqu (1986), Brockwell e Davis (2006), Sowell (1992), Beran (1994), Robinson (1995a), Baillie (1996), Peiris e Singh (1996), Reisen e Lopes (1999), Lopes et al. (2004), Lopes (2007) entre outros.

Capítulo 4

Processos Fracionários Generalizados

Os processos ARFIMA(p, d, q) são utilizados por um grande número de trabalhos, porém estes processos não são suficientes para resolver todos os problemas sobre longa dependência. Deste modo, é necessária alguma generalização que possa resolver uma quantidade maior desses casos.

Um exemplo de generalização sugerido por Hosking (1981), é o modelo:

$$\phi(\mathcal{B})(1 - 2u\mathcal{B} + \mathcal{B}^2)^\lambda(X_t - \mu) = \theta(\mathcal{B})\varepsilon_t, \quad \text{para todo } t \in \mathbb{Z}, \quad (4.1)$$

no qual estão incluídos os processos ARFIMA(p, d, q), pois, quando $u = 1$, temos que $d = 2\lambda$.

Este modelo somente foi estudado recentemente, devido à complexidade na inversão do fator $(1 - 2u\mathcal{B} + \mathcal{B}^2)^\lambda$.

Inicialmente, ele foi investigado por Gray et al. (1989) que utilizou as funções geradoras dos polinômios Gegenbauer para sua definição (Ver Definição 4.1). Os processos indicados por Hosking, dados pela expressão (4.1), como dito anteriormente, são uma generalização dos processos ARFIMA(p, d, q), em que a função densidade espectral é ilimitada para alguma frequência do intervalo $[0, \pi]$, enquanto que no modelo ARFIMA(p, d, q) ela é ilimitada somente na origem.

A seguir, definiremos os polinômios Gegenbauer os quais são de grande importância para a definição dos processos GARMA e k -Factor GARMA, modelos de interesse para este trabalho.

4.1 Polinômios Gegenbauer

Os polinômios de Gegenbauer são muito aplicados na Matemática, tanto por sua ortogonalidade, quanto por suas propriedades recursivas. A seguir, definiremos os polinômios Gegenbauer os quais são de grande importância para a definição dos processos GARMA e k -Factor GARMA, modelos de interesse para este trabalho.

Definição 4.1. Os *polinômios Gegenbauer*, denotados por $C_j^{(\lambda)}(u)$, são definidos como os coeficientes na expansão em série de potência da seguinte função:

$$(1 - 2uZ + Z^2)^{-\lambda} = \sum_{j \in \mathbb{Z}_\geq} C_j^{(\lambda)}(u)Z^j, \quad (4.2)$$

para $\lambda \neq 0$, $|u| \leq 1$ e $|Z| < 1$, em que:

$$C_j^{(\lambda)}(u) = \sum_{k=0}^{\lfloor j/2 \rfloor} \frac{(-1)^k \Gamma(\lambda - k + j) (2u)^{j-2k}}{\Gamma(\lambda) \Gamma(k+1) \Gamma(j-2k+1)}, \quad (4.3)$$

para todo $j \geq 0$, com $\lfloor x \rfloor$ sendo a parte inteira de x , em que $\Gamma(\cdot)$ é a função Gama definida por:

- i) $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \quad x > 0;$
- ii) $\Gamma(x) = \infty, \quad x = 0;$
- iii) $\Gamma(x) = x^{-1} \Gamma(x+1), \quad x < 0;$
- iv) $\Gamma(x) \Gamma(1-x) = \frac{\pi}{\text{sen}(\pi x)}.$

Os polinômios Gegenbauer podem ser aproximados por:

$$C_j^{(\lambda)}(u) \sim \frac{\cos[(j+\lambda)G - (\lambda\pi/2)]}{\Gamma(\lambda) \text{sen}^\lambda(G)} \left(\frac{2}{j}\right)^{1-\lambda}, \quad \text{quando } j \rightarrow \infty,$$

em que a constante G é dada por $G = \cos^{-1}(u)$.

Assim, quando $\lambda < 1$, $C_j^{(\lambda)}(\cdot)$ decresce numa taxa hiperbólica, quando $j \rightarrow \infty$.

Computacionalmente, podemos calcular $C_j^{(\lambda)}(\cdot)$ usando a seguinte fórmula recursiva:

$$C_j^{(\lambda)}(u) = 2u \left(\frac{\lambda-1}{j} + 1\right) C_{j-1}^{(\lambda)}(u) - \left(2\frac{\lambda-1}{j} + 1\right) C_{j-2}^{(\lambda)}(u),$$

para todo $j > 2$, com $C_0^{(\lambda)}(u) = 1$, $C_1^{(\lambda)}(u) = 2\lambda u$, $C_2^{(\lambda)}(u) = 2\lambda(\lambda+1)u^2 - \lambda$.

Tendo definido os polinômios Gegenbauer, podemos definir os processos Gegenbauer e, conseqüentemente, os processos GARMA.

4.2 Processos GARMA

Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico dado pela expressão:

$$(1 - 2u\mathcal{B} + \mathcal{B}^2)^\lambda (X_t - \mu) = \varepsilon_t, \quad t \in \mathbb{Z}, \quad (4.4)$$

em que $|u| \leq 1$, μ é a média do processo, \mathcal{B} é o operador de defasagem ou de retardo, isto é, $\mathcal{B}^j(X_t) = X_{t-j}$, para todo $j \in \mathbb{N}$ e $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um processo ruído branco.

Se o fator $(1 - 2u\mathcal{B} + \mathcal{B}^2)^\lambda$ é invertível, então podemos escrever formalmente:

$$X_t = \mu + (1 - 2u\mathcal{B} + \mathcal{B}^2)^{-\lambda} \varepsilon_t, \quad \text{para todo } t \in \mathbb{Z}.$$

Utilizando-se a equação (4.2), o processo $\{X_t\}_{t \in \mathbb{Z}}$ pode ser escrito na sua representação média móvel infinita:

$$X_t = \mu + \sum_{j \in \mathbb{Z}_\geq} C_j^{(\lambda)}(u) \varepsilon_{t-j}, \quad (4.5)$$

em que $(C_j^{(\lambda)}(\cdot))_{j \in \mathbb{Z}}$ são os polinômios Gegenbauer definidos na equação (4.3). Assim, temos a seguinte definição.

Definição 4.2. O processo $\{X_t\}_{t \in \mathbb{Z}}$, dado pela equação (4.5), é chamado de *processo Gegenbauer* com parâmetros u e λ e pode ser escrito pela expressão dada em (4.4).

Observação 4.1. Podemos notar que se $u = 1$, em (4.4), temos:

$$(1 - \mathcal{B})^{2\lambda}(X_t - \mu) = \varepsilon_t. \quad (4.6)$$

Neste caso, $\{X_t\}_{t \in \mathbb{Z}}$ é justamente o processo fracionariamente integrado de grau ou parâmetro de diferenciação 2λ , isto é, um processo ARFIMA(p, d, q), com $p = 0$, $d = 2\lambda$ e $q = 0$, e suas propriedades podem ser encontradas em Hosking (1981).

Quando $u = -1$, a equação (4.4) torna-se:

$$(1 + \mathcal{B})^{2\lambda}(X_t - \mu) = \varepsilon_t.$$

A seguir, citamos algumas propriedades dos processos Gegenbauer cujas demonstrações podem ser encontradas em Gray et al. (1989) e Chung (1996).

Observação 4.2. As Definições 3.3 e 3.4 de longa dependência e intermediária dependência, respectivamente, podem ser estendidas para um processo estacionário qualquer da seguinte maneira:

- i) Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estacionário para o qual existe um número real $b \in (0, 1)$, uma constante $C_f > 0$ e uma frequência $G \in [0, \pi]$ (ou um número finito de frequências) tal que:

$$f_x(w) \sim C_f |w - G|^{-b}, \quad \text{quando } w \rightarrow G.$$

Então, $\{X_t\}_{t \in \mathbb{Z}}$ é chamado processo estacionário com *longa dependência*.

- ii) Quando $b \in (-1, 0)$, dizemos que o processo possui *dependência intermediária*.
iii) No item i), quando $\{X_t\}_{t \in \mathbb{Z}}$ é um processo ARFIMA(p, d, q), com $p = 0 = q$, temos que $b = 2d$ e $G = 0$. Assim, os processos ARFIMA(p, d, q), com $p = 0 = q$, possuem longa dependência quando $d \in (0, 0.5)$.

Teorema 4.1. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo Gegenbauer dado pela Definição 4.2. Então:

- a) O processo $\{X_t\}_{t \in \mathbb{Z}}$ é *estacionário* se:

- i) $|u| < 1$ e $\lambda < \frac{1}{2}$, ou
ii) $|u| = 1$ e $\lambda < \frac{1}{4}$;

- b) O processo $\{X_t\}_{t \in \mathbb{Z}}$ é *invertível* se:

- i) $|u| < 1$ e $\lambda > -\frac{1}{2}$, ou
ii) $|u| = 1$ e $\lambda > -\frac{1}{4}$;

- c) Um processo Gegenbauer estacionário possui a propriedade de *longa dependência* se:

- i) $|u| < 1$ e $0 < \lambda < \frac{1}{2}$, ou

ii) $|u| = 1$ e $0 < \lambda < \frac{1}{4}$;

d) A função densidade espectral do processo $\{X_t\}_{t \in \mathbb{Z}}$ é dada por:

$$\begin{aligned} f_x(w) &= \frac{\sigma_\varepsilon^2}{2\pi} |1 - 2ue^{iw} + e^{2iw}|^{-2\lambda} \\ &= \frac{\sigma_\varepsilon^2}{2\pi} [2(\cos(w) - u)]^{-2\lambda}, \quad \text{para todo } w \in (0, \pi], \end{aligned} \quad (4.7)$$

em que $G = \cos^{-1}(u)$ é chamada *frequência de Gegenbauer* ou *frequência G* , isto é, a função densidade espectral torna-se ilimitada em G .

Quando $w \rightarrow G$,

$$f_x(w) \sim C_f |w - G|^{-2\lambda}, \quad (4.8)$$

em que $C_f = \frac{\sigma_\varepsilon^2}{2\pi} [2|\text{sen}(G)|]^{-2\lambda} > 0$.

e) A função de autocovariância do processo $\{X_t\}_{t \in \mathbb{Z}}$, quando $|u| < 1$ e $k \in \mathbb{Z}_{\geq}$, é dada por:

$$\gamma_x(k) = \frac{\sigma_\varepsilon^2}{2\sqrt{\pi}} \Gamma(1 - 2\lambda) [2\text{sen}(G)]^{1/2 - 2\lambda} [P_{k-1/2}^{2\lambda-1/2}(u) + (-1)^k P_{k-1/2}^{2\lambda-1/2}(-u)],$$

em que $P_a^b(\cdot)$ são funções de Legendre (ver Observação 4.4) e pode ser aproximada por:

$$\begin{aligned} \gamma_x(k) &= \frac{2^{1-2\lambda} \sigma_\varepsilon^2}{\pi} \text{sen}^{-2\lambda}(G) \text{sen}(\lambda\pi) \Gamma(1 - 2\lambda) \cos(kG) \\ &\quad \times \frac{\Gamma(k + 2\lambda)}{\Gamma(k + 1)} [1 + O(k^{-1})]. \end{aligned}$$

f) Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo Gegenbauer estacionário com longa dependência. Então:

i) Quando $u = 1$ e $0 < \lambda < \frac{1}{4}$, a *função de autocorrelação* do processo $\{X_t\}_{t \in \mathbb{Z}}$ é dada por:

$$\rho_x(k) = \frac{\Gamma(1 - 2\lambda)\Gamma(k + 2\lambda)}{\Gamma(2\lambda)\Gamma(k - 2\lambda + 1)}, \quad k \in \mathbb{Z}_{\geq},$$

em que $\rho_x(k) \sim k^{4\lambda-1}$, quando $k \rightarrow \infty$.

ii) Quando $u = -1$ e $0 < \lambda < \frac{1}{4}$, a *função de autocorrelação* do processo $\{X_t\}_{t \in \mathbb{Z}}$ é dada por:

$$\rho_x(k) = (-1)^k \frac{\Gamma(1 - 2\lambda)\Gamma(k + 2\lambda)}{\Gamma(2\lambda)\Gamma(k - 2\lambda + 1)}, \quad k \in \mathbb{Z}_{\geq},$$

em que $\rho_x(k) \sim (-1)^k k^{4\lambda-1}$, quando $k \rightarrow \infty$.

iii) Quando $|u| < 1$ e $0 < \lambda < \frac{1}{2}$,

$$\rho_x(k) \sim k^{2\lambda-1} \text{sen}(\pi\lambda - kG), \text{ quando } k \rightarrow \infty, \quad (4.9)$$

em que $G = \cos^{-1}(u)$.

Observação 4.3. A expressão (4.9), acima, explica o comportamento senoidal da função de autocorrelação dos processos Gegenbauer.

Observação 4.4. As *funções de Legendre*, associadas a expressão exata da função de autocovariância do processo $\{X_t\}_{t \in \mathbb{Z}}$ dada no item e) da Proposição (4.1), podem ser calculadas utilizando a fórmula recursiva abaixo:

$$P_a^b(x) = \frac{2a-1}{a-b} x P_{a-1}^b(x) - \frac{a+b-1}{a-b} P_{a-2}^b(x). \quad (4.10)$$

A fórmula recursiva (4.10) requer os seguintes termos iniciais:

$$P_{-1/2}^{2\lambda-1/2}(u) = \left(\frac{1+u}{1-u} \right)^{\lambda-1/4} \frac{1}{\Gamma\left(\frac{3}{2} - 2\lambda\right)} \cdot F\left(\frac{1}{2}, \frac{1}{2}; \frac{3}{2} - 2\lambda; \frac{1-u}{2}\right),$$

$$P_{1/2}^{2\lambda-1/2}(u) = \left(\frac{1+u}{1-u} \right)^{\lambda-1/4} \frac{1}{\Gamma\left(\frac{3}{2} - 2\lambda\right)} \cdot F\left(-\frac{1}{2}, \frac{3}{2}; \frac{3}{2} - 2\lambda; \frac{1-u}{2}\right),$$

em que $F(\cdot, \cdot; \cdot; \cdot)$ é a *função hipergeométrica* definida por:

$$F(a, b; c; x) = \sum_{j \geq 1} \left[\frac{\Gamma(c)\Gamma(a+j)\Gamma(b+j)}{\Gamma(a)\Gamma(b)\Gamma(c+j)\Gamma(j+1)} \right] x^j.$$

Observação 4.5. Woodward et al. (1998), baseado na equação (4.5), apresenta a *função de autocovariância* dada por:

$$\gamma_x(k) = \sigma_\varepsilon^2 \sum_{j \in \mathbb{Z}_{\geq}} C_j^{(\lambda)}(u) C_{j+k}^{(\lambda)}(u), \text{ para todo } |u| \leq 1, \quad (4.11)$$

em que $k \in \mathbb{Z}_{\geq}$ e $C_j^{(\lambda)}(\cdot)$ é dado pela expressão (4.3).

Da mesma forma que podemos incluir componentes autorregressivas e médias móveis aos processos fracionalmente integrados puros, obtendo então os processos ARFIMA(p, d, q), podemos estender os processos Gegenbauer combinando-os com os processos ARMA(p, q). Isto nos conduz a seguinte definição.

Definição 4.3. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico que satisfaz a equação:

$$\phi(\mathcal{B})(1 - 2u\mathcal{B} + \mathcal{B}^2)^\lambda (X_t - \mu) = \theta(\mathcal{B})\varepsilon_t, \quad (4.12)$$

em que μ é a média do processo, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um processo ruído branco, $\phi(\cdot)$ e $\theta(\cdot)$ são os polinômios de graus p e q , respectivamente, definidos por:

$$\phi(z) = \sum_{\ell=0}^p (-\phi_\ell) z^\ell \quad \text{e} \quad \theta(z) = \sum_{m=0}^q \theta_m z^m, \quad (4.13)$$

em que ϕ_ℓ , $1 \leq \ell \leq p$, e θ_m , $1 \leq m \leq q$, são constantes reais tais que $\phi_0 = -1$, e $\theta_0 = 1$.

Então, $\{X_t\}_{t \in \mathbb{Z}}$ é dito um *processo autorregressivo de média móvel Gegenbauer* de ordem (p, u, λ, q) , denotado por $\text{GARMA}(p, u, \lambda, q)$.

As propriedades dos processos $\text{GARMA}(p, u, \lambda, q)$, a seguir apresentadas, são demonstradas em Gray et al. (1989).

Proposição 4.1. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo $\text{GARMA}(p, u, \lambda, q)$, ($\lambda \neq 0$), dado na Definição 4.3, com todas as raízes das equações $\phi(z) = 0$ e $\theta(z) = 0$ fora do círculo unitário. Então, valem as seguintes afirmações:

- a) O processo $\{X_t\}_{t \in \mathbb{Z}}$ é *estacionário* se:
 - i) $|u| < 1$ e $\lambda < \frac{1}{2}$, ou
 - ii) $|u| = 1$ e $\lambda < \frac{1}{4}$.
- b) O processo $\{X_t\}_{t \in \mathbb{Z}}$ é *invertível* se:
 - i) $|u| < 1$ e $\lambda > -\frac{1}{2}$, ou
 - ii) $|u| = 1$ e $\lambda > -\frac{1}{4}$.
- c) $\{X_t\}_{t \in \mathbb{Z}}$ é um processo com *longa dependência* se:
 - i) $|u| < 1$ e $0 < \lambda < \frac{1}{2}$, ou
 - ii) $|u| = 1$ e $0 < \lambda < \frac{1}{4}$.
- d) A *função densidade espectral* do processo $\{X_t\}_{t \in \mathbb{Z}}$, denotada por $f_X(\cdot)$, é dada por:

$$f_X(w) = \frac{\sigma_\varepsilon^2 |\theta(e^{-iw})|^2}{2\pi |\phi(e^{-iw})|^2} [2(\cos(w) - u)]^{-2\lambda}, \quad \text{para todo } w \in (0, \pi], \quad (4.14)$$

em que $G = \cos^{-1}(u)$ é a frequência de Gegenbauer.

Podemos notar que:

- i) Se $|u| < 1$, $\lim_{w \rightarrow G} (w - G)^{2\lambda} f_X(w)$ existe e é finito.
 - ii) Se $|u| = 1$, $\lim_{w \rightarrow G} (w - G)^{4\lambda} f_X(w)$ existe e é finito.
- e) Quando $k \rightarrow \infty$,
- i) $\rho_X(k) \sim k^{4\lambda-1}$, quando $u = 1$ e $0 < \lambda < \frac{1}{4}$;
 - ii) $\rho_X(k) \sim (-1)^k k^{4\lambda-1}$, quando $u = -1$ e $0 < \lambda < \frac{1}{4}$;
 - iii) $\rho_X(k) \sim k^{2\lambda-1} \text{sen}(\pi\lambda - kG)$, quando $|u| < 1$ e $0 < \lambda < \frac{1}{2}$.

Observação 4.6.

- i) A função densidade espectral dos processos $\text{GARMA}(p, u, \lambda, q)$ não é necessariamente ilimitada na origem, como nos processos $\text{ARFIMA}(p, d, q)$, mas para alguma frequência G do intervalo $[0, \pi]$.

ii) Estamos interessados em encontrar a expressão assintótica da função densidade espectral dos processos GARMA(p, u, λ, q). Para isso, escrevemos a expressão (4.14) como:

$$f_X(w) = f_Y(w) \frac{|\theta(e^{-iw})|^2}{|\phi(e^{-iw})|^2}, \quad (4.15)$$

em que $f_Y(\cdot)$ é a função densidade espectral de um processo GARMA (p, u, λ, q), com $p = 0 = q$, dada pela expressão (4.7), com $G = \cos^{-1}(u)$.

Vamos supor que o processo GARMA(p, u, λ, q) seja causal e invertível. Sabemos que $\lim_{z \rightarrow 0} \cos(z) = 1$, ou seja, $\cos(z) \sim 1$, quando $z \rightarrow 0$ (ver Bisognin (2007), definição A.5). Pelas expressões (4.15) e pela equação de Sowell (ver Bisognin (2007), equação (B.5), ou Sowell (1992)), quando $z \rightarrow 0$, temos que:

$$\begin{aligned} f_X(z + G) &= f_Y(z + G) \frac{|\theta(e^{-i(z+G)})|^2}{|\phi(e^{-i(z+G)})|^2} \\ &= f_Y(z + G) \left[\frac{\prod_{m=1}^q (1 - 2\rho_{m,1} \cos(z + G) + \rho_{m,1}^2)}{\prod_{\ell=1}^p (1 - 2\rho_{\ell,2} \cos(z + G) + \rho_{\ell,2}^2)} \right] \\ &\sim \frac{\sigma_\varepsilon^2}{2\pi} [2|\text{sen}(G)|]^{-2\lambda} |z|^{-2\lambda} \\ &\quad \times \left[\frac{\prod_{m=1}^q (1 - 2\rho_{m,1} \cos(G) + \rho_{m,1}^2)}{\prod_{\ell=1}^p (1 - 2\rho_{\ell,2} \cos(G) + \rho_{\ell,2}^2)} \right] \end{aligned} \quad (4.16)$$

$$\begin{aligned} &= \frac{\sigma_\varepsilon^2}{2\pi} [2|\text{sen}(G)|]^{-2\lambda} |z|^{-2\lambda} \frac{|\theta(e^{-iG})|^2}{|\phi(e^{-iG})|^2} \\ &= C_f |z|^{-2\lambda}, \end{aligned} \quad (4.17)$$

em que $G = \cos^{-1}(u)$, e além disso:

$$C_f = \frac{\sigma_\varepsilon^2}{2\pi} [2|\text{sen}(G)|]^{-2\lambda} \frac{|\theta(e^{-iG})|^2}{|\phi(e^{-iG})|^2}. \quad (4.18)$$

A expressão (4.16) é validada pela equação (4.8).

Fazendo a mudança de variável $w = z + G$, na equação (4.17), temos que:

$$f_X(w) \sim C_f |w - G|^{-2\lambda}, \quad (4.19)$$

quando $w \rightarrow G$, onde C_f é dada pela expressão (4.18).

Para maiores detalhes sobre os processos GARMA(p, u, λ, q), ver Chung (1996) e Ferrara e Guégan (1999).

4.3 Processos k -Factor GARMA

Os processos ARFIMA(p, d, q), em que $d \in (-0.5, 0.5)$, podem ser tratados como uma generalização dos processos ARIMA(p, d, q), em que $d \in \mathbb{N}$, para modelar dados com a propriedade de longa dependência, isto é, quando a função densidade espectral é ilimitada na frequência zero. Similarmente, os processos GARMA(p, u, λ, q) são tratados como uma generalização dos processos ARFIMA(p, d, q) cuja função densidade espectral torna-se ilimitada em alguma frequência G do intervalo $[0, \pi]$, não necessariamente a frequência zero. Entretanto, uma limitação dos processos ARFIMA(p, d, q) e do processo mais geral GARMA(p, u, λ, q) é que as suas funções densidade espectral tornam-se ilimitadas em apenas uma frequência do intervalo $[0, \pi]$. Por esse motivo, Gray et al. (1989) sugere a inclusão de mais de um fator Gegenbauer nos modelos GARMA.

Giraitis e Leipus (1995) e, depois, Woodward et al. (1998) estendem os modelos Gegenbauer e GARMA, respectivamente, aos modelos k -Factor Gegenbauer e k -Factor GARMA, para os quais a função densidade espectral é ilimitada para um número finito k de frequências, chamadas de *frequências de Gegenbauer* (ou frequências G), do intervalo $[0, \pi]$.

Definição 4.4. Seja $\{X_t\}_{t \in \mathbb{Z}}$ o processo estocástico que satisfaz a equação:

$$\prod_{j=1}^k (1 - 2u_j \mathcal{B} + \mathcal{B}^2)^{\lambda_j} (X_t - \mu) = \varepsilon_t, \quad (4.20)$$

em que k é um inteiro finito, $|u_j| \leq 1$ e λ_j é um número fracionário, para $j = 1, \dots, k$, μ é a média do processo e $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um processo ruído branco. Então, $\{X_t\}_{t \in \mathbb{Z}}$ é um *processo k -Factor Gegenbauer* de ordem $(p, \mathbf{u}, \boldsymbol{\lambda}, q)$, com $p = 0 = q$, denotado por k -Factor Gegenbauer($p, \mathbf{u}, \boldsymbol{\lambda}, q$), com $p = 0 = q$, em que $\mathbf{u} = (u_1, \dots, u_k)^T$ e $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)^T$.

Observação 4.7. As Definições 2.39 e 2.40 de causalidade e invertibilidade, respectivamente, também são válidas para os processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), os quais veremos mais adiante.

A seguir, apresentamos algumas propriedades dos processos k -Factor Gegenbauer($p, \mathbf{u}, \boldsymbol{\lambda}, q$), com $p = 0 = q$. As demonstrações dessas propriedades podem ser encontradas em Giraitis e Leipus (1995), e em Woodward et al. (1998).

Teorema 4.2. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo k -Factor Gegenbauer($p, \mathbf{u}, \boldsymbol{\lambda}, q$), com $p = 0 = q$ (ver Definição 4.4). Então:

i) O processo $\{X_t\}_{t \in \mathbb{Z}}$ é *estacionário* se u_j são distintos e, ainda:

$$\lambda_j < \begin{cases} \frac{1}{2}, & \text{quando } |u_j| < 1; \\ \frac{1}{4}, & \text{quando } |u_j| = 1, \end{cases} \quad (4.21)$$

para $j = 1, \dots, k$;

ii) O processo $\{X_t\}_{t \in \mathbb{Z}}$ é *estacionário e tem longa dependência* se u_j são distintos, e ainda:

$$\begin{cases} 0 < \lambda_j < \frac{1}{2}, & \text{quando } |u_j| < 1; \\ 0 < \lambda_j < \frac{1}{4}, & \text{quando } |u_j| = 1, \end{cases} \quad (4.22)$$

para $j = 1, \dots, k$;

iii) Seja

$$|\lambda_j| < \begin{cases} \frac{1}{2}, & \text{se } |u_j| < 1; \\ \frac{1}{4}, & \text{se } |u_j| = 1, \end{cases} \quad (4.23)$$

com $\lambda_j \neq 0$, para $j = 1, \dots, k$. Então, existe uma única solução estacionária, X_t de (4.20) a qual é causal e invertível e tem representação média móvel:

$$X_t = \sum_{n \in \mathbb{Z}_{\geq}} \psi_n \varepsilon_{t-n} \equiv \prod_{j=1}^k (1 - 2u_j \mathcal{B} + \mathcal{B}^2)^{-\lambda_j} \varepsilon_t.$$

Os coeficientes de $\{\psi_n\}_{n \in \mathbb{Z}_{\geq}}$ são dados por:

$$\psi_n = \sum_{j_1=0}^n \sum_{j_2=0}^n \cdots \sum_{j_k=0}^n C_{j_1}^{\lambda_1}(u_1) C_{j_2}^{\lambda_2}(u_2) \cdots C_{j_k}^{\lambda_k}(u_k).$$

em que $j_1 + j_2 + \cdots + j_k = n$

Equivalentemente podemos escrever:

$$\psi_n = \sum_A C_{j_1}^{\lambda_1}(u_1) \cdots C_{j_k}^{\lambda_k}(u_k), \quad (4.24)$$

em que $A := \{0 \leq j_1, \dots, j_k \leq n; j_1 + \cdots + j_k = n\}$.

Os coeficientes $\{\psi_n\}_{n \in \mathbb{Z}_{\geq}}$ apresentam a expansão assintótica:

$$\begin{aligned} \psi_n &= 2 \sum_{\ell: 0 < G_\ell < \pi} D(\ell) \frac{\Gamma(n + \lambda_\ell)}{\Gamma(n + 1)\Gamma(\lambda_\ell)} \cos(nG_\ell + \nu_\ell) \\ &+ \sum_{\ell: G_\ell = 0 \text{ ou } \pi} D(\ell) \frac{\Gamma(n + 2\lambda_\ell)}{\Gamma(n + 1)\Gamma(2\lambda_\ell)} \cos(nG_\ell) + O(n^{-2 + \max \lambda_1^*, \dots, \lambda_k^*}), \end{aligned}$$

quando $n \rightarrow \infty$, em que $\ell : 1, \dots, k$, $G_j = \cos^{-1}(u_j)$, para $j = 1, \dots, k$,

$$\lambda_\ell^* = \begin{cases} \lambda_\ell, & 0 < G_\ell < \pi; \\ 2\lambda_\ell, & G_\ell = 0 \text{ ou } G_\ell = \pi; \end{cases} \quad (4.25)$$

$$\nu_\ell = G_\ell \sum_1^k \lambda_j - \pi \sum_1^{\ell-1} \lambda_j - \lambda_\ell \frac{\pi}{2}$$

e

$$D(\ell) = \begin{cases} |2\text{sen}(G_\ell)|^{-\lambda_\ell} \prod_{\substack{j=0 \\ j \neq \ell}}^k [2|\cos(G_\ell) - \cos(G_j)|]^{-\lambda_j}, & \text{para } 0 < G_\ell < \pi; \\ \prod_{\substack{j=0 \\ j \neq \ell}}^k [2|\cos(G_\ell) - \cos(G_j)|]^{-\lambda_j}, & \text{para } G_\ell = 0 \text{ ou } G_\ell = \pi. \end{cases} \quad (4.26)$$

iv) O processo $\{X_t\}_{t \in \mathbb{Z}}$ possui função densidade espectral dada por:

$$f_X(w) = \frac{\sigma_\varepsilon^2}{2\pi} \prod_{j=1}^k |2(\cos(w) - u_j)|^{-2\lambda_j}, \quad \text{para } w \in (0, \pi],$$

em que $f_X(\cdot)$ é ilimitada nas frequências $G_j = \cos^{-1}(u_j)$, $j = 1, \dots, k$ (frequências de Gegenbauer).

v) A função de autocovariância $\rho_X(n) = \mathbb{E}(X_n X_0)$ no caso em que o $\max_{j=1, \dots, k} \lambda_j > 0$ tem expressão assintótica: $\rho_X(n) = \sum_{\ell=1, \dots, k: d_\ell > 0} a_\ell |n|^{2\lambda_\ell^* - 1} [\cos(nG_\ell) + o(1)]$,

quando $n \rightarrow \infty$,

$$a_\ell = \begin{cases} 2a_\ell^i, & 0 < G_\ell < \pi; \\ a_\ell^i, & G_\ell = 0 \text{ ou } G_\ell = \pi, \end{cases} \quad (4.27)$$

em que $a_\ell^i = \frac{\sigma_\varepsilon^2}{\pi} \Gamma(1 - 2\lambda_\ell^*) \text{sen}(\pi \lambda_\ell^*) D^2(\ell)$, e λ_ℓ^* dada pela expressão (4.25).

A expressão assintótica da função densidade espectral é dada por:

$$f_X(w) \sim M(\ell) |w - G_\ell|^{-2\lambda_\ell^*}, \quad \text{quando } w \rightarrow G_\ell,$$

para $\ell = 1, \dots, k$, em que:

$$M(\ell) = \frac{\sigma_\varepsilon^2}{2\pi} (D(\ell))^2,$$

com $D(\ell)$ dado pela expressão (4.26).

A seguir definimos os modelos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$).

Definição 4.5. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico que satisfaz a equação:

$$\phi(\mathcal{B}) \prod_{j=1}^k (1 - 2u_j \mathcal{B} + \mathcal{B}^2)^{\lambda_j} (X_t - \mu) = \theta(\mathcal{B}) \varepsilon_t, \quad (4.28)$$

em que k é um inteiro finito, $|u_j| \leq 1$ e λ_j é um número fracionário, para $j = 1, \dots, k$, μ é a média do processo, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um processo ruído branco e $\phi(\cdot)$ e $\theta(\cdot)$ são os polinômios de graus p e q , respectivamente, definidos em (3.8). Então, $\{X_t\}_{t \in \mathbb{Z}}$ é um *processo autorregressivo de média móvel k -Factor Gegenbauer* de ordem $(p, \mathbf{u}, \boldsymbol{\lambda}, q)$, denotado por k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), em que $\mathbf{u} = (u_1, \dots, u_k)^T$ e $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)^T$.

Na proposição a seguir, apresentamos alguns resultados sobre os processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) estabelecidos e provados em Giraitis e Leipus (1995) e Woodward et al. (1998).

Teorema 4.3. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) conforme a Definição 4.5. Então:

- i) O processo $\{X_t\}_{t \in \mathbb{Z}}$ é *estacionário* se todas as raízes da equação $\phi(z) = 0$ estão fora do círculo unitário, e além disso, u_j e λ_j , para $1 \leq j \leq k$, satisfazem a condição do item i) do Teorema 4.2;
- ii) O processo $\{X_t\}_{t \in \mathbb{Z}}$ é *estacionário e tem longa dependência* se todas as raízes da equação $\phi(z) = 0$ estão fora do círculo unitário, e além disso, u_j e λ_j , para $1 \leq j \leq k$, satisfazem a condição do item ii) do Teorema 4.2;
- iii) O processo estacionário $\{X_t\}_{t \in \mathbb{Z}}$ é *causal*, se e somente se, $\phi(z) \neq 0$, para $|z| \leq 1$. Sua representação média móvel é dada por:

$$X_t = \sum_{n \in \mathbb{Z}_{\geq}} \varphi_n \varepsilon_{t-n} \equiv \frac{\theta(\mathcal{B})}{\phi(\mathcal{B})} \prod_{j=1}^k (1 - 2u_j \mathcal{B} + \mathcal{B}^2)^{-\lambda_j} \varepsilon_t.$$

Podemos escrever:

$$\sum_{n \in \mathbb{Z}_{\geq}} \varphi_n \varepsilon_{t-n} \equiv \frac{\theta(\mathcal{B})}{\phi(\mathcal{B})} \sum_{n \in \mathbb{Z}_{\geq}} \psi_n \varepsilon_{t-n},$$

Assim:

$$\varphi(\mathcal{B}) \varepsilon_t \equiv \frac{\theta(\mathcal{B})}{\phi(\mathcal{B})} \psi(\mathcal{B}) \varepsilon_t.$$

Logo temos a seguinte relação:

$$\varphi(z) = \frac{\theta(z)}{\phi(z)}\psi(z),$$

em que $\theta(z)$, $\phi(z)$, são polinômios dados pelas expressões (3.4), (3.3), e:

$$\psi(z) = 1 + \psi_1 z^1 + \psi_2 z^2 + \dots = \sum_{n \in \mathbb{Z}_{\geq}} \psi_n z^n,$$

em que os coeficientes $\{\psi_n\}_{n \in \mathbb{Z}_{\geq}}$ são dados pela expressão (4.24).

Resolvendo as equações polinomiais para $\{\varphi_n\}_{n \in \mathbb{Z}_{\geq}}$, obtemos os coeficientes $\{\varphi_n\}_{n \in \mathbb{Z}_{\geq}}$ da representação média móvel de um processo k -Factor GARMA causal:

$$\varphi_0 = \psi_0 = 1;$$

$$\varphi_1 = \psi_1 + \theta_1 + \phi_1;$$

$$\varphi_2 = \psi_2 + \psi_1 \theta_1 + \theta_2 + \varphi_1 \phi_1 + \phi_2;$$

$$\varphi_3 = \psi_3 + \psi_2 \theta_1 + \psi_1 \theta_2 + \theta_3 + \varphi_2 \phi_1 + \varphi_1 \phi_2 + \phi_3;$$

\vdots

$$\varphi_n = \psi_n + \psi_{n-1} \theta_1 + \psi_{n-2} \theta_2 + \dots + \psi_1 \theta_{n-1} + \theta_n + \varphi_{n-1} \phi_1 + \varphi_{n-2} \phi_2 + \dots + \varphi_1 \phi_{n-1} + \phi_n.$$

Em que $\phi_\ell = 0$ para $\ell > p$, e $\theta_\ell = 0$ para $\ell > q$.

Observação 4.8. Note que os coeficientes $\{\psi_n\}_{n \in \mathbb{Z}_{\geq}}$ são os coeficientes da representação média móvel de um processo k -Factor Gegenbauer causal (k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), com $p = 0 = q$).

- iv) O processo estacionário $\{X_t\}_{t \in \mathbb{Z}}$ é *invertível* se e somente se $\theta(z) \neq 0$, para $|z| \leq 1$;
- v) A *função densidade espectral* do processo k -Factor GARMA, definido pela expressão (4.28), é dada por:

$$f_x(w) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\theta(e^{-iw})|^2}{|\phi(e^{-iw})|^2} \prod_{j=1}^k [2(\cos(w) - u_j)]^{-2\lambda_j}, \quad (4.29)$$

em que $w \in (0, \pi]$ e $G_j = \cos^{-1}(u_j)$ são as chamadas frequências de Gegenbauer.

- vi) A função de autocovariância no caso em que o $\max_{j=1, \dots, k} \lambda_j > 0$ tem *expressão assintótica*:

$$\rho_x(n) = \sum_{\ell=1, \dots, k; d_\ell > 0} a_\ell \frac{|\theta(e^{-iw})|^2}{|\phi(e^{-iw})|^2} |n|^{2\lambda_\ell^* - 1} [\cos(nG_\ell) + o(1)],$$

quando $n \rightarrow \infty$, em que a_ℓ é dado pela expressão (4.27), λ_ℓ^* é dado pela expressão (4.25), e $o(1)$ é uma função de ordem estritamente menor a função constante 1 (ver Definição 2.44).

Lema 4.1. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) dado pela Definição 4.5, causal e invertível cuja função densidade espectral $f_x(\cdot)$ é dada pela equação (4.29). Então:

$$\int_{-\pi}^{\pi} \ln [f_x(w)] dw = 2\pi \ln \left(\frac{\sigma_\varepsilon^2}{2\pi} \right), \quad (4.30)$$

em que $\sigma_\varepsilon^2/2\pi$ é a função densidade espectral do processo ruído branco, denotado por $\{\varepsilon_t\}_{t \in \mathbb{Z}}$.

Demonstração: Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) dado pela expressão (4.28), cuja função densidade espectral é dada pela expressão (4.29).

Então, podemos reescrever a função densidade espectral $f_x(\cdot)$ da seguinte forma:

$$f_x(w) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\theta(e^{-iw})|^2}{|\phi(e^{-iw})|^2} \prod_{j=1}^k |1 - 2u_j e^{-iw} + e^{-i2w}|^{-2\lambda_j}. \quad (4.31)$$

De forma semelhante,

$$\begin{aligned} \frac{2\pi}{\sigma_\varepsilon^2} f_x(w) &= \frac{|\theta(e^{-iw})|^2}{|\phi(e^{-iw})|^2} \prod_{j=1}^k |1 - 2u_j e^{-iw} + e^{-i2w}|^{-2\lambda_j} \\ &= \left| \left[\frac{\theta(e^{-iw})}{\phi(e^{-iw})} \right] \prod_{j=1}^k (1 - 2u_j e^{-iw} + e^{-i2w})^{-\lambda_j} \right|^2 \\ &= \left| \left[\frac{\theta(z)}{\phi(z)} \right] \prod_{j=1}^k (1 - 2u_j z + z^2)^{-\lambda_j} \right|^2, \end{aligned}$$

em que $z = e^{-iw}$, para $w \in (0, \pi]$.

Seja $g(w) = \frac{2\pi}{\sigma_\varepsilon^2} f_x(w)$ e:

$$h(z) = \left[\frac{\theta(z)}{\phi(z)} \right] \prod_{j=1}^k (1 - 2u_j z + z^2)^{-\lambda_j}. \quad (4.32)$$

Então,

$$g(w) = |h(e^{-iw})|^2 = |h(z)|^2, \quad (4.33)$$

em que $z = e^{-iw}$, para $w \in (0, \pi]$.

Temos que $h(0) = 1$ e como o processo é causal e invertível, a função $h(z) \neq 0$, para todo $|z| \leq 1$. Logo, as funções $g(\cdot)$ e $h(\cdot)$ satisfazem as afirmações do Teorema de Fejér para funções polinomiais trigonométricas (ver Bisognin (2007), expressão (A.13) do Teorema 5), assim temos que:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log(g(w)) dw = 2 \log(h(0)) = 2 \log(1) = 0.$$

Por outro lado,

$$\begin{aligned}
\frac{1}{2\pi} \int_{-\pi}^{\pi} \log(g(w)) dw &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left(\frac{2\pi}{\sigma_\varepsilon^2} f_X(w)\right) dw \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left(\frac{2\pi}{\sigma_\varepsilon^2}\right) dw + \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(f_X(w)) dw \\
&= \log\left(\frac{2\pi}{\sigma_\varepsilon^2}\right) + \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(f_X(w)) dw.
\end{aligned}$$

Desta forma, temos que:

$$\log\left(\frac{2\pi}{\sigma_\varepsilon^2}\right) + \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(f_X(w)) dw = 0,$$

ou seja,

$$\begin{aligned}
\int_{-\pi}^{\pi} \log(f_X(w)) dw &= -2\pi \log\left(\frac{2\pi}{\sigma_\varepsilon^2}\right) \\
&= 2\pi \log\left(\frac{\sigma_\varepsilon^2}{2\pi}\right).
\end{aligned}$$

□

A seguir, apresentamos um quadro dos modelos fracionários generalizados e fracionário estudados até aqui. O quadro indica os mais importantes comportamentos desses processos no domínio do tempo (caracterizado pela função de autocorrelação, $\rho_X(\cdot)$) e no domínio da frequência (caracterizado pela função densidade espectral, $f_X(\cdot)$).

Modelos	Domínio do Tempo	Domínio da Frequência
ARFIMA (p, d, q)	$\rho_x(\cdot)$ decai hiperbolicamente (suavemente) para zero, quando $d \in (0, \frac{1}{2})$;	$f_x(\cdot)$ é ilimitada(tende ao infinito) quando as frequências se aproximam de zero;
Gegenbauer (u, λ)	$\rho_x(\cdot)$ decai hiperbolicamente e/ou senoidalmente(ou oscilante) para zero, com $\lambda \in (0, \frac{1}{2})$ e $ u < 1$, ou $\lambda \in (0, \frac{1}{4})$ e $ u = 1$;	$f_x(\cdot)$ é ilimitada(tende ao infinito) em alguma frequência do intervalo $[0, \pi]$;
GARMA (p, u, λ, q)	$\rho_x(\cdot)$ decai hiperbolicamente e/ou senoidalmente(ou oscilante) para zero, com $\lambda \in (0, \frac{1}{2})$ e $ u < 1$, ou $\lambda \in (0, \frac{1}{4})$ e $ u = 1$;	$f_x(\cdot)$ é ilimitada(tende ao infinito) em alguma frequência do intervalo $[0, \pi]$;
k-Factor Gegenbauer ($p, \mathbf{u}, \boldsymbol{\lambda}, q$) com $p = 0 = q$	$\rho_x(\cdot)$ decresce variando lentamente (hiperbolicamente) para zero, quando $\text{máx}\lambda_j > 0$;	$f_x(\cdot)$ é ilimitada(tende ao infinito) em k frequências do intervalo $[0, \pi]$;
k-Factor GARMA ($p, \mathbf{u}, \boldsymbol{\lambda}, q$)	$\rho_x(\cdot)$ decresce variando lentamente (hiperbolicamente) para zero, quando $\text{máx}\lambda_j > 0$;	$f_x(\cdot)$ é ilimitada(tende ao infinito) em k frequências do intervalo $[0, \pi]$.

Podem ser encontrados em Giraitis e Leipus (1995), Woodward et al. (1998) e Ferrara e Guégan (2001), maiores detalhes sobre os processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$).

Na seção a seguir, apresentamos um estimador para as frequências de Gegenbauer e o estimador de Fox e Taquq (1986) para os parâmetros dos processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$).

4.4 Estimação

Nesta seção, apresentamos diversos métodos de estimação para o parâmetro dos processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$).

Na classe paramétrica, a estimação de todos os parâmetros do processo ocorre conjuntamente através do estimador de máxima verossimilhança. Dentre os estimadores paramétricos, citamos o estimador proposto por Fox e Taquq (1986) no qual é utilizado uma aproximação proposta por Whittle (1951) para a matriz de autocovariância do processo.

No caso dos estimadores semiparamétricos e, conseqüentemente, nos estimadores robustos, apenas o parâmetro de longa dependência é estimado. Ou seja, estes estimadores necessitam da localização das frequências de Gegenbauer para estimar o parâmetro de longa dependência.

Para os processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), teríamos que expandir em série de potências o termo $\ln[2(\cos(w) - u_j)]^2$, para cada $j = 1, \dots, k$, que aparece na expressão (4.34) que é mostrada a seguir. Além disso, deveríamos truncar cada série em um certo valor m_j , para $j = 1, \dots, k$. Em seguida, teríamos que aplicar a técnica dos mínimos quadrados para estimar os parâmetros desse modelo, o que seria extremamente complicado e inviável.

Tendo em vista esta dificuldade, Yajima (1996) propõe um estimador gráfico para as frequências de Gegenbauer. As frequências de Gegenbauer são aquelas em que a função periodograma tem um máximo local.

A seguir, apresentamos um estimador para as frequências de Gegenbauer de um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$).

Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) dado pela expressão (4.28) cuja função densidade espectral tem k singularidades, nas frequências $G_j = \cos^{-1}(u_j)$, para $j = 1, \dots, k$, as quais são chamadas de frequências de Gegenbauer. A função densidade espectral tem a seguinte expressão:

$$f_X(w) = f_U(w) \prod_{j=1}^k [2(\cos(w) - u_j)]^{-2\lambda_j}, \quad (4.34)$$

em que

$$f_U(w) = \frac{\sigma_\varepsilon^2 |\theta(e^{-iw})|^2}{2\pi |\phi(e^{-iw})|^2}, \quad \text{para todo } w \in (0, \pi].$$

Os parâmetros u_j e λ_j , para todo $j = 1, \dots, k$, satisfazem as condições do Teorema (4.3).

Seja $\{V_\nu\}_{\nu=1}^k$ uma sequência de intervalos definidos como segue: para todo $\varepsilon > 0$ fixado e para todo $\nu = 1, \dots, k$,

$$V_\nu = (G_\nu - a_\nu, G_\nu - \varepsilon) \cup (G_\nu + \varepsilon, G_\nu + b_\nu),$$

em que $\{a_\nu\}_{\nu=1}^k$ e $\{b_\nu\}_{\nu=1}^k$ são sequências reais, tais que $\{V_\nu\}_{\nu=1}^k$ é uma cobertura do intervalo $(0, \pi]$ quando $\varepsilon \rightarrow 0$, isto é,

$$\bigcup_{\nu=1}^k V_\nu = (0, \pi] \setminus \{G_1, \dots, G_k\}.$$

Definição 4.6. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) dado pela expressão (4.28). Seja $\{X_t\}_{t=1}^n$ uma série temporal obtida a partir do processo $\{X_t\}_{t \in \mathbb{Z}}$ e seja:

$$I(w) := \frac{1}{2\pi n} \left| \sum_{t=1}^n X_t e^{-itw} \right|^2, \quad (4.35)$$

a função periodograma. O estimador para as frequências de Gegenbauer G_j , denotado por \widehat{G}_j , para $j = 1, \dots, k$ é definido por:

$$\widehat{G}_j = \frac{2\pi}{n} \operatorname{argmax}_{w \in V_j} I(w). \quad (4.36)$$

Suposição 4.1. As funções $f_X(\cdot)$ e $f_U(\cdot)$ satisfazem as seguintes afirmações:

- i) $f_X(\cdot)$ é uma função par;
- ii) $f_U(\cdot)$ é uma função positiva no intervalo $[0, \pi]$;
- iii) Para todo $\nu = 1, \dots, k$ e $w \in V_\nu$, $f_U(\cdot)$ satisfaz:

$$\left| \frac{f'_U(w)}{f_U(w)} \right| = \mathcal{O}(|w - G_\nu|^{-1}),$$

para todo $w \in (0, \pi]$. A definição de $f(x) = \mathcal{O}(g(x))$ é encontrada na Definição 2.43.

O seguinte Teorema foi estabelecido por Yajima (1996).

Teorema 4.4. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) com média $\mu = 0$, cuja função densidade espectral satisfaz a equação (4.34). Sob a Suposição 4.1, para todo $\alpha \in (0, 1)$, quando $n \rightarrow \infty$, temos que:

$$n^\alpha(\widehat{G}_j - G_j) \xrightarrow{\mathbb{P}} 0, \quad \text{para cada } j = 1, \dots, k.$$

A seguir, apresentamos os estimadores para o parâmetro, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)$, de longa dependência para um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$).

Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) com média $\mu = 0$, cuja função densidade espectral é dada pela expressão (4.34).

Os parâmetros $u_j = \cos(G_j)$, para $j = 1, \dots, k$, são previamente estimados por $\widehat{u}_j = \cos(\widehat{G}_j)$, em que $\widehat{G}_j, j = 1, \dots, k$, são dados através da expressão (4.36). Pela expressão (4.34), uma vez estimados os parâmetros u_j , a função densidade espectral para este processo é dada por:

$$f_x(w) = f_U(w) \prod_{j=1}^k [2(\cos(w) - \widehat{u}_j)]^{-2\lambda_j}. \quad (4.37)$$

Aplicando a função logarítmica a ambos os lados da equação (4.37) temos:

$$\ln(f_x(w)) = \ln(f_U(w)) - \sum_{j=1}^k \lambda_j \ln[2(\cos(w) - \widehat{u}_j)]^2. \quad (4.38)$$

Adicionando $\ln(f_U(0))$ e $\ln(I(w))$, a ambos os lados da equação (4.38), em que $I(\cdot)$ é a função periodograma dada pela expressão (4.35), obtemos:

$$\begin{aligned} \ln(I(w)) &= \ln(f_U(0)) - \sum_{j=1}^k \lambda_j \ln[2(\cos(w) - \widehat{u}_j)]^2 + \ln \left[\frac{f_U(w)}{f_U(0)} \right] \\ &+ \ln \left[\frac{I(w)}{f_x(w)} \right]. \end{aligned} \quad (4.39)$$

Seja $B = \{0, 1, \dots, g(n) \mid \nu \neq v_j = \operatorname{argmax}_{w \in V_j} I(w), j = 1, \dots, k\}$, em que $g(n)$ é tal que $g(n) \rightarrow \infty$ e $\frac{g(n)}{n} \rightarrow 0$, quando $n \rightarrow \infty$. Substituindo a frequência w pelas frequências de Fourier $w_\nu = \frac{2\pi\nu}{n}$, $\nu \in B$, obtemos uma forma aproximada para a equação (4.39), dada por:

$$\ln(I(w_\nu)) \sim \ln(f_U(0)) - \sum_{j=1}^k \lambda_j \ln[2(\cos(w_\nu) - \widehat{u}_j)]^2 + \ln \left[\frac{I(w_\nu)}{f_x(w_\nu)} \right]. \quad (4.40)$$

O termo $\ln \left[\frac{f_U(w)}{f_U(0)} \right]$, que aparece na equação (4.39), é desprezível se comparado com os outros termos daquela equação, sendo então desconsiderado. Desta forma, podemos observar que a equação (4.40) é uma forma aproximada de uma equação de regressão linear múltipla dada por:

$$y_\nu \sim \beta_0 + \sum_{j=1}^k \beta_j x_{\nu j} + \varepsilon_\nu, \quad \text{para todo } \nu \in B, \quad (4.41)$$

em que

$$y_\nu = \ln(I(w_\nu)), \quad x_{\nu j} = \ln[2(\cos(w_\nu) - \hat{u}_j)]^2,$$

$$\varepsilon_\nu = \ln \left[\frac{I(w_\nu)}{f_x(w_\nu)} \right] + c, \quad \beta_0 = \ln f_\varepsilon(0) - c, \quad c = \mathbb{E} \left(\ln \left[\frac{I(w_\nu)}{f_x(w_\nu)} \right] \right),$$

$$\beta_j = -\lambda_j,$$

com ε_ν variáveis aleatórias não correlacionadas com média zero e variância constante.

Geweke e Porter-Hudak (1983) propõe o estimador semiparamétrico dos mínimos quadrados para o vetor de parâmetros $\beta = (\beta_0, \beta_1, \dots, \beta_k) \in \mathbb{R}^{k+1}$, dado por $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k) \in \mathbb{R}^{k+1}$ os quais minimizam a função perda abaixo:

$$\mathfrak{L}_1(g(n)) = \sum_{\nu \in B} r_\nu^2, \quad (4.42)$$

em que $r_\nu = y_\nu - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{\nu j}$ é o ν -ésimo resíduo.

Denotaremos os estimadores das componentes do parâmetro $\lambda = (\lambda_1, \dots, \lambda_k)$, respectivamente, por:

$$\hat{\lambda}_{GPH-MQ}^j, \quad \text{para } j = 1, \dots, k,$$

para indicar os estimadores de λ_j , obtidos através do método de estimação proposto por Geweke e Porter-Hudak (1983).

Para estimar os parâmetros dos processos ARFIMA(p, d, q), Lopes et al. (2004), Lopes et al. (2006) e Lopes e Mendes (2006) consideram o estimador proposto por Geweke e Porter-Hudak (1983), Robinson (1995a), Hurvich e Ray (1995) e o estimador baseado na função periodograma suavizado de covariâncias (ver Bisognin (2007), Seção A.6), utilizando as janelas de Parzen e Bartlett. Além dos estimadores semi-paramétricos, Lopes e Mendes (2006) utiliza as respectivas versões robustas *MM* e *MQP* destes estimadores.

Da mesma forma que estendemos o estimador proposto por Geweke e Porter-Hudak (1983) para os processos k -Factor GARMA($p, \mathbf{u}, \lambda, q$), podemos estender os outros estimadores utilizados para a estimação nos processos ARFIMA(p, d, q). Maiores detalhes sobre os procedimentos de estimação podem ser encontrados em Bisognin (2007).

Na estimação das frequências de Gegenbauer e do parâmetro de longa dependência, Collet et al. (2003) analisa dois casos particulares.

Assumindo que existam dois picos no gráfico da função periodograma de uma série temporal $\{X_t\}_{t=1}^n$, denotam, respectivamente, por G_1 e G_2 as frequências nas quais estes picos ocorrem. Collet et al. (2003) considera um primeiro caso quando os picos estão distantes um do outro, isto é, $|G_1 - G_2| > 0$ e um segundo caso quando os picos estão próximos um do outro, isto é, $|G_1 - G_2| \sim 0$.

Observação 4.9. Para a situação quando $|G_1 - G_2| \sim 0$, Collet et al. (2003) analisa o vício do estimador proposto por Fox e Taqqu (1986), como pode ser visto nos itens a seguir.

- i) Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) definido através da expressão (4.28). Considere a situação do Caso II e seja $\{X_t\}_{t=1}^n$ uma série temporal gerada a partir do processo $\{X_t\}_{t \in \mathbb{Z}}$. Ajustamos um processo $(k-1)$ -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) à série temporal por hipótese. Seja $\hat{\lambda}_j$ o estimador de máxima verossimilhança para o parâmetro λ_j , para $j = 1, \dots, p, p+2, \dots, k$. Então, o estimador de máxima verossimilhança para o parâmetro de longa dependência λ_p é tal que $\mathbb{E}(\hat{\lambda}_p) > \lambda_p$.
- ii) Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), com frequências de Gegenbauer λ_j , $j = 1, \dots, k$. Assumimos que este processo é um $(k+1)$ -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), com λ_j , $j = 1, \dots, k$ e λ_{k+1} frequências de Gegenbauer, tais que $\lambda_{k+1} \neq \lambda_j$, para todo $j = 1, \dots, k$. Então, o estimador de máxima verossimilhança do parâmetro de longa dependência λ_j , denotado por $\hat{\lambda}_j$, para $j = 1, \dots, k+1$, é tal que $\mathbb{E}(\hat{\lambda}_j) = \lambda_j$, para $j = 1, \dots, k$ e $\mathbb{E}(\hat{\lambda}_{k+1}) = 0$.

Observação 4.10. Collet et al. (2003) propõe um procedimento para detectar o número k de picos na função densidade espectral de um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$). Neste mesmo artigo, os autores definem um critério de parada para o procedimento de estimação de k , fornecem o teste de Kolmogorov e Smirnov (ver Brockwell e Davis (2006)) e um teste para decidir se a estimação do parâmetro de longa dependência λ é significativamente diferente de zero ou não (teste de t-Student).

A seguir, apresentamos os coeficientes das representações média móvel infinita e autorregressiva infinita as quais são de grande importância na demonstração do Lema 4.2 abaixo.

Um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) é dito causal, ver Proposição 4.3, item iii), se existe uma sequência $\{\varphi_\ell\}_{\ell \in \mathbb{Z}_{\geq}}$ tal que $\sum_{\ell \in \mathbb{Z}_{\geq}} |\varphi_\ell| < \infty$ e:

$$\varphi(z) = \sum_{\ell \in \mathbb{Z}_{\geq}} \varphi_\ell z^\ell = \frac{\theta(z)}{\phi(z)} \prod_{j=1}^k (1 - 2u_j \mathcal{B} + \mathcal{B}^2)^{-\lambda_j}. \quad (4.43)$$

Um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) é dito invertível, ver Proposição 4.3, item iv), se existe uma sequência $\{\pi_l\}_{l \in \mathbb{Z}_{\geq}}$ tal que $\sum_{l \in \mathbb{Z}_{\geq}} |\pi_l| < \infty$ e:

$$\pi(z) = \sum_{l \in \mathbb{Z}_{\geq}} \pi_l z^l = \frac{\phi(z)}{\theta(z)} \prod_{j=1}^k (1 - 2u_j \mathcal{B} + \mathcal{B}^2)^{\lambda_j}. \quad (4.44)$$

O Lema 4.2 a seguir, apresenta um resultado sobre a soma finita do produto entre os coeficientes das representações média móvel infinita e autorregressiva infinita, respectivamente, para os processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) causal e invertível. Este resultado é muito importante na previsão de séries temporais utilizando estes processos.

Lema 4.2. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), causal e invertível, dado pela expressão (4.28). Então:

$$\sum_{l=0}^{\ell} \varphi_l \pi_{\ell-l} = 0, \quad (4.45)$$

em que φ_n e π_n são, respectivamente, os coeficientes da representação média móvel infinita e autorregressiva infinita dados nas expressões (4.44) e (4.43), respectivamente.

Demonstração: Por hipótese, o processo $\{X_t\}_{t \in \mathbb{Z}}$ é causal e invertível. Como consequência, temos que:

$$\begin{aligned} \sum_{\ell \geq 0} \sum_{l=0}^{\ell} \varphi_{\ell} \pi_{\ell-l} &= \left(\sum_{\ell \geq 0} \varphi_{\ell} z^{\ell} \right) \times \left(\sum_{l \geq 0} \pi_l z^l \right) \\ &= \left(\frac{\theta(z)}{\phi(z)} \prod_{j=1}^k (1 - 2u_j \mathcal{B} + \mathcal{B}^2)^{-\lambda_j} \right) \\ &\quad \times \left(\frac{\phi(z)}{\theta(z)} \prod_{j=1}^k (1 - 2u_j \mathcal{B} + \mathcal{B}^2)^{\lambda_j} \right) = 1. \end{aligned}$$

Portanto,

$$\sum_{\ell \geq 1} \sum_{l=0}^{\ell} \varphi_{\ell} \pi_{\ell-l} = 0.$$

Ou seja,

$$\sum_{l=0}^{\ell} \varphi_{\ell} \pi_{\ell-l} = 0, \quad \text{para todo } \ell \geq 1.$$

□

O Lema 4.2 acima, nos fornece o comportamento dos coeficientes da série numérica obtida pelo produto das representações média móvel e autorregressiva infinitas dos processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) causal e invertível. Este resultado será extremamente importante na previsão utilizando estes processos. Maiores detalhes sobre previsão, utilizando os processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), podem ser encontrados em Ferrara e Guégan (2001).

Na Seção a seguir, apresentamos o estimador de máxima verossimilhança, proposto por Fox e Taquq (1986) para modelos ARFIMA(p, d, q).

Utilizamos, durante este trabalho, a seguinte notação para a álgebra maricial:

Notação: A matriz B^T é a matriz transposta da matriz B .

4.4.1 ESTIMADOR FT

Para modelos ARFIMA(p, d, q), Fox e Taquq (1986), utilizando a aproximação sugerida por Whittle (1951), apresenta condições que permitem que este estimador seja consistente e tenha distribuição normal assintótica para sequências com forte dependência. Sowell (1992), por sua vez, apresenta resultados para o estimador de máxima verossimilhança exata. Ooms (1995) mostra alguns resultados para modelos com longa dependência sazonais fazendo uma comparação entre três estimadores: o estimador de máxima verossimilhança exata proposto por

Sowell (1992), o estimador de máxima verossimilhança aproximada proposto por Fox e Taquq (1986) e o estimador baseado na regressão da função periodograma proposto por Geweke e Porter-Hudak (1983).

A seguir, descrevemos o procedimento para encontrar o estimador de máxima verossimilhança aproximada.

Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo gaussiano com média $\mu > 0$ e função densidade espectral dada por:

$$f_X(w, \eta, \sigma^2) = \sigma^2 f(w, \eta), \quad w \in (0, \pi],$$

em que $\sigma^2 > 0$ (não precisa ser a variância do processo X_t) e $f(\cdot, \eta)$ é uma função da forma:

$$f(w, \eta) \simeq |w|^{-\alpha(\eta)} G_\eta(w), \quad \text{quando } w \rightarrow 0,$$

com $0 < \alpha(\eta) < 1$ e $G_\eta(w)$ uma função que varia suavemente quando $w \rightarrow 0$ (ver Observação 3.5, ou Observação A.1 em Bisognin (2007)) e η é o vetor de parâmetros desconhecidos os quais desejamos estimar.

A estimação dos parâmetros do vetor η e σ^2 é obtida através da série temporal X_1, X_2, \dots, X_n cuja função de distribuição conjunta, ou função de verossimilhança, é dada por:

$$h(x, \eta, \sigma^2) = (2\pi)^{-\frac{n}{2}} |Q_n(\eta)|^{-\frac{1}{2}} \exp \left\{ -\frac{z^T Q_n^{-1}(\eta) z}{2} \right\}, \quad (4.46)$$

em que $x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ e $Q_n(\eta)$ é a matriz de Toeplitz contendo os valores da função de autocovariância do processo $\{X_t\}_{t \in \mathbb{Z}}$ cujos elementos são dados por $\{Q_n(\eta)\}_{i,j} = \gamma_X(|i-j|)$, $1 \leq i, j \leq n$.

Aplicando a função logarítmica na equação (4.46) obtemos a seguinte expressão:

$$\mathcal{L}(x, \eta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |Q_n(\eta)| - \frac{1}{2} z^T Q_n^{-1}(\eta) z. \quad (4.47)$$

O estimador de máxima verossimilhança de η e de σ^2 é obtido maximizando-se a função $\mathcal{L}(x, \eta, \sigma^2)$ com respeito ao vetor de parâmetros η e σ^2 .

Computacionalmente, é complicado obter a inversa da matriz de autocovariâncias $Q_n(\eta)$. Whittle (1951) estabelece a forma aproximada para $Q_n^{-1}(\eta)$. Fox e Taquq (1986), fazendo uso desta aproximação, aplica o método da máxima verossimilhança aproximada para estimar η e σ^2 maximizando a função:

$$h(x, \eta, \sigma^2) = \left(\frac{1}{\sqrt{2\pi} \sigma} \right)^n \exp \left\{ -\frac{Z^T A_n(\eta) Z}{2 n \sigma^2} \right\}, \quad (4.48)$$

em que $Z = (X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X})$ com $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$ e $A_n(\eta)$ é uma matriz de ordem $n \times n$, proposta por Whittle, para aproximar $Q_n^{-1}(\eta)$, em que o (i, j) -ésimo elemento é dado por:

$$\{A_n(\eta)\}_{i,j} = a_{i-j}(\eta) = a_\ell(\eta) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} e^{i w \ell} [f_X(w, \eta)]^{-1} dw.$$

Os estimadores de máxima verossimilhança aproximada de η e σ^2 que maximizam a equação (4.48), são denotados por $\hat{\eta}_n$ e $\hat{\sigma}_n^2$.

Maximizar a função $h(x, \eta, \sigma^2)$ é equivalente a encontrar $\hat{\eta}_n$ que minimize:

$$\hat{\sigma}_n^2(\eta) = \frac{Z^T A_n(\eta) Z}{n}.$$

Computacionalmente, o estimador de máxima verossimilhança é obtido minimizando a função, na forma discreta:

$$\mathcal{L}_n(\eta) = \frac{1}{2n} \sum_{j=1}^K \left(\ln f_X(w_j, \eta) + \frac{I(w_j)}{f_X(w_j, \eta)} \right), \quad (4.49)$$

em que η denota o vetor de parâmetros desconhecidos, $w_j = \frac{2\pi j}{n}$, $j \in \{0, 1, \dots, K | j \neq \frac{in}{s}\}$, são as frequências de Fourier e $K = \lfloor \frac{n-1}{2} \rfloor$ ($\lfloor x \rfloor$ significa a parte inteira de x).

Denotado por *FT*, o estimador de máxima verossimilhança aproximada proposto por Fox e Taqqu (1986), sob certas condições de regularidade, é consistente e tem distribuição assintótica normal. Um estudo completo desse estimador pode ser encontrado em Fox e Taqqu (1986) e Beran (1994).

Capítulo 5

Modelos de Espaço de Estados

Neste capítulo, apresentamos os modelos de espaço de estados e o algoritmo recursivo de estimação chamado filtro de Kalman na estimação dos parâmetros dos modelos k -Factor GARMA. Os modelos de espaço de estados são uma classe de modelos que vem obtendo grande espaço na análise de séries temporais a partir do final da década de 80. Essa classe tem sido muito estudada nas últimas décadas, a qual foi apresentada inicialmente por Kalman (1960), e Kalman e Bucy (1961) com trabalhos relacionados a controle de sistemas lineares. Akaike (1974a) foi um precursor a aplicar os modelos de espaço de estados em modelos ARMA. Esses modelos têm sido largamente usados para modelar dados em Economia (Harrison e Stevens, 1976; Shumway e Stoffer, 1982; Harvey e Pierse, 1984; Kitagawa e Gersch, 1984), dados da área médica com Jones (1984), e também das ciências do solo com Shumway (1985), dentre outras áreas. A proposta da análise de espaço de estados é inferir relevantes propriedades de séries de dados não observados (estados), a partir do conhecimento de séries de dados observados (observações).

Os processos lineares (ver Definição 2.41) podem ser expressos em termos de um sistema linear de espaço de estados, assim como podem ser descritos por uma decomposição de Wold. Essas representações de um processo, segundo Palma (2007), são equivalentes e estão relacionadas.

Com o intuito de representar os processos GARMA e k -Factor GARMA em um sistema de espaço de estados, necessitamos saber sob quais condições podemos encontrar tais representações. Na observação abaixo, comentamos duas importantes condições para a representação em espaço de estados que são utilizadas neste trabalho.

Observação 5.1.

- i) Se $\{X_t\}_{t \in \mathbb{Z}}$ é um processo estacionário puramente não determinístico, então, pelo Teorema da Decomposição de Wold (ver Teorema 2.4) o processo pode ser expresso por:

$$X_t = \sum_{j \in \mathbb{Z}_{\geq}} \psi_j \varepsilon_{t-j}, \text{ em que } \psi_0 = 1, \sum_{j \in \mathbb{Z}_{\geq}} \psi_j^2 < \infty, \text{ e } \{\varepsilon_t\}_{t \in \mathbb{Z}} \text{ é um processo ruído branco de variância } \sigma^2.$$

Note que o processo $\{X_t\}_{t \in \mathbb{Z}}$ estacionário puramente não determinístico é escrito como um processo linear, logo pode ser representado por um sistema de espaço de estados.

- ii) Se $\{X_t\}_{t \in \mathbb{Z}}$ é um processo Causal, então existe uma sequência $\{\psi_n\}_{n \in \mathbb{Z}_{\geq}}$ tal que:

$$X_t = \sum_{n \in \mathbb{Z}_{\geq}} \psi_n \varepsilon_{t-n}, \quad t \in \mathbb{Z} \text{ e } \sum_{n \in \mathbb{Z}_{\geq}} |\psi_n| < \infty.$$

Percebemos que o processo $\{X_t\}_{t \in \mathbb{Z}}$ causal é escrito como um processo linear (representação média móvel), logo pode ser representado por um sistema linear de espaço de estados. Dessa maneira, podemos representar os processos k-Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) causal por um sistema de espaço de estados.

Muitas séries temporais são representadas por um sistema linear de espaço de estados. Um modelo geral de espaço de estados que representa uma série $\{Y_t\}_{t \in \mathbb{Z}_{\geq}}$ é da forma:

$$Y_t = G_t \mathbf{X}_t + W_t, \quad t \in \mathbb{Z}_{\geq} \text{ (equação de observação),} \quad (5.1)$$

$$\mathbf{X}_{t+1} = F_t \mathbf{X}_t + V_t, \quad t \in \mathbb{Z}_{\geq} \text{ (equação de estado),} \quad (5.2)$$

A equação de estado (5.2) pode ser interpretada como a descrição da evolução do vetor de estado \mathbf{X}_t do sistema (vetor $v \times 1$), no tempo t , em termos de uma conhecida sequência de matrizes $v \times v$, F_0, F_1, F_2, \dots , e uma sequência de vetores aleatórios $\mathbf{X}_0, V_0, V_1, V_2, \dots$, em que os vetores V_0, V_1, V_2, \dots têm ordem $v \times 1$, para $v \in \mathbb{N}$.

A equação de observação (5.1) define a sequência de observações, Y_t , $t \in \mathbb{Z}$, as quais são obtidas pela aplicação de uma transformação linear ao vetor \mathbf{X}_t , adicionada a um vetor ruído W_t , de ordem $w \times 1$, para cada tempo $t \in \mathbb{Z}_{\geq}$, em que $w \in \mathbb{N}$. Deste modo, o comportamento de $\{Y_t\}_{t \in \mathbb{Z}_{\geq}}$ é determinado a partir do vetor de estado \mathbf{X}_t usando a equação de observação.

Em seguida, listamos as hipóteses que são usadas na análise da equação de estado (5.2), e na análise da equação de observação (5.1):

Suposição 5.1.

- a) F_0, F_1, F_2, \dots é uma sequência de matrizes $v \times v$ conhecidas;
- b) G_0, G_1, G_2, \dots é uma sequência de matrizes $w \times v$ conhecidas;
- c) $\{\mathbf{X}_0, (V_t^T, W_t^T)^T; t \in \mathbb{Z}_{\geq}\}$ é uma sequência ortogonal de vetores aleatórios com momentos de segunda ordem finitos (ver Observação 5.2 item iii));
- d) $\mathbb{E}(V_t) = 0$ e $\mathbb{E}(W_t) = 0$, para todo $t \in \mathbb{Z}_{\geq}$;
- e) $\mathbb{E}(V_t V_t^T) = Q_t$, $\mathbb{E}(W_t W_t^T) = R_t$, $\mathbb{E}(V_t W_t^T) = S_t$, em que $\{Q_t\}_{t \in \mathbb{Z}_{\geq}}$, $\{R_t\}_{t \in \mathbb{Z}_{\geq}}$ e $\{S_t\}_{t \in \mathbb{Z}_{\geq}}$ são sequências de matrizes conhecidas de ordens $v \times v$, $w \times w$ e $v \times w$, respectivamente, para $v, w \in \mathbb{N}$.

Observação 5.2.

- i) Sistema de equações da forma (5.1) e (5.2) é conhecido como um sistema de espaço de estados o qual é denominado modelo de espaço de estados.
- ii) A equação de estado (5.2) é generalizada na teoria de controle pela inclusão de um termo adicional $H_t \mathbf{u}_t$ na direita da equação, representando o efeito de aplicar um controle \mathbf{u}_t , no tempo t , para o propósito de influenciar o estado \mathbf{X}_{t+1} .

- iii) Os vetores aleatórios \mathbf{X} e \mathbf{Y} são ditos ortogonais, denotamos $\mathbf{X} \perp \mathbf{Y}$, se $\mathbb{E}(\mathbf{X}\mathbf{Y}^T)$ for igual a zero.
- iv) Se a sequência $\{\mathbf{X}_0, V_0, V_1, V_2, \dots\}$ for independente, então o processo \mathbf{X}_t tem a propriedade de Markov (a distribuição de \mathbf{X}_{t+1} dado $\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_t$ é igual a distribuição de \mathbf{X}_{t+1} dado \mathbf{X}_t), presente em muitas aplicações na área da Física. Em outras palavras, dado o estado presente, o futuro do sistema independe do passado. E é devido a essa propriedade que os cálculos necessários para implementar os modelos de espaço de estados podem ser colocados de forma recursiva. Por esse motivo, a representação de espaço de estados apresenta uma estrutura simples a qual permite uma simples análise da série temporal.
- v) Em diversos casos especiais, como veremos a seguir, as matrizes F_t, G_t, Q_t, R_t e S_t são independentes do tempo t , isto é, são matrizes constantes invariantes no tempo. Em tais casos, suprimimos o subscrito t , assim as matrizes são denotadas por F, G, Q, R e S .
- vi) Temos as seguintes relações de ortogonalidade:

$$V_t \perp \mathbf{X}_s, \quad V_t \perp Y_s, \quad 0 \leq s \leq t, \text{ e}$$

$$W_t \perp \mathbf{X}_s, \quad 0 \leq s \leq t, \quad W_t \perp Y_s, \quad 0 \leq s < t.$$

Para mais detalhes, ver Brockwell e Davis (2006), capítulo 12.

- vii) A equação de estado (5.2) também é chamada, na literatura, de equação de transição, enquanto a equação de observação (5.1) é denominada de equação de medida.

Um *modelo de espaço de estados estacionário de dimensão finita* é dado pelas seguintes equações:

$$Y_t = G\mathbf{X}_t + W_t, \quad t \in \mathbb{Z} \text{ (equação de observação)}, \quad (5.3)$$

$$\mathbf{X}_{t+1} = F\mathbf{X}_t + V_t, \quad t \in \mathbb{Z} \text{ (equação de estado)}, \quad (5.4)$$

em que F e G são matrizes de ordens finitas $v \times v$ e $w \times v$ respectivamente, em que $v, w \in \mathbb{N}$. Além disso, temos que $V_t \sim RB(0, Q)$, em que Q é a matriz de covariância do ruído V_t , de ordem $v \times v$, $W_t \sim RB(0, R)$, em que R é a matriz de covariância do ruído W_t , de ordem $w \times w$, $\mathbb{E}(V_t W_t^T) = S$ para todo $t \in \mathbb{Z}$, a matriz S é de ordem $v \times w$, e V_t é ortogonal a W_t para todo $s \neq t$ com $s, t \in \mathbb{Z}$.

A seguir, definimos a causalidade da equação de estados (5.4). Essa propriedade permite que as observações $\{Y_t\}_{t \in \mathbb{Z}}$ possam ser escritas em uma decomposição de Wold.

Definição 5.1. A equação (5.4) é *dita estável ou causal* se a matriz F tem todos os seus autovalores no interior do círculo unitário, ou equivalentemente, se $\det(I - Fz) \neq 0$, para todo $z \in \mathbb{C}$ tal que $|z| \leq 1$. Assim, a matriz F é dita estável.

No caso estável, a equação de estado (5.4) tem única solução estacionária dada por:

$$\mathbf{X}_t = \sum_{j \in \mathbb{Z}_{>}} F^j \mathbf{V}_{t-j-1}, \quad t \in \mathbb{Z}. \quad (5.5)$$

Logo, obtemos a correspondente sequência de observações também estacionária:

$$Y_t = \mathbf{W}_t + \sum_{j \in \mathbb{Z}_{\geq}} GF^j \mathbf{V}_{t-j-1}, \quad t \in \mathbb{Z}.$$

A seguir, apresentamos algumas representações, via espaço de estados, de modelos para séries temporais. Salientamos que essas representações não são únicas para cada modelo (ver Brockwell e Davis (2006)). Desta maneira, é interessante procurarmos aquela que apresenta a menor representação, isto é, o vetor de estados com número mínimo de elementos. Isto diminui o número de cálculos, o que computacionalmente é muito vantajoso. Por isso, mais tarde, será definido os conceitos de controlabilidade e observabilidade os quais nos fornecem critérios para tal escolha.

Para a discussão desses conceitos, apresentamos uma subclasse de modelos de espaço de estados estacionários de dimensão finita para o processo $\{Y_t\}_{t \in \mathbb{Z}}$ definido pelas equações:

$$Y_t = G\mathbf{X}_t + \varepsilon_t, \quad t \in \mathbb{Z} \text{ (equação de observação);} \tag{5.6}$$

$$\mathbf{X}_{t+1} = F\mathbf{X}_t + H\varepsilon_t, \quad t \in \mathbb{Z} \text{ (equação de estado),}$$

$\{\varepsilon_t\} \sim RB(0, \Sigma)$, vetor de dimensão $v \times 1$, em que Σ é a matriz de covariância de ε_t , e H é uma matriz de dimensão $v \times w$ e F uma matriz estável de dimensão $v \times v$, em que $v, w \in \mathbb{N}$.

Caso o processo $\{Y_t\}_{t \in \mathbb{Z}}$ seja invertível, o modelo (5.6) é dito representação de inovações. Podemos observar que, se $\{Y_t\}_{t \in \mathbb{Z}}$ tem representação de inovações, então $\{Y_t\}_{t \in \mathbb{Z}}$ tem uma representação da forma descrita pelas equações (5.4) e (5.3).

Em seguida, apresentamos a recíproca da afirmação anterior a qual é bastante utilizada para o estudo da controlabilidade e observabilidade de um sistema de espaço de estados.

Proposição 5.1. O modelo de espaço de estados dado pelas equações (5.4) e (5.3) tem representação de inovações da forma (5.6) se a equação (5.4) for estável.

A demonstração desta proposição pode ser encontrada em Brockwell e Davis (2006) no Capítulo 12.

5.1 Representação em Espaço de Estados

A seguir, apresentamos a representação em espaço de estados de um processo ARMA(p, q). Podemos escrever um processo ARMA(p, q) causal da seguinte forma:

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \tag{5.7}$$

em que $\varepsilon_t \sim RB(0, \sigma^2)$.

Assumimos que o processo $\{Y_t\}_{t \in \mathbb{Z}}$ é causal, e definimos:

- a) $m = \max(p, q)$ e $\phi_j = 0$ para $j > p$;

b) $Y_{t|s} = \mathbb{E}(Y_t | Y_j, j \leq s) = \mathbb{E}(Y_t | Y_s, Y_{s-1}, Y_{s-2}, \dots)$;

c) $\mathbf{X}_{t|t-1} = (Y_{t|t-1}, Y_{t+1|t-1}, \dots, Y_{t+m-1|t-1})^T$ é o vetor de estados (vetor coluna).

Assim, da hipótese de causalidade temos:

$$\begin{aligned} Y_{t+j|t} &= \sum_{k=j}^{\infty} \psi_k \varepsilon_{t+j-k} = \sum_{k=j+1}^{\infty} \psi_k \varepsilon_{t+j-k} + \psi_j \varepsilon_t, \quad j = 0, \dots, m; \\ &= Y_{t+j|t-1} + \psi_j \varepsilon_t. \end{aligned}$$

Trocando t por $t+m$ na equação (5.7), tomando a Esperança condicionada até o tempo t em ambos os lados dessa equação ($\mathbb{E}(Y_{t+m|t})$), e utilizando os coeficientes da representação

média móvel, $\psi_j = \sum_{k=1}^j \phi_k \psi_{j-k} + \theta_j$, para $j = 1, 2, \dots, m$ obtemos as relações:

$$\begin{aligned} \mathbb{E}(Y_{t+m} | Y_t, Y_{t-1}, \dots) &= \mathbb{E}(\phi_1 Y_{t+m-1} | Y_t, Y_{t-1}, \dots) + \mathbb{E}(\phi_2 Y_{t+m-2} | Y_t, Y_{t-1}, \dots) \\ &\quad + \dots + \mathbb{E}(\phi_m Y_t | Y_t, Y_{t-1}, \dots); \\ Y_{t+m|t} &= \phi_1 Y_{t+m-1|t} + \phi_2 Y_{t+m-2|t} + \dots + \phi_m Y_{t|t}; \\ &= \sum_{k=1}^m \phi_k Y_{t+m-k|t} \quad e, \\ Y_{t+m|t} &= \sum_{k=m}^{\infty} \psi_k \varepsilon_{t+m-k} = \sum_{k=m+1}^{\infty} \psi_k \varepsilon_{t+m-k} + \psi_m \varepsilon_t; \\ &= Y_{t+m|t-1} + \psi_m \varepsilon_t; \\ &= \sum_{k=1}^m \phi_k Y_{t+m-k|t-1} + \psi_m \varepsilon_t. \end{aligned} \tag{5.8}$$

Observação 5.3.

i) Da equação (5.8), obtemos a seguinte expressão:

$$Y_{t+m|t} = \phi_1 Y_{t+m-1|t-1} + \phi_2 Y_{t+m-2|t-1} + \dots + \phi_m Y_{t|t-1} + \psi_m \varepsilon_t. \tag{5.9}$$

A expressão (5.9) juntamente com a equação a seguir:

$$Y_{t+j|t} = Y_{t+j|t-1} + \psi_j \varepsilon_t, \quad \text{para } j = 1, \dots, m-1,$$

originam a equação de estados.

Para $j = 0$, temos que $Y_{t|t} = Y_{t|t-1} + \varepsilon_t$. Essa equação origina a equação de observação, notando que $Y_{t|t} = Y_t$.

ii) Note que $Y_{t+m|t}$ é função de $Y_{t|t-1}, Y_{t+1|t-1}, \dots$, e $Y_{t+m-1|t-1}$. Então, o vetor de estados é $\mathbf{X}_{t|t-1} = (Y_{t|t-1}, Y_{t+1|t-1}, \dots, Y_{t+m-1|t-1})^T$.

O vetor de estados descrito no item *ii*) da observação (5.3) satisfaz $\mathbf{X}_{t+1|t} = F\mathbf{X}_{t|t-1} + H\varepsilon_t$. Assim, obtemos a equação de estados em que $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ é sua única solução estacionária:

$$\underbrace{\begin{bmatrix} Y_{t+1|t} \\ Y_{t+2|t} \\ \vdots \\ Y_{t+m-1|t} \\ Y_{t+m|t} \end{bmatrix}}_{\mathbf{X}_{t+1}} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \cdots & 1 & \\ \phi_m & \phi_{m-1} & \phi_{m-2} & \cdots & \phi_1 & \end{bmatrix}}_F \underbrace{\begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \\ \vdots \\ Y_{t+m-2|t-1} \\ Y_{t+m-1|t-1} \end{bmatrix}}_{\mathbf{X}_t} + \underbrace{\begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{m-1} \\ \psi_m \end{bmatrix}}_H \varepsilon_t, t \in \mathbb{Z}, \quad (5.10)$$

e também a equação de observação (no tempo presente t):

$$Y_t = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \end{bmatrix}}_G \underbrace{\begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \\ \vdots \\ Y_{t+m-2|t-1} \\ Y_{t+m-1|t-1} \end{bmatrix}}_{\mathbf{X}_t} + \varepsilon_t, t \in \mathbb{Z}. \quad (5.11)$$

Note que essa representação satisfaz as equações:

$$Y_t = G\mathbf{X}_t + \varepsilon_t; \quad (5.12)$$

$$\mathbf{X}_{t+1} = F\mathbf{X}_t + H\varepsilon_t,$$

em que tomamos $\mathbf{X}_t = \mathbf{X}_{t|t-1}$ e $\mathbf{X}_{t+1} = \mathbf{X}_{t+1|t}$.

Portanto, a representação está de acordo com as equações que definem o sistema de espaço de estados dado na expressão (5.6).

Teorema 5.1. Se $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ é a única solução de (5.10) e se $\{Y_t\}_{t \in \mathbb{Z}}$ é dado por (5.11), então $\{Y_t\}_{t \in \mathbb{Z}}$ é a única solução da equação do modelo ARMA(p, q).

Demonstração:

Seja F e G as matrizes coeficientes do vetor de estados \mathbf{X}_t nas equações (5.10) e (5.11) respectivamente, e seja $H = (\psi_1, \psi_2, \dots, \psi_m)^T$. Então, a partir do modelo de espaço de estados, definidos pelo sistema de equações (5.6), podemos notar que:

$$\begin{aligned}
Y_t &= GX_t + \varepsilon_t; \\
Y_{t+1} &= GX_{t+1} + \varepsilon_{t+1} = G[FX_t + H\varepsilon_t] + \varepsilon_{t+1} = GFX_t + GH\varepsilon_t + \varepsilon_{t+1}; \\
Y_{t+2} &= GF^2X_t + GFH\varepsilon_t + GH\varepsilon_{t+1} + \varepsilon_{t+2}; \\
&\vdots \\
Y_{t+m-1} &= GF^{m-1}X_t + GF^{m-2}H\varepsilon_t + \cdots + GH\varepsilon_{t+m-2} + \varepsilon_{t+m-1}; \\
Y_{t+m} &= GF^mX_t + GF^{m-1}H\varepsilon_t + \cdots + GH\varepsilon_{t+m-1} + \varepsilon_{t+m}.
\end{aligned}$$

Assim, temos que:

$$\begin{aligned}
-\phi_m Y_t &= -\phi_m GX_t - \phi_m \varepsilon_t; \\
-\phi_{m-1} Y_{t+1} &= -\phi_{m-1} GFX_t - \phi_{m-1} GH\varepsilon_t - \phi_{m-1} \varepsilon_{t+1}; \\
-\phi_{m-2} Y_{t+2} &= -\phi_{m-2} GF^2X_t - \phi_{m-2} GFH\varepsilon_t - \phi_{m-2} GH\varepsilon_{t+1} - \phi_{m-2} \varepsilon_{t+2}; \\
&\vdots \\
-\phi_1 Y_{t+m-1} &= -\phi_1 GF^{m-1}X_t - \phi_1 GF^{m-2}H\varepsilon_t - \cdots - \phi_1 GH\varepsilon_{t+m-2} - \phi_1 \varepsilon_{t+m-1}; \\
Y_{t+m} &= GF^mX_t + GF^{m-1}H\varepsilon_t + \cdots + GH\varepsilon_{t+m-1} + \varepsilon_{t+m}.
\end{aligned}$$

Somando as equações anteriores, e reagrupando os termos temos que:

$$\begin{aligned}
&Y_{t+m} - \phi_1 Y_{t+m-1} - \cdots - \phi_{m-2} Y_{t+2} - \phi_{m-1} Y_{t+1} - \phi_m Y_t \\
&= G(F^m - \phi_1 F^{m-1} - \cdots - \phi_{m-1} F - \phi_m) X_t \\
&\quad + [G(F^{m-1} - \phi_1 F^{m-2} - \cdots - \phi_{m-2} F - \phi_{m-1}) H - \phi_m] \varepsilon_t \\
&\quad + [G(F^{m-2} - \phi_1 F^{m-3} - \cdots - \phi_{m-3} F - \phi_{m-2}) H - \phi_{m-1}] \varepsilon_{t+1} \\
&\quad \vdots \\
&\quad + [G(F - \phi_1) H - \phi_2] \varepsilon_{t+m-2} \\
&\quad + [GH - \phi_1] \varepsilon_{t+m-1} \\
&\quad + \varepsilon_{t+m}.
\end{aligned}$$

A primeira parcela da soma é zero, visto que o polinômio característico de F é dado por $\det(zI - F) = z^m - \phi_1 z^{m-1} - \dots - \phi_m$. Logo pelo Teorema de Cayley-Hamilton (ver a observação 5.4) temos que $(F^m - \phi_1 F^{m-1} - \dots - \phi_m = 0)$.

Tomando $\psi_i = GF^{i-1}H$, para $i = 1, \dots, m$ temos que:

$$\begin{aligned}
Y_{t+m} & - \phi_1 Y_{t+m-1} - \dots - \phi_{m-2} Y_{t+2} - \phi_{m-1} Y_{t+1} - \phi_m Y_t \\
& = \varepsilon_{t+m} \\
& \quad + \underbrace{[\psi_1 - \phi_1]}_{\theta_1} \varepsilon_{t+m-1} \\
& \quad + \underbrace{[\psi_2 - \phi_1 \psi_1 - \phi_2]}_{\theta_2} \varepsilon_{t+m-2} \\
& \quad \vdots \\
& \quad + \underbrace{[\psi_{m-1} - \phi_1 \psi_{m-2} - \dots - \phi_{m-2} \psi_1 - \phi_{m-1}]}_{\theta_{m-1}} \varepsilon_{t+1} \\
& \quad + \underbrace{[\psi_m - \phi_1 \psi_{m-1} - \dots - \phi_{m-2} \psi_2 - \phi_{m-1} \psi_1 - \phi_m]}_{\theta_m} \varepsilon_t.
\end{aligned}$$

Como os coeficientes da representação média móvel deste modelo são dados por: $\psi_j =$

$$\sum_{k=1}^j \phi_k \psi_{j-k} + \theta_j, \text{ para } j = 1, 2, \dots, m, \text{ podemos definir } \theta_j = \psi_j - \sum_{k=1}^j \phi_k \psi_{j-k} + \theta_m, \text{ com } j = 1, 2, \dots, m.$$

Portanto,

$$Y_{t+m} - \phi_1 Y_{t+m-1} - \dots - \phi_{m-2} Y_{t+2} - \phi_{m-1} Y_{t+1} - \phi_m Y_t = \varepsilon_{t+m} + \theta_1 \varepsilon_{t+m-1} + \dots + \theta_{m-1} \varepsilon_{t+1} + \theta_m \varepsilon_t.$$

Logo, $\{Y_t\}_{t \in \mathbb{Z}}$ satisfaz as equações do modelo ARMA(p, q), isto é, $\{Y_t\}_{t \in \mathbb{Z}}$ definido em (5.11), satisfaz a equação $\phi(\mathcal{B})Y_t = \theta(\mathcal{B})\varepsilon_t$, para $t \in \mathbb{Z}$, coincidindo com a única solução estacionária destas equações. \square

Observação 5.4. O Teorema de Cayley-Hamilton tem o seguinte enunciado:

Seja A uma matriz $n \times n$ sobre o corpo K e seja p o polinômio característico de A . Então, $p(A) = 0$.

Em geral, tomamos $K = \mathbb{R}$ neste trabalho.

A demonstração deste Teorema encontra-se em Hoffman e Kunze (1970).

A seguir, apresentamos alguns exemplos de representações em espaço de estados para processo ARMA(p, q):

Exemplo 5.1. Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo ARMA(p, q), com $p = 2 = q$. Então:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t, \text{ com } m = \text{máx}(p, q) = 2.$$

Assim:

$$\mathbf{X}_{t+1} = \begin{bmatrix} 0 & 1 \\ \phi_2 & \phi_1 \end{bmatrix} \mathbf{X}_t + \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} \varepsilon_t \quad (\text{equação de estado}),$$

e

$$Y_t = [1 \ 0] \mathbf{X}_t + \varepsilon_t \quad (\text{equação de observação}),$$

em que $\psi_1 = \phi_1 + \theta_1$ e $\psi_2 = \phi_1 \psi_1 + \phi_2 + \theta_2$.

Analogamente, podemos escrever:

$$\begin{bmatrix} Y_{t+1|t} \\ Y_{t+2|t} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \phi_2 & \phi_1 \end{bmatrix} \begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \end{bmatrix} + \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} \varepsilon_t \quad (\text{equação de estado}),$$

e

$$Y_t = [1 \ 0] \begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \end{bmatrix} + \varepsilon_t \quad (\text{equação de observação}).$$

Exemplo 5.2. Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo ARMA(p, q), com $p = 1$, e com $q = 3$. Então:

$$Y_t = \phi_1 Y_{t-1} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \varepsilon_t, \text{ com } m = \text{máx}(p, q) = 3.$$

Assim:

$$\mathbf{X}_{t+1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \phi_1 \end{bmatrix} \mathbf{X}_t + \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} \varepsilon_t \quad (\text{equação de estado}),$$

e

$$Y_t = [1 \ 0 \ 0] \mathbf{X}_t + \varepsilon_t \quad (\text{equação de observação})$$

em que $\psi_1 = \phi_1 + \theta_1$, $\psi_2 = \phi_1 \psi_1 + \theta_2$ e $\psi_3 = \phi_1 \psi_2 + \theta_3$.

Analogamente, podemos escrever:

$$\begin{bmatrix} Y_{t+1|t} \\ Y_{t+2|t} \\ Y_{t+3|t} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \phi_1 \end{bmatrix} \begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \\ Y_{t+2|t-1} \end{bmatrix} + \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} \varepsilon_t \quad (\text{equação de estado}),$$

e

$$Y_t = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \\ Y_{t+2|t-1} \end{bmatrix} + \varepsilon_t \quad (\text{equação de observação}).$$

Exemplo 5.3. Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo ARMA(p, q), com $p = 3$, e com $q = 1$. Então:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \theta_1 \varepsilon_{t-1} + \varepsilon_t, \text{ com } m = \max(p, q) = 3.$$

Assim:

$$\mathbf{X}_{t+1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \phi_3 & \phi_2 & \phi_1 \end{bmatrix} \mathbf{X}_t + \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} \varepsilon_t \quad (\text{equação de estado}),$$

e

$$Y_t = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \mathbf{X}_t + \varepsilon_t \quad (\text{equação de observação})$$

em que $\psi_1 = \phi_1 + \theta_1$, $\psi_2 = \phi_1 \psi_1 + \phi_2$ e $\psi_3 = \phi_1 \psi_2 + \phi_2 \psi_1 + \phi_3$.

Analogamente, podemos escrever:

$$\begin{bmatrix} Y_{t+1|t} \\ Y_{t+2|t} \\ Y_{t+3|t} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \phi_3 & \phi_2 & \phi_1 \end{bmatrix} \begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \\ Y_{t+2|t-1} \end{bmatrix} + \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} \varepsilon_t \quad (\text{equação de estado}),$$

e

$$Y_t = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \\ Y_{t+2|t-1} \end{bmatrix} + \varepsilon_t \quad (\text{equação de observação}).$$

Para um modelo não estacionário ARIMA(p, d, q) (ver Definição 3.2) é possível encontrar uma representação de espaço de estados de dimensão finita. Segundo Brockwell e Davis (2006), o processo ARIMA(p, d, q) pode ser representado da seguinte maneira:

Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo ARIMA(p, d, q) com $(1 - \mathcal{B})^d Y_t$ satisfazendo um modelo ARMA(p, q) (ver Definição 3.2), então o processo $(1 - \mathcal{B})^d Y_t$ tem a representação de espaço de estados:

$$(1 - \mathcal{B})^d Y_t = G \mathbf{X}_t + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (5.13)$$

em que $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ é a única solução estacionária da equação:

$$\mathbf{X}_{t+1} = F \mathbf{X}_t + H \varepsilon_t,$$

e sejam G , F e H as matrizes coeficientes da representação de espaço de estados dados pelas equações (5.11) e (5.10), em que A e B são matrizes de dimensão $d \times 1$ e $d \times d$ respectivamente, definidas por:

$$A = B = 1, \text{ se } d = 1, \text{ e } A = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ (-1)^{d+1} \binom{d}{d} & (-1)^d \binom{d}{d-1} & (-1)^{d-1} \binom{d}{d-2} & \cdots & d \end{bmatrix},$$

se $d > 1$.

Então, obtemos a expressão:

$$Y_t = (1 - \mathcal{B})^d Y_t - \sum_{j=1}^d \binom{d}{j} (-1)^j Y_{t-j},$$

com o vetor $\mathbf{Y}_{t-1} := (Y_{t-d}, \dots, Y_{t-1})^T$ satisfazendo a equação:

$$\mathbf{Y}_t = A(1 - \mathcal{B})^d Y_t + B\mathbf{Y}_{t-1} = AG\mathbf{X}_t + B\mathbf{Y}_{t-1} + A\varepsilon_t.$$

Assim, definimos um novo vetor de estado \mathbf{T}_t pela concatenação dos vetores \mathbf{X}_t e \mathbf{Y}_{t-1} , e a partir disso, obtemos a equação de estado, dada por:

$$\mathbf{T}_{t+1} := \begin{bmatrix} \mathbf{X}_{t+1} \\ \mathbf{Y}_t \end{bmatrix} = \begin{bmatrix} F & 0 \\ AG & B \end{bmatrix} \mathbf{T}_t + \begin{bmatrix} H \\ A \end{bmatrix} \varepsilon_t, \quad t \in \mathbb{N}, \quad (5.14)$$

Deste modo, obtemos a nova equação de observação a partir das equações e (5.13) e (5.14), dada por:

$$Y_t = [G \quad (-1)^{d+1} \binom{d}{d} \quad (-1)^d \binom{d}{d-1} \quad (-1)^{d-1} \binom{d}{d-2} \quad \cdots \quad d] \begin{bmatrix} \mathbf{X}_t \\ \mathbf{Y}_{t-1} \end{bmatrix} + \varepsilon_t, \quad t \in \mathbb{N},$$

Com a condição inicial $\mathbf{T}_1 = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{Y}_0 \end{bmatrix} = \begin{bmatrix} \sum_{j=0}^{\infty} F^j H \varepsilon_{-j} \\ \mathbf{Y}_0 \end{bmatrix}$, e a condição de ortogonalidade $\mathbf{Y}_0 \perp \varepsilon_t$ para $t \in \mathbb{Z}$.

A seguir, mostramos um exemplo de representação em espaço de estados de um processo ARIMA(p, d, q).

Exemplo 5.4. Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo ARIMA(p, d, q), com $p = 2$, $d=1$ e com $q=1$. Então:

$$(1 - \phi_1 \mathcal{B} - \phi_2 \mathcal{B}^2)(1 - \mathcal{B})Y_t = (1 + \theta_1)\varepsilon_t$$

Desta maneira, temos:

$$(1 - \mathcal{B})Y_t = \phi_1(1 - \mathcal{B})Y_{t-1} + \phi_2(1 - \mathcal{B})Y_{t-2} + \theta_1\varepsilon_t + \varepsilon_t$$

Assim:

$$\begin{bmatrix} \mathbf{X}_{t+1} \\ Y_t \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ \phi_2 & \phi_1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_t \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} H \\ A \end{bmatrix} \varepsilon_t \quad (\text{equação de estado}),$$

e

$$Y_t = [1 \ 0 \ 1] \begin{bmatrix} \mathbf{X}_t \\ Y_{t-1} \end{bmatrix} + \varepsilon_t \quad (\text{equação de observação}).$$

Analogamente, podemos escrever:

$$\begin{bmatrix} Y_{t+1|t} \\ Y_{t+2|t} \\ Y_t \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ \phi_2 & \phi_1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \psi_1 \\ \psi_2 \\ 1 \end{bmatrix} \varepsilon_t \quad (\text{equação de estado}),$$

e

$$Y_t = [1 \ 0 \ 1] \begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \\ Y_{t-1} \end{bmatrix} + \varepsilon_t \quad (\text{equação de observação}),$$

em que $\psi_1 = \phi_1 + \theta_1$, $\psi_2 = \phi_1\psi_1 + \phi_2$.

Na próxima seção, apresentamos as propriedades de controlabilidade e observabilidade de um sistema de espaço de estados.

5.2 Controlabilidade e Observabilidade

Os conceitos de controlabilidade e observabilidade permitem encontrar condições para a obtenção da representação mínima em espaço de estados, isto é, aquela que tem o menor número de elementos para o vetor de estado. Encontrar essa representação significa um ganho computacional grande, à medida que ela reduz a quantidade de cálculos.

Em seguida, apresentamos a matriz de Hankel cujo posto está relacionado com a dimensão da representação de um processo estocástico em sistema de espaço de estados.

Seja o modelo de espaço de estados dado pelo sistema de equações (5.6), com as suposições de que $\psi_0 = 1$ e $\psi_j = GF^{j-1}H \in \mathbb{R}$, para todo $j > 0$, tal que $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, então o processo $\{Y_t\}_{t \in \mathbb{Z}}$ pode ser escrito na expansão de Wold:

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}. \quad (5.15)$$

Assim, o processo $\{Y_t\}_{t \in \mathbb{Z}}$ pode ser caracterizado pela Matriz de Hankel \mathbf{H}^* , a qual é formada pelos coeficientes $\{\psi_j\}_{j \geq 0}$ da representação média móvel infinita do processo (dada pela equação (5.15)). Dessa forma, a matriz de Hankel do processo $\{Y_t\}_{t \in \mathbb{Z}}$ é dada por:

$$\mathbf{H}^* = \begin{bmatrix} \psi_1 & \psi_2 & \psi_3 & \cdots \\ \psi_2 & \psi_3 & \psi_4 & \cdots \\ \psi_3 & \psi_4 & \psi_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (5.16)$$

Note que a matriz de Hankel determina a expansão de Wold e vice-versa.

A seguir, apresentamos o conceito de controlabilidade para um sistema de espaço de estados, com a seguinte definição:

Definição 5.2. (Controlabilidade). Seja $C = (H, FH, F^2H, \dots)$ o operador controlabilidade. O sistema dado pelas equações (5.6) é dito *controlável* se, e somente se, C tem posto completo ou, equivalentemente, $C^T C$ é invertível.

Observação 5.5. Uma matriz A , de dimensão $m \times n$, é dita ter *posto completo* se $\text{posto}(A) = \min\{m, n\}$, isto é, seu posto tem o maior valor possível. Em outras palavras, A tem posto completo se apresentar número máximo possível de linhas (colunas) linearmente independentes.

A idéia principal do conceito de controlabilidade é: dado a história, ou passado do processo ruído no tempo t , $\mathcal{E}_{t-1} = (\dots, \varepsilon_{t-2}, \varepsilon_{t-1})$, queremos escolher uma sequência adequada de ruídos passados \mathcal{E}_{t-1} para chegar a um particular valor de \mathbf{X}_t . Para responder essa questão, podemos escrever o estado no tempo t como:

$$\begin{aligned} \mathbf{X}_t &= H\varepsilon_{t-1} + FH\varepsilon_{t-2} + F^2H\varepsilon_{t-3} + \dots, \\ &= C\mathcal{E}_{t-1}. \end{aligned}$$

Então, se o sistema é controlável, $C^T C$ tem posto completo e podemos escrever:

$$\mathcal{E}_{t-1} = (C^T C)^{-1} C^T X_t.$$

Em seguida, apresentamos o conceito de observabilidade para um sistema de espaço de estados, com a seguinte definição:

Definição 5.3. (Observabilidade). Seja $O = (G^T, F^T G^T, (F^T)^2 G^T, \dots)^T$ o operador observabilidade. O sistema dado pelas equações (5.6) é dito *observável* se, e somente se, O tem posto completo ou, equivalentemente, $O^T O$ é invertível.

A definição de observabilidade está relacionada com o problema de determinar os valores do estado inicial não observado X_0 da trajetória do processo observado $\{Y_0, Y_1, \dots\}$ na ausência de ruídos de estado e de ruídos observacionais. Vamos considerar o sistema determinístico de espaço de estados:

$$\begin{aligned} Y_t &= G\mathbf{X}_t, \quad t \in \mathbb{Z}; \\ \mathbf{X}_{t+1} &= F\mathbf{X}_t, \quad t \in \mathbb{Z}, \end{aligned}$$

em que $Y_t = (Y_0, Y_1, \dots)^T$ é a trajetória do processo. Caso tenhamos:

$$\begin{aligned} Y_0 &= G\mathbf{X}_0 \\ Y_1 &= GF\mathbf{X}_0 \\ Y_2 &= GF^2\mathbf{X}_0 \\ &\vdots \end{aligned} \tag{5.17}$$

podemos escrever $Y = O\mathbf{X}_0$. Assim, se $O^T O$ tem posto completo podemos determinar, explicitamente, o valor do estado inicial como $\mathbf{X}_0 = (O^T O)^{-1} O^T O\mathbf{X}_0$, ou equivalentemente, $\mathbf{X}_0 = (O^T O)^{-1} O^T Y$.

Observação 5.6. Para um sistema de dimensão finita, o Teorema de Cayley- Hamilton estabelece que os operadores de observabilidade e de controlabilidade podem ser escritos, respectivamente, do seguinte modo:

$$O = (G^T, F^T G^T, (F^T)^2 G^T, \dots, (F^T)^{n-1} G^T)^T,$$

$$C = (H, FH, F^2 H, \dots, F^{n-1} H),$$

em que $n = \text{posto}(O) = \text{posto}(C)$.

A seguir, definimos sistema de espaço de estados minimal.

Definição 5.4. Um sistema de espaço de estados é dito Minimal se, e somente se, F tem dimensão mínima entre todas as representações do processo linear (ver equação (5.15)).

Observação 5.7. Muitas vezes, dizemos que o sistema de espaço de estados de dimensão mínima é aquele que possui o vetor de estados de menor dimensão. Isto está de acordo com a Definição 5.4, pois, quanto menor a dimensão da matriz F menor será a dimensão do vetor de estado \mathbf{X}_t no modelo de espaço de estados, dado pelo sistema de equações (5.6).

Para um sistema finito-dimensional, o problema de encontrar a representação minimal é bastante importante. Isso ocorre, visto que, quanto menor a dimensão, mais fácil é sua interpretação, ou ainda, mais simples de trabalhar numericamente. No entanto, para um sistema de dimensão infinita, encontrar a menor representação é um problema irrelevante.

O Teorema abaixo, encontrado em Palma (2007), estabelece uma condição para que um sistema de espaço de estados seja minimal.

Teorema 5.2. Um sistema de espaço de estados é dito *minimal* se, e somente se, ele é observável e controlável (ver Hannan e Deistler (1988)).

Para mais detalhes, ver Brocwell e Davis (2006), Capítulo 12; Wei (2006), Capítulo 18; e Palma (2007), Capítulo 2.

Em geral, considerando o sistema de espaço de estados :

$$Y_t = G\mathbf{X}_t + \varepsilon_t, \quad t \in \mathbb{Z};$$

$$\mathbf{X}_{t+1} = F\mathbf{X}_t + H\varepsilon_t, \quad t \in \mathbb{Z},$$

em que supomos $\psi_0 = 1$ e $\psi_j = GF^{j-1}H \in \mathbb{R}$, para todo $j > 0$, tal que $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Temos

que o processo $\{Y_t\}_{t \in \mathbb{Z}}$ pode ser escrito em decomposição de Wold, isto é, $Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$.

Desta maneira, podemos representar o processo $\{Y_t\}_{t \in \mathbb{Z}}$ em espaço de estados como:

$$\mathbf{X}_t = \begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \\ Y_{t+2|t-1} \\ \vdots \end{bmatrix}, \quad Y_{t+i|j} = \mathbb{E}[Y_{t+i} | Y_j, Y_{j-1}, \dots],$$

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{bmatrix}, \quad H = [\psi_1 \quad \psi_2 \quad \psi_3 \quad \cdots]^T \quad \text{e} \quad G = [1 \quad 0 \quad 0 \quad \cdots].$$

A seguir, veremos resultados bastantes importantes para nosso trabalho. Eles permitem afirmar que existe somente representação de espaço de estados de dimensão infinita para processos de longa dependência.

Vamos considerar um sistema de espaço de estados de dimensão finita dado pelas equações:

$$Y_t = G\mathbf{X}_t + \eta_t, \quad t \in \mathbb{Z}, \tag{5.18}$$

$$\mathbf{X}_{t+1} = F\mathbf{X}_t + \varepsilon_t, \quad t \in \mathbb{Z},$$

em que $Y_t \in \mathbb{R}$ e $\mathbf{X}_t \in \mathbb{R}^m$, e m é a dimensão da representação de espaço de estados.

O próximo resultado caracteriza a dimensão da representação de espaço de estados para qualquer processo estocástico causal $\{Y_t\}_{t \in \mathbb{Z}}$ dado.

Lema 5.1. Considere o sistema de espaço de estados finito dimensional dado pelo sistema de equações (5.18). Se esse sistema é observável e $\{Y_t\}_{t \in \mathbb{Z}}$ é estacionário, então a sequência de estados, $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$, é também estacionária.

O teorema a seguir relaciona a dimensão da representação em espaços de estados de um processo estocástico com a relação de dependência entre as observações do processo.

Teorema 5.3. Qualquer processo estocástico causal com representação de espaço de estados finito-dimensional é um processo de memória curta, isto é, $\gamma_k \sim a^k$, $|a| < 1$, em que γ_k é a função de autocovariância do processo.

As demonstrações do Lema 5.1 e do Teorema 5.3 podem ser encontradas em Chan e Palma (1998).

Deste modo, como consequência do Teorema 5.3 temos o seguinte corolário:

Corolário 5.1. Não existe representação em espaço de estados de dimensão finita de um processo ARFIMA(p, d, q) com parâmetro de longa dependência $d \neq 0$.

Consideremos $\{Y_t\}_{t \in \mathbb{Z}}$ um processo causal ARFIMA(p, d, q) com representação média móvel infinita dada por:

$$Y_t = \frac{\theta(\mathcal{B})}{\phi(\mathcal{B})} (1 - \mathcal{B})^{-d} \varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

em que $\{\psi_j\}_{j \geq 0}$ satisfazem $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)} (1 - z)^{-d}$.

Uma representação de espaço de estados de dimensão infinita para este processo pode ser expressa do seguinte modo:

$$Y_t = G\mathbf{X}_t + \varepsilon_t, \quad t \in \mathbb{Z};$$

$$\mathbf{X}_{t+1} = F\mathbf{X}_t + H\varepsilon_t, \quad t \in \mathbb{Z},$$

em que

$$\mathbf{X}_t = \begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \\ Y_{t+2|t-1} \\ \vdots \end{bmatrix}, \quad Y_{t+i|j} = \mathbb{E}[Y_{t+i} | Y_j, Y_{j-1}, \dots],$$

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ \vdots & \vdots & \ddots & & \end{bmatrix}, \quad H = [\psi_1 \quad \psi_2 \quad \psi_3 \quad \dots]^T \quad \text{e} \quad G = [1 \quad 0 \quad 0 \quad \dots].$$

Em seguida, enunciamos três corolários que são consequências diretas do Teorema (5.3) que indicam como é a representação, em espaço de estados, dos processos fracionários generalizados com a propriedade de longa dependência.

Corolário 5.2. Não existe representação em espaço de estados de dimensão finita do processo GARMA(p, u, λ, q) (causal) estacionário de longa dependência, isto é, quando $|u| < 1$ e $0 < \lambda < \frac{1}{2}$, ou quando $|u| = 1$ e $0 < \lambda < \frac{1}{4}$.

Corolário 5.3. Não existe representação em espaço de estados de dimensão finita do processo k -Factor Gegenbauer($p, \mathbf{u}, \boldsymbol{\lambda}, q$) (causal) estacionário de longa dependência, com $p=q$, isto é, quando $|u_j| < 1$ e $0 < \lambda_j < \frac{1}{2}$, ou quando $|u_j| = 1$ e $0 < \lambda_j < \frac{1}{4}$ para $j = 1, 2, \dots, k$, em que $\mathbf{u} = (u_1, \dots, u_k)^T$ e $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)^T$.

Corolário 5.4. Não existe representação em espaço de estados de dimensão finita do processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) (causal) estacionário de longa memória, isto é, quando $|u_j| < 1$ e $0 < \lambda_j < \frac{1}{2}$, ou quando $|u_j| = 1$ e $0 < \lambda_j < \frac{1}{4}$ para $j = 1, 2, \dots, k$, em que $\mathbf{u} = (u_1, \dots, u_k)^T$ e $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)^T$.

Pelo Teorema 5.4 abaixo, a dimensionalidade da matriz de Hankel \mathbf{H}^* , dada pela equação (5.16), está relacionada com a racionalidade da função densidade espectral do processo $\{Y_t\}_{t \in \mathbb{Z}}$ causal dado pela equação (5.15). Assim, podemos perceber que a dimensionalidade do sistema de espaço de estados está relacionada com a dimensionalidade da matriz de Hankel.

A seguir, apresentamos a definição de função racional.

Definição 5.5. Uma função $f(\cdot)$ é dita *racional* se ela é escrita como razão de dois polinômios.

O seguinte Teorema, encontrado em Palma (2007), determina a relação entre o posto da Matriz de Hankel \mathbf{H}^* e a racionalidade do processo.

Teorema 5.4. O posto de \mathbf{H}^* é finito se, e somente se, a densidade espectral do processo $\{Y_t\}_{t \in \mathbb{Z}}$, dado pela expressão (5.15), é racional.

Os processos ARMA(p, q) apresentam função densidade espectral racional (ver equação (3.6)), portanto, pelo Teorema 5.4, o posto de \mathbf{H}^* é finito para esses modelos. Isto significa que todo processo ARMA(p, q) pode ser representado por um sistema de espaço de estados de dimensão finita. Por outro lado, a classe de processos de longa-memória (modelos ARFIMA(p, d, q)), não tem função densidade espectral racional (ver equação (3.9)), portanto toda representação em sistema de espaço de estados para esses modelos tem dimensão infinita, como visto no Corolário 5.1.

Queremos representar os processos fracionários generalizados em espaço de estados, com o objetivo posterior de estudarmos o filtro de Kalman para que tenhamos a possibilidade de estimarmos os parâmetros desses processos via espaço de estados. Indicamos como representar tais processos, baseados em Palma (2007), a partir dos exemplos abaixo:

Exemplo 5.5. Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo causal GARMA(p, u, λ, q) com representação média móvel infinita dada por:

$$Y_t = \frac{\theta(\mathcal{B})}{\phi(\mathcal{B})} (1 - 2u\mathcal{B} + \mathcal{B}^2)^{-\lambda} \varepsilon_t = \sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-j},$$

em que $\{\psi_j\}_{j \geq 0}$ satisfazem $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)} (1 - 2uz + z^2)^{-\lambda}$.

Uma representação de espaço de estados de dimensão infinita para este processo pode ser expressa como:

$$Y_t = G\mathbf{X}_t + \varepsilon_t, \quad t \in \mathbb{Z};$$

$$\mathbf{X}_{t+1} = F\mathbf{X}_t + H\varepsilon_t, \quad t \in \mathbb{Z},$$

em que

$$\mathbf{X}_t = \begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \\ Y_{t+2|t-1} \\ \vdots \end{bmatrix}, \quad Y_{t+i|j} = \mathbb{E}[Y_{t+i} | Y_j, Y_{j-1}, \dots],$$

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \ddots & & \end{bmatrix}, \quad H = [\psi_1 \ \psi_2 \ \psi_3 \ \cdots]^T \quad \text{e} \quad G = [1 \ 0 \ 0 \ \cdots].$$

Exemplo 5.6. Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo causal k -Factor GARMA(p, u, λ, q), em que $u = (u_1, \dots, u_k)^T$ e $\lambda = (\lambda_1, \dots, \lambda_k)^T$. Esse processo tem uma representação média móvel infinita dada por:

$$Y_t = \frac{\theta(\mathcal{B})}{\phi(\mathcal{B})} \prod_{j=1}^k (1 - 2u_j \mathcal{B} + \mathcal{B}^2)^{-\lambda_j} \varepsilon_t = \sum_{n=0}^{+\infty} \varphi_n \varepsilon_{t-n},$$

em que $\{\varphi_n\}_{n \geq 0}$ satisfazem $\varphi(z) = \sum_{n=0}^{\infty} \varphi_n z^n = \frac{\theta(z)}{\phi(z)} \prod_{j=1}^k (1 - 2u_j z + z^2)^{-\lambda_j}$.

Uma representação de espaço de estados de dimensão infinita para este processo pode ser expressa como:

$$Y_t = G\mathbf{X}_t + \varepsilon_t, \quad t \in \mathbb{Z};$$

$$\mathbf{X}_{t+1} = F\mathbf{X}_t + H\varepsilon_t, \quad t \in \mathbb{Z},$$

em que

$$\mathbf{X}_t = \begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \\ Y_{t+2|t-1} \\ \vdots \end{bmatrix}, \quad Y_{t+i|j} = \mathbb{E}[Y_{t+i} | Y_j, Y_{j-1}, \dots],$$

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \ddots & & \end{bmatrix}, \quad H = [\varphi_1 \ \varphi_2 \ \varphi_3 \ \cdots]^T \quad \text{e} \quad G = [1 \ 0 \ 0 \ \cdots].$$

Chan e Palma (1998) mostra que, embora um processo de longa dependência tem representação infinita em espaço de estados, a função de máxima verossimilhança exata pode ser computada recursivamente em um número finito de passos pelo uso do filtro de Kalman.

A seguir, introduzimos o conceito de filtro de Kalman para, posteriormente, vermos os resultados que nos permitem computar, recursivamente, a função de verossimilhança em um número finito de passos.

Vimos que em um sistema de espaço de estados, como por exemplo, o sistema definido pelas equações (5.6), o vetor de estados não é diretamente observado. Ele, por sua vez, pode representar uma variável física, uma dinâmica econômica, uma variável instrumental, entre outras. Por outro lado, temos que o processo $\{Y_t\}_{t \in \mathbb{Z}}$ é observado. Deste modo, um problema importante é estimar o estado não observado \mathbf{X}_t a partir do processo observado $\{Y_t\}_{t \in \mathbb{Z}}$.

Para a estimação do estado \mathbf{X}_t , dependendo de qual informação avaliamos, temos os seguintes casos:

- i) **Predição:** Passado do processo avaliado $(\dots, Y_{t-2}, Y_{t-1})$;
- ii) **Filtragem:** Passado e presente do processo avaliados $(\dots, Y_{t-2}, Y_{t-1}, Y_t)$;
- iii) **Suavização:** Passado, presente e futuro do processo avaliados $(\dots, Y_{t-1}, Y_t, Y_{t+1}, \dots)$.

As equações recursivas de Kalman permitem encontrar os estados preditores, filtros e suavizadores. Essas equações tem por objetivo simplificar os cálculos numéricos dos estimadores.

Neste trabalho, damos ênfase ao estudo do filtro de Kalman o qual será utilizado para fazermos estimações dos parâmetros dos modelos fracionários generalizados.

5.3 Filtro de Kalman

O filtro de Kalman é um algoritmo de estimação recursiva, introduzido por Rudolph E. Kalman em 1960 (com Kalman, 1960 e Kalman e Bucy, 1961), o qual, inicialmente, foi relacionado com controle de sistemas lineares. Esse procedimento permite descrever o comportamento de um sistema dinâmico (sistema que pode ser descrito por um sistema de espaço de estados). Ele calcula uma estimativa para o estado não-observável \mathbf{X}_t em um dado instante de um sistema linear que evolui no tempo, perturbado por um ruído (erro). E essa estimativa, por sua vez, é baseada em observações (Y_t) linearmente relacionadas com esse estado, porém também corrompidas por um ruído.

No filtro de Kalman, avaliamos o passado e o presente do processo para encontrar um estimador para o estado não observado \mathbf{X}_t (Filtragem).

Podemos dizer que Filtragem é a estimação do estado corrente de um sistema. O motivo da expressão está no fato de que é um processo para obter a melhor estimação (que minimiza o erro quadrático médio) a partir de dados ruidosos. Então, esse termo filtragem é usado no sentido da tentativa de eliminar um sinal indesejado: o ruído do sistema.

O filtro de Kalman é, essencialmente, um conjunto de equações matemáticas que, para um modelo gaussiano de espaço de estados, implementa um estimador de mínimos quadrados do tipo preditor-corretor.

A simplicidade de implementação e robustez do filtro, juntamente com o avanço da era digital, proporcionaram sua ampla divulgação, visto que ele possibilitou a implementação de

soluções recursivas para diversas aplicações em tempo real. Assim, ganhou espaço em importantes áreas da engenharia de controle, da física e de outras ciências. Como exemplo de aplicações, temos: processamento de sinais, navegação aeroespacial, sonares subaquáticos, controle estatístico de qualidade, previsão de séries temporais, computação gráfica, robótica (predição de movimento), acústica inercial (sensores).

A seguir, descrevemos as equações recursivas do filtro de Kalman e seu funcionamento para um sistema de espaço de estados linear gaussiano.

Considere o modelo de espaço de estados linear gaussiano geral que pode ser descrito na seguinte forma:

$$Y_t = G_t \mathbf{X}_t + \varepsilon_t, \quad t \in \mathbb{Z} \text{ (equação de observação);} \quad (5.19)$$

$$\mathbf{X}_{t+1} = F_t \mathbf{X}_t + R_{t+1} \eta_{t+1}, \quad t \in \mathbb{Z} \text{ (equação de estado),} \quad (5.20)$$

em que $\varepsilon_t \sim N(0, H_t)$ e $\eta_t \sim N(0, Q_t)$ para $t \in \mathbb{Z}$.

A equação do estado (5.20) descreve a dinâmica do vetor de estado \mathbf{X}_t que é dirigido pela entrada estocástica $\{\eta_t\}_{t \in \mathbb{Z}}$, o qual é chamado de erro ou ruído do sistema. A equação de observação (5.19) relaciona linearmente a resposta observada Y_t com o vetor de estado não observado \mathbf{X}_t , e com o ruído de observação $\{\varepsilon_t\}_{t \in \mathbb{Z}}$.

Assumimos que as matrizes G_t, F_t, R_t, H_t e Q_t são conhecidas e podem depender de um vetor de parâmetros sendo variantes no tempo, ou apenas serem constantes no tempo.

Os erros $\{\eta_t\}_{t \in \mathbb{Z}}$ e $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ são assumidos serialmente independentes e independentes entre si para todo tempo $t \in \mathbb{Z}$. Isso implica que $\mathbb{E}[\eta_t \varepsilon_s^T] = 0$, para todo $t, s \in \mathbb{Z}$.

Para especificarmos a distribuição dos vetores de estado, \mathbf{X}_t para $t = 1, 2, \dots$, os quais são dados pela equação (5.20), assumimos que o vetor inicial do estado não-observável \mathbf{X}_0 tem média $\hat{\mathbf{X}}_0$ e matriz de autocovariância P_0 conhecidas e sua distribuição é normal, isto é, $\mathbf{X}_0 \sim N(\hat{\mathbf{X}}_0, P_0)$. Além disso, supomos que \mathbf{X}_0 independe de $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ e de $\{\eta_t\}_{t \in \mathbb{Z}}$ para todo $t \in \mathbb{Z}$. Assim, com essas hipóteses, segundo Brockwell e Davis (2006), o sistema tem a propriedade de Markov, isto é, a distribuição de \mathbf{X}_{t+1} dado $\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_t$ é igual a distribuição de \mathbf{X}_{t+1} dado \mathbf{X}_t . Em outras palavras, a propriedade de Markov significa que dado o presente do sistema, o futuro do sistema independe do passado, como podemos observar na equação de estado (5.20).

Assim, temos que $\{\eta_t\}_{t \in \mathbb{Z}}$ e $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ são assumidos ter distribuição conjunta normal multivariada com média $\mathbf{0}$ e matriz de autocovariância bloco diagonal, denotado por:

$$\begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} \sim \mathbf{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q_t & 0 \\ 0 & H_t \end{bmatrix} \right), \quad t \in \mathbb{Z}.$$

Observação 5.8.

- i) Temos que $Y_1, Y_2, \dots, Y_{t-1}, Y_t$ denotam os valores observados do processo (ou variável) de interesse nos tempos $1, 2, \dots, t-1, t$.
- ii) Como o vetor das observações Y_t depende de uma quantidade não observável \mathbf{X}_t , o vetor de estado, temos por objetivo fazer inferências sobre \mathbf{X}_t a partir do conhecimento das observações Y_t .

- iii) A equação de observação (5.19) tem estrutura de um modelo de Regressão Linear Múltipla em que o vetor de coeficientes \mathbf{X}_t varia no tempo.
- iv) A equação de estado (5.20) representa um modelo autoregressivo de primeira ordem, com natureza Markoviana da qual decorrem muitas propriedades do Modelo de espaço de estados.

A seguir, obtemos o filtro de Kalman para o modelo de espaço de estados dado pelas equações (5.19) e (5.20) em que $\mathbf{X}_0 \sim N(\widehat{\mathbf{X}}_0, P_0)$, e com $\widehat{\mathbf{X}}_0$ e P_0 conhecidos.

Para isso, queremos obter a distribuição condicional de \mathbf{X}_{t+1} , para $t \in \mathbb{Z}_{\geq}$, dado as observações Y_1, Y_2, \dots, Y_t . Denotaremos por $\mathbf{Y}_t^* = \{Y_1, Y_2, \dots, Y_t\}$.

Construído o modelo de espaço de estados podemos calcular as previsões ℓ -passos a frente, da seguinte maneira:

$$\widehat{\mathbf{X}}_{t+\ell|t} = \mathbb{E}(\mathbf{X}_{t+\ell} | Y_1, Y_2, \dots, Y_t) = \mathbb{E}(\mathbf{X}_{t+\ell} | \mathbf{Y}_t^*);$$

Em particular, $\widehat{\mathbf{X}}_t = \mathbb{E}(\mathbf{X}_t | \mathbf{Y}_t^*)$ e $\widehat{\mathbf{X}}_{t+1|t} = \mathbb{E}(\mathbf{X}_{t+1} | \mathbf{Y}_t^*)$. Também temos que $\widehat{\mathbf{Y}}_{t+\ell|t} = \mathbb{E}(\mathbf{Y}_{t+\ell} | \mathbf{Y}_t^*)$.

As estimativas $\widehat{\mathbf{X}}_t$ do vetor de estados \mathbf{X}_t resumem a informação do passado necessária para prever o futuro. Quando avaliadas, novas informações são utilizadas para atualizar o vetor de estado assim como, as previsões.

A distribuição dos vetores de estado, denotada por $\mathbb{P}(\mathbf{X}_t)$, para $t = 1, 2, \dots$, é determinada pela hipótese de que, no tempo zero, $\mathbf{X}_0 \sim N(\widehat{\mathbf{X}}_0, P_0)$, e pela equação de estado (5.20). Essas distribuições são referidas como *distribuições à priori*. Depois de observar os dados, atualizamos nossa distribuição *à priori* obtendo uma *distribuição à posteriori* do vetor de estado. Essa distribuição *à posteriori* é uma distribuição condicional denotada por $\mathbb{P}(\mathbf{X}_t | \mathbf{Y}_t^*)$.

Como anteriormente citado, o filtro de Kalman é um procedimento recursivo que faz inferências sobre um estado \mathbf{X}_t . A noção fundamental é que dadas as observações $\{Y_1, Y_2, \dots, Y_t\}$, a inferência sobre \mathbf{X}_t pode ser obtida pela aplicação direta do Teorema de Bayes.

Informalmente, temos pelo Teorema de Bayes que a probabilidade do estado dadas as observações é proporcional a probabilidade das observações dado o estado. Assim podemos escrever:

$$\mathbb{P}(\mathbf{X}_{t+1} | \mathbf{Y}_{t+1}^*) \propto \mathbb{P}(Y_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*) \mathbb{P}(\mathbf{X}_{t+1} | \mathbf{Y}_t^*). \quad (5.21)$$

Portanto, uma distribuição *à posteriori* $\mathbb{P}(\mathbf{X}_{t+1} | \mathbf{Y}_t^*)$, um passo a frente, torna-se uma distribuição *à priori* para calcular uma nova distribuição *à posteriori* $\mathbb{P}(\mathbf{X}_{t+1} | \mathbf{Y}_{t+1}^*)$, quando uma nova observação Y_{t+1} torna-se avaliada.

Observação 5.9. Na expressão (5.21), $\mathbb{P}(\mathbf{X}_{t+1} | \mathbf{Y}_{t+1}^*)$ denota a distribuição *à posteriori* do processo $\{\mathbf{X}_t\}_{t \in \mathbb{Z}_{\geq}}$, no tempo $t + 1$, equanto que, $\mathbb{P}(Y_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*)$ denota a função de verossimilhança, e $\mathbb{P}(\mathbf{X}_{t+1} | \mathbf{Y}_t^*)$ é a distribuição *à priori* para $\{\mathbf{X}_t\}_{t \in \mathbb{Z}_{\geq}}$.

Agora, supomos que a distribuição *à posteriori* $\mathbb{P}(\mathbf{X}_t | \mathbf{Y}_t^*)$, do vetor de estados \mathbf{X}_t , no tempo t , é uma distribuição Normal com média $\widehat{\mathbf{X}}_t$ e matriz de covariância P_t , temos que:

$$(\mathbf{X}_t | \mathbf{Y}_t^*) \sim N(\widehat{\mathbf{X}}_t, P_t). \quad (5.22)$$

Observamos que a recursão de Kalman inicia a partir da escolha, no tempo zero, de $\widehat{\mathbf{X}}_0$ e P_0 para serem as estimativas sobre a média e a variância de \mathbf{X}_0 , respectivamente.

Em seguida, temos duas etapas da recursão, tendo em vista o tempo $t + 1$.

1) Antes de observar Y_{t+1} :

A escolha de \mathbf{X}_{t+1} é governada pela equação (5.20) e como a distribuição *à posteriori* $\mathbb{P}(\mathbf{X}_t|\mathbf{Y}_t^*)$, é uma Normal com média $\widehat{\mathbf{X}}_t$ e matriz de covariância P_t , temos que a estimativa do estado não observado, no tempo $t + 1$, dado as observações até o tempo t , é dada por:

$$\begin{aligned}\widehat{\mathbf{X}}_{t+1|t} &= \mathbb{E}(\mathbf{X}_{t+1}|\mathbf{Y}_t^*) = \mathbb{E}(F_t\mathbf{X}_t + R_{t+1}\eta_{t+1}|\mathbf{Y}_t^*) \\ &= F_t\mathbb{E}(\mathbf{X}_t|\mathbf{Y}_t^*) + R_{t+1}\mathbb{E}(\eta_{t+1}|\mathbf{Y}_t^*) \\ &= F_t\mathbb{E}(\mathbf{X}_t|\mathbf{Y}_t^*) = F_t\widehat{\mathbf{X}}_t.\end{aligned}$$

Portanto, obtemos:

$$\widehat{\mathbf{X}}_{t+1|t} = F_t\widehat{\mathbf{X}}_t. \quad (5.23)$$

E também, a matriz de covariância do estado não observado, no tempo $t + 1$, dado as observações até o tempo t , é dada por:

$$\begin{aligned}P_{t+1|t} &= \text{Var}(\mathbf{X}_{t+1}|\mathbf{Y}_t^*) = \mathbb{E}(\mathbf{X}_{t+1}^2|\mathbf{Y}_t^*) - \mathbb{E}^2(\mathbf{X}_{t+1}|\mathbf{Y}_t^*) \\ &= \mathbb{E}([F_t\mathbf{X}_t + R_{t+1}\eta_{t+1}]^2|\mathbf{Y}_t^*) - [F_t\mathbb{E}(\mathbf{X}_t|\mathbf{Y}_t^*)]^2 \\ &= \mathbb{E}([F_t\mathbf{X}_t + R_{t+1}\eta_{t+1}][F_t\mathbf{X}_t + R_{t+1}\eta_{t+1}]^T|\mathbf{Y}_t^*) - [F_t\mathbb{E}(\mathbf{X}_t|\mathbf{Y}_t^*)][F_t\mathbb{E}(\mathbf{X}_t|\mathbf{Y}_t^*)]^T \\ &= \mathbb{E}([F_t\mathbf{X}_t + R_{t+1}\eta_{t+1}][(F_t\mathbf{X}_t)^T + (R_{t+1}\eta_{t+1})^T]|\mathbf{Y}_t^*) - F_t\mathbb{E}(\mathbf{X}_t|\mathbf{Y}_t^*)\mathbb{E}^T(\mathbf{X}_t|\mathbf{Y}_t^*)F_t^T \\ &= \mathbb{E}([F_t\mathbf{X}_t + R_{t+1}\eta_{t+1}][\mathbf{X}_t^T F_t^T + \eta_{t+1}^T R_{t+1}^T]|\mathbf{Y}_t^*) - F_t\mathbb{E}^2(\mathbf{X}_t|\mathbf{Y}_t^*)F_t^T \\ &= \mathbb{E}(F_t\mathbf{X}_t\mathbf{X}_t^T F_t^T + F_t\mathbf{X}_t\eta_{t+1}^T R_{t+1}^T + R_{t+1}\eta_{t+1}\mathbf{X}_t^T F_t^T + R_{t+1}\eta_{t+1}\eta_{t+1}^T R_{t+1}^T|\mathbf{Y}_t^*) \\ &\quad - F_t\mathbb{E}^2(\mathbf{X}_t|\mathbf{Y}_t^*)F_t^T \\ &= F_t\mathbb{E}(\mathbf{X}_t\mathbf{X}_t^T|\mathbf{Y}_t^*)F_t^T + F_t\mathbb{E}(\mathbf{X}_t\eta_{t+1}^T|\mathbf{Y}_t^*)R_{t+1}^T + R_{t+1}\mathbb{E}(\eta_{t+1}\mathbf{X}_t^T|\mathbf{Y}_t^*)F_t^T \\ &\quad + R_{t+1}\mathbb{E}(\eta_{t+1}\eta_{t+1}^T|\mathbf{Y}_t^*)R_{t+1}^T - F_t\mathbb{E}^2(\mathbf{X}_t|\mathbf{Y}_t^*)F_t^T \\ &= F_t\mathbb{E}(\mathbf{X}_t^2|\mathbf{Y}_t^*)F_t^T - F_t\mathbb{E}^2(\mathbf{X}_t|\mathbf{Y}_t^*)F_t^T + R_{t+1}\mathbb{E}(\eta_{t+1}\eta_{t+1}^T|\mathbf{Y}_t^*)R_{t+1}^T \\ &= F_t[\mathbb{E}(\mathbf{X}_t^2|\mathbf{Y}_t^*) - \mathbb{E}^2(\mathbf{X}_t|\mathbf{Y}_t^*)]F_t^T + R_{t+1}\mathbb{E}(\eta_{t+1}^2|\mathbf{Y}_t^*)R_{t+1}^T \\ &= F_t[\text{Var}(\mathbf{X}_t|\mathbf{Y}_t^*)]F_t^T + R_{t+1}Q_{t+1}R_{t+1}^T \\ &= F_tP_tF_t^T + R_{t+1}Q_{t+1}R_{t+1}^T.\end{aligned}$$

Logo temos que:

$$P_{t+1|t} = F_tP_tF_t^T + R_{t+1}Q_{t+1}R_{t+1}^T \quad (5.24)$$

Observação 5.10. No cálculo da matriz de covariância acima, utilizamos os seguintes argumentos: os ruídos $\{\eta_t\}_{t \in \mathbb{Z}}$ e $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ são não-correlacionados entre si e não-correlacionados com o vetor de estado inicial \mathbf{X}_0 , juntamente com o fato de que os estados \mathbf{X}_t e as observações

Y_t dependem linearmente, respectivamente, dos vetores $(\mathbf{X}_0, \eta_1, \eta_2, \dots, \eta_t)$ e $(\mathbf{X}_0, \eta_1, \eta_2, \dots, \eta_t, \varepsilon_t)$, para $t \in \mathbb{Z}_{\geq}$, então podemos concluir baseados em Brockwell e Davis (2006), que $\mathbb{E}(\eta_t \mathbf{X}_s^T) = 0$, $\mathbb{E}(\eta_t Y_s^T) = 0$, para $0 \leq s \leq t$, e $\mathbb{E}(\varepsilon_t \mathbf{X}_s^T) = 0$, $\mathbb{E}(\varepsilon_t Y_s^T) = 0$ para $t \in \mathbb{Z}_{\geq}$.

É por este motivo que, por exemplo, obtemos: $\mathbb{E}(\mathbf{X}_t \eta_{t+1}^T | \mathbf{Y}_t^*) = 0$ e $\mathbb{E}(\eta_{t+1} \mathbf{X}_t^T | \mathbf{Y}_t^*) = 0$.

Portanto, no tempo $t + 1$, tendo observado Y_1, Y_2, \dots, Y_t , porém antes de observar Y_{t+1} , temos pela equação de estado (5.20) a distribuição *à priori*:

$$(\mathbf{X}_{t+1} | \mathbf{Y}_t^*) \sim N(F_t \widehat{\mathbf{X}}_t, P_{t+1|t}). \quad (5.25)$$

2) Depois de observar Y_{t+1} :

A partir do conhecimento da observação Y_{t+1} , queremos computar a distribuição *à posteriori* de \mathbf{X}_{t+1} , isto é, a distribuição de $(\mathbf{X}_{t+1} | \mathbf{Y}_{t+1}^*)$, usando o Teorema de Bayes na expressão (5.21). No entanto, precisamos, previamente, conhecer a função de verossimilhança de $(Y_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*)$.

A obtenção dessa distribuição *à posteriori* é dada pelos argumentos a seguir:

Seja e_{t+1} o erro de previsão da observação Y_{t+1} dadas as medidas até o tempo t . Então:

$$\begin{aligned} e_{t+1} &= Y_{t+1} - \widehat{Y}_{t+1|t} \\ &= Y_{t+1} - G_{t+1} \widehat{\mathbf{X}}_{t+1|t} = Y_{t+1} - G_{t+1} F_t \widehat{\mathbf{X}}_t \\ &= G_{t+1} \mathbf{X}_{t+1} + \varepsilon_{t+1} - G_{t+1} \widehat{\mathbf{X}}_{t+1|t} \\ &= G_{t+1} (\mathbf{X}_{t+1} - \widehat{\mathbf{X}}_{t+1|t}) + \varepsilon_{t+1} \\ &= G_{t+1} (\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t) + \varepsilon_{t+1}. \end{aligned}$$

Assim, conhecidos G_{t+1} , F_t e $\widehat{\mathbf{X}}_t$ podemos dizer que observar Y_{t+1} é equivalente a observar e_{t+1} . Portanto obtemos a seguinte relação:

$$\mathbb{P}(\mathbf{X}_{t+1} | \mathbf{Y}_{t+1}^*) = \mathbb{P}(\mathbf{X}_{t+1} | Y_{t+1}, \mathbf{Y}_t^*) = \mathbb{P}(\mathbf{X}_{t+1} | e_{t+1}, \mathbf{Y}_t^*).$$

Com essa expressão, podemos reescrever a expressão (5.21) como:

$$\mathbb{P}(\mathbf{X}_{t+1} | e_{t+1}, \mathbf{Y}_t^*) \propto \mathbb{P}(e_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*) \mathbb{P}(\mathbf{X}_{t+1} | \mathbf{Y}_t^*), \quad (5.26)$$

em que $\mathbb{P}(e_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*)$ é a função de verossimilhança.

Observação 5.11.

- i) Pela expressão (5.26), computar a distribuição *à posteriori* $\mathbb{P}(\mathbf{X}_{t+1} | \mathbf{Y}_{t+1}^*)$ é equivalente a computar a distribuição $\mathbb{P}(\mathbf{X}_{t+1} | e_{t+1}, \mathbf{Y}_t^*)$.
- ii) Para obter a distribuição $\mathbb{P}(\mathbf{X}_{t+1} | e_{t+1}, \mathbf{Y}_t^*)$ percebemos, pela expressão (5.26), que necessitamos conhecer as distribuições $\mathbb{P}(e_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*)$ e $\mathbb{P}(\mathbf{X}_{t+1} | \mathbf{Y}_t^*)$. Esta última distribuição é dada pela expressão (5.25).

Agora, vamos encontrar a distribuição $\mathbb{P}(e_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*)$. Calculando a esperança de e_{t+1} dado \mathbf{X}_{t+1} e dadas as observações até o tempo t , temos que:

$$\begin{aligned}
\mathbb{E}(e_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*) &= \mathbb{E}[G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t) + \varepsilon_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*] \\
&= \mathbb{E}[G_{t+1} \mathbf{X}_{t+1} - G_{t+1} F_t \widehat{\mathbf{X}}_t + \varepsilon_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*] \\
&= G_{t+1} \mathbb{E}(\mathbf{X}_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*) - G_{t+1} F_t \mathbb{E}(\widehat{\mathbf{X}}_t | \mathbf{X}_{t+1}, \mathbf{Y}_t^*) + \mathbb{E}(\varepsilon_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*) \\
&= G_{t+1} \mathbf{X}_{t+1} - G_{t+1} F_t \widehat{\mathbf{X}}_t \\
&= G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t).
\end{aligned}$$

Portanto, temos que $\mathbb{E}(e_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*) = G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t)$.

Calculando a variância de e_{t+1} dado \mathbf{X}_{t+1} e dadas as observações até o tempo t , temos que:

$$\begin{aligned}
&Var(e_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*) \\
&= \mathbb{E}(e_{t+1}^2 | \mathbf{X}_{t+1}, \mathbf{Y}_t^*) - \mathbb{E}^2(e_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*) \\
&= \mathbb{E}[e_{t+1} e_{t+1}^T | \mathbf{X}_{t+1}, \mathbf{Y}_t^*] - [G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t)]^2 \\
&= \mathbb{E}\{[G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t) + \varepsilon_{t+1}][G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t) + \varepsilon_{t+1}]^T | \mathbf{X}_{t+1}, \mathbf{Y}_t^*\} \\
&\quad - [G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t)][G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t)]^T \\
&= \mathbb{E}\{[G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t) + \varepsilon_{t+1}]\{[G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t)]^T + \varepsilon_{t+1}^T\} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*\} \\
&\quad - [G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t)][(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t)^T G_{t+1}^T] \\
&= \mathbb{E}\{[G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t) + \varepsilon_{t+1}][(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t)^T G_{t+1}^T + \varepsilon_{t+1}^T] | \mathbf{X}_{t+1}, \mathbf{Y}_t^*\} \\
&\quad - [G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t)][(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t)^T G_{t+1}^T] \\
&= \mathbb{E}\{[G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t) + \varepsilon_{t+1}][(\mathbf{X}_{t+1}^T - (F_t \widehat{\mathbf{X}}_t)^T) G_{t+1}^T + \varepsilon_{t+1}^T] | \mathbf{X}_{t+1}, \mathbf{Y}_t^*\} \\
&\quad - [G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t)][(\mathbf{X}_{t+1}^T - (F_t \widehat{\mathbf{X}}_t)^T) G_{t+1}^T] \\
&= \mathbb{E}\{[G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t) + \varepsilon_{t+1}][\mathbf{X}_{t+1}^T G_{t+1}^T - (F_t \widehat{\mathbf{X}}_t)^T G_{t+1}^T + \varepsilon_{t+1}^T] | \mathbf{X}_{t+1}, \mathbf{Y}_t^*\} \\
&\quad - [G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t)][\mathbf{X}_{t+1}^T G_{t+1}^T - (F_t \widehat{\mathbf{X}}_t)^T G_{t+1}^T] \\
&= \mathbb{E}\{[G_{t+1} \mathbf{X}_{t+1} - G_{t+1} F_t \widehat{\mathbf{X}}_t + \varepsilon_{t+1}][\mathbf{X}_{t+1}^T G_{t+1}^T - \widehat{\mathbf{X}}_t^T F_t^T G_{t+1}^T + \varepsilon_{t+1}^T] | \mathbf{X}_{t+1}, \mathbf{Y}_t^*\} \\
&\quad - [G_{t+1} \mathbf{X}_{t+1} - G_{t+1} F_t \widehat{\mathbf{X}}_t][\mathbf{X}_{t+1}^T G_{t+1}^T - \widehat{\mathbf{X}}_t^T F_t^T G_{t+1}^T] \\
&= \mathbb{E}\{G_{t+1} \mathbf{X}_{t+1} \mathbf{X}_{t+1}^T G_{t+1}^T - G_{t+1} \mathbf{X}_{t+1} \widehat{\mathbf{X}}_t^T F_t^T G_{t+1}^T + G_{t+1} \mathbf{X}_{t+1} \varepsilon_{t+1}^T - G_{t+1} F_t \widehat{\mathbf{X}}_t \mathbf{X}_{t+1}^T G_{t+1}^T \\
&\quad + G_{t+1} F_t \widehat{\mathbf{X}}_t \widehat{\mathbf{X}}_t^T F_t^T G_{t+1}^T - G_{t+1} F_t \widehat{\mathbf{X}}_t \varepsilon_{t+1}^T + \varepsilon_{t+1} \mathbf{X}_{t+1}^T G_{t+1}^T - \varepsilon_{t+1} \widehat{\mathbf{X}}_t^T F_t^T G_{t+1}^T + \varepsilon_{t+1} \varepsilon_{t+1}^T | \mathbf{X}_{t+1}, \mathbf{Y}_t^*\} \\
&\quad - [G_{t+1} \mathbf{X}_{t+1} \mathbf{X}_{t+1}^T G_{t+1}^T - G_{t+1} \mathbf{X}_{t+1} \widehat{\mathbf{X}}_t^T F_t^T G_{t+1}^T - G_{t+1} F_t \widehat{\mathbf{X}}_t \mathbf{X}_{t+1}^T G_{t+1}^T + G_{t+1} F_t \widehat{\mathbf{X}}_t \widehat{\mathbf{X}}_t^T F_t^T G_{t+1}^T] \\
&= \mathbb{E}\{G_{t+1} \mathbf{X}_{t+1}^2 G_{t+1}^T - G_{t+1} \mathbf{X}_{t+1} \widehat{\mathbf{X}}_t^T F_t^T G_{t+1}^T + G_{t+1} \mathbf{X}_{t+1} \varepsilon_{t+1}^T - G_{t+1} F_t \widehat{\mathbf{X}}_t \mathbf{X}_{t+1}^T G_{t+1}^T \\
&\quad + G_{t+1} F_t \widehat{\mathbf{X}}_t^T F_t^T G_{t+1}^T - G_{t+1} F_t \widehat{\mathbf{X}}_t \varepsilon_{t+1}^T + \varepsilon_{t+1} \mathbf{X}_{t+1}^T G_{t+1}^T - \varepsilon_{t+1} \widehat{\mathbf{X}}_t^T F_t^T G_{t+1}^T + \varepsilon_{t+1}^2 | \mathbf{X}_{t+1}, \mathbf{Y}_t^*\} \\
&\quad - [G_{t+1} \mathbf{X}_{t+1}^2 G_{t+1}^T - G_{t+1} \mathbf{X}_{t+1} \widehat{\mathbf{X}}_t^T F_t^T G_{t+1}^T - G_{t+1} F_t \widehat{\mathbf{X}}_t \mathbf{X}_{t+1}^T G_{t+1}^T + G_{t+1} F_t \widehat{\mathbf{X}}_t^2 F_t^T G_{t+1}^T] \\
&= G_{t+1} \mathbb{E}(\mathbf{X}_{t+1}^2 | \mathbf{X}_{t+1}, \mathbf{Y}_t^*) G_{t+1}^T - G_{t+1} \mathbb{E}(\mathbf{X}_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*) \widehat{\mathbf{X}}_t^T F_t^T G_{t+1}^T
\end{aligned}$$

$$\begin{aligned}
& -G_{t+1}F_t\widehat{\mathbf{X}}_t\mathbb{E}(\mathbf{X}_{t+1}^T|\mathbf{X}_{t+1},\mathbf{Y}_t^*)G_{t+1}^T + G_{t+1}F_t\widehat{\mathbf{X}}_t^2F_t^TG_{t+1}^T + G_{t+1}\mathbb{E}(\mathbf{X}_{t+1}\varepsilon_{t+1}^T|\mathbf{X}_{t+1},\mathbf{Y}_t^*) \\
& -G_{t+1}F_t\widehat{\mathbf{X}}_t\mathbb{E}(\varepsilon_{t+1}^T|\mathbf{X}_{t+1},\mathbf{Y}_t^*) + \mathbb{E}(\varepsilon_{t+1}\mathbf{X}_{t+1}^T|\mathbf{X}_{t+1},\mathbf{Y}_t^*)G_{t+1}^T - \mathbb{E}(\varepsilon_{t+1}|\mathbf{X}_{t+1},\mathbf{Y}_t^*)\widehat{\mathbf{X}}_t^TF_t^TG_{t+1}^T \\
& + \mathbb{E}(\varepsilon_{t+1}^2|\mathbf{X}_{t+1},\mathbf{Y}_t^*) \\
& -[G_{t+1}\mathbf{X}_{t+1}^2G_{t+1}^T - G_{t+1}\mathbf{X}_{t+1}\widehat{\mathbf{X}}_t^TF_t^TG_{t+1}^T - G_{t+1}F_t\widehat{\mathbf{X}}_t\mathbf{X}_{t+1}^TG_{t+1}^T + G_{t+1}F_t\widehat{\mathbf{X}}_t^2F_t^TG_{t+1}^T] \\
& = G_{t+1}\mathbf{X}_{t+1}^2G_{t+1}^T - G_{t+1}\mathbf{X}_{t+1}\widehat{\mathbf{X}}_t^TF_t^TG_{t+1}^T - G_{t+1}F_t\widehat{\mathbf{X}}_t\mathbf{X}_{t+1}^TG_{t+1}^T + G_{t+1}F_t\widehat{\mathbf{X}}_t^2F_t^TG_{t+1}^T + H_{t+1} \\
& -[G_{t+1}\mathbf{X}_{t+1}^2G_{t+1}^T - G_{t+1}\mathbf{X}_{t+1}\widehat{\mathbf{X}}_t^TF_t^TG_{t+1}^T - G_{t+1}F_t\widehat{\mathbf{X}}_t\mathbf{X}_{t+1}^TG_{t+1}^T + G_{t+1}F_t\widehat{\mathbf{X}}_t^2F_t^TG_{t+1}^T] \\
& = H_{t+1}.
\end{aligned}$$

Portanto, temos que $Var(e_{t+1}|\mathbf{X}_{t+1},\mathbf{Y}_t^*) = H_{t+1}$.

Observação 5.12. A penúltima igualdade é obtida devido aos seguintes fatos:

- i) $\mathbb{E}(\varepsilon_{t+1}|\mathbf{X}_{t+1},\mathbf{Y}_t^*) = 0$, $\mathbb{E}(\varepsilon_{t+1}^T|\mathbf{X}_{t+1},\mathbf{Y}_t^*) = 0$, e $\mathbb{E}(\varepsilon_{t+1}^2|\mathbf{X}_{t+1},\mathbf{Y}_t^*) = H_{t+1}$;
- ii) $\mathbb{E}(\mathbf{X}_{t+1}|\mathbf{X}_{t+1},\mathbf{Y}_t^*) = \mathbf{X}_{t+1}$ o que implica que $\mathbb{E}(\mathbf{X}_{t+1}^2|\mathbf{X}_{t+1},\mathbf{Y}_t^*) = \mathbf{X}_{t+1}^2$;
- iii) Utilizamos também os mesmos argumentos dados na observação (5.10).

Desta maneira, a distribuição $\mathbb{P}(e_{t+1}|\mathbf{X}_{t+1},\mathbf{Y}_t^*)$ é dada por:

$$(e_{t+1}|\mathbf{X}_{t+1},\mathbf{Y}_t^*) \sim N(G_{t+1}(\mathbf{X}_{t+1} - F_t\widehat{\mathbf{X}}_t), H_{t+1}). \quad (5.27)$$

Assim, no tempo $t + 1$, quando uma nova observação Y_{t+1} torna-se avaliada, podemos atualizar o vetor de estado obtendo a nova distribuição *à posteriori* $\mathbb{P}(\mathbf{X}_{t+1}|\mathbf{Y}_{t+1}^*)$ a qual é equivalente à distribuição $\mathbb{P}(\mathbf{X}_{t+1}|e_{t+1},\mathbf{Y}_t^*)$. E essa última distribuição é obtida a partir do Teorema de Bayes, utilizando-se as distribuições de $(\mathbf{X}_{t+1}|\mathbf{Y}_t^*)$, e de $(e_{t+1}|\mathbf{X}_{t+1},\mathbf{Y}_t^*)$, as quais são dadas, respectivamente, pelas expressões (5.25) e (5.27). Usamos o Teorema de Bayes da seguinte forma:

$$\mathbb{P}(\mathbf{X}_{t+1}|Y_{t+1},\mathbf{Y}_t^*) = \frac{\mathbb{P}(e_{t+1}|\mathbf{X}_{t+1},\mathbf{Y}_t^*) \times \mathbb{P}(\mathbf{X}_{t+1}|\mathbf{Y}_t^*)}{\int_{\mathbf{X}_{t+1}} \mathbb{P}(e_{t+1},\mathbf{X}_{t+1}|\mathbf{Y}_t^*)d\mathbf{X}_{t+1}}. \quad (5.28)$$

Essa equação descreve a distribuição *à posteriori* quando avaliada a observação Y_{t+1} . Quando computada a distribuição *à posteriori* $\mathbb{P}(\mathbf{X}_{t+1}|\mathbf{Y}_{t+1}^*)$, retornamos ao passo inicial de recursão, dado na expressão (5.22), considerando-a como uma distribuição *à posteriori* inicial para o tempo $t + 1$. Iniciando, assim, um novo ciclo para o cálculo de outra nova distribuição *à posteriori* $\mathbb{P}(\mathbf{X}_{t+2}|\mathbf{Y}_{t+2}^*)$ utilizando para seu cálculo a distribuição *à priori* $\mathbb{P}(\mathbf{X}_{t+2}|\mathbf{Y}_{t+1}^*)$.

No entanto, calcular a distribuição $\mathbb{P}(\mathbf{X}_{t+1}|\mathbf{Y}_{t+1}^*)$ a partir da equação (5.28) é bastante complicado. Para driblar essa dificuldade, utilizamos um resultado conhecido da estatística multivariada, o qual é enunciado pela proposição abaixo.

Proposição 5.2. Sejam \mathbf{X}_1 e \mathbf{X}_2 vetores aleatórios com distribuição Normal multivariada conjunta, então:

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right) \quad (5.29)$$

se, e somente se,

$$\begin{cases} \mathbf{X}_2 \sim N(\mu_2, \Sigma_{22}) \\ \mathbf{X}_1 | \mathbf{X}_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X}_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}), \end{cases} \quad (5.30)$$

em que Σ_{ij} são as matrizes da covariância entre os vetores \mathbf{X}_i e \mathbf{X}_j , isto é, $\Sigma_{ij} = Cov(\mathbf{X}_i, \mathbf{X}_j)$.

E o vetor $\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ é o vetor das médias de \mathbf{X}_1 e \mathbf{X}_2 .

Agora, com o intuito de usar as distribuições de $(\mathbf{X}_{t+1} | \mathbf{Y}_t^*)$, e de $(e_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*)$, para obter a distribuição de $(\mathbf{X}_{t+1} | e_{t+1}, \mathbf{Y}_t^*)$ aplicaremos a Proposição 5.2, tomando \mathbf{X}_1 igual a e_{t+1} condicionado às observações Y_1, Y_2, \dots, Y_t , e tomando \mathbf{X}_2 igual a \mathbf{X}_{t+1} condicionado às observações Y_1, Y_2, \dots, Y_t . Deste modo, juntamente com a distribuição de $(\mathbf{X}_{t+1} | \mathbf{Y}_t^*)$ dada pela expressão (5.25) e com a primeira equação do sistema (5.30) temos que:

$$\begin{cases} \mu_2 = F_t \widehat{\mathbf{X}}_t \\ \Sigma_{22} = P_{t+1|t} \end{cases} \quad (5.31)$$

Com a distribuição de $(e_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*)$, dada pela expressão (5.27), em conjunto com a segunda equação da expressão (5.30) e com a expressão (5.31) obtemos:

$$\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X}_2 - \mu_2) = \mu_1 + \Sigma_{12}P_{t+1|t}^{-1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t) = G_{t+1}(\mathbf{X}_{t+1} - F_t \widehat{\mathbf{X}}_t).$$

Assim, concluímos que:

$$\begin{cases} \mu_1 = 0 \\ \Sigma_{12} = G_{t+1}P_{t+1|t}, \end{cases} \quad (5.32)$$

pois $\Sigma_{12}P_{t+1|t}^{-1} = G_{t+1}$.

Também, pela distribuição de $(e_{t+1} | \mathbf{X}_{t+1}, \mathbf{Y}_t^*)$, dada pela expressão (5.27), em conjunto com a segunda equação do sistema (5.30) e com a expressão (5.31) obtemos:

$$\begin{aligned} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} &= H_{t+1} \\ \Leftrightarrow \Sigma_{11} - G_{t+1}P_{t+1|t}P_{t+1|t}^{-1}(G_{t+1}P_{t+1|t})^T &= H_{t+1} \\ \Leftrightarrow \Sigma_{11} - G_{t+1}P_{t+1|t}P_{t+1|t}^{-1}P_{t+1|t}^T G_{t+1}^T &= H_{t+1} \\ \Leftrightarrow \Sigma_{11} - G_{t+1}P_{t+1|t}P_{t+1|t}^{-1}P_{t+1|t} G_{t+1}^T &= H_{t+1} \\ \Leftrightarrow \Sigma_{11} - G_{t+1}P_{t+1|t}G_{t+1}^T &= H_{t+1}. \end{aligned}$$

Logo, $\Sigma_{11} = H_{t+1} + G_{t+1}P_{t+1|t}G_{t+1}^T$.

Observação 5.13. Para obtenção de Σ_{11} , nos cálculos acima, utilizamos o fato de que a matriz de covariância $P_{t+1|t}$ é simétrica, isto é, $P_{t+1|t} = P_{t+1|t}^T$.

Deste modo, aplicando a Proposição 5.2, em que tomamos \mathbf{X}_1 igual a e_{t+1} , \mathbf{X}_2 igual a \mathbf{X}_{t+1} , e condicionando às observações $\mathbf{Y}_t^* = (Y_1, Y_2, \dots, Y_t)$ obtemos:

$$\left(\begin{bmatrix} e_{t+1} \\ \mathbf{X}_{t+1} \end{bmatrix} \middle| \mathbf{Y}_t^* \right) \sim N \left(\begin{bmatrix} 0 \\ F_t \widehat{\mathbf{X}}_t \end{bmatrix}, \begin{bmatrix} H_{t+1} + G_{t+1} P_{t+1|t} G_{t+1}^T & G_{t+1} P_{t+1|t} \\ P_{t+1|t} G_{t+1}^T & P_{t+1|t} \end{bmatrix} \right), \quad (5.33)$$

ou equivalentemente:

$$\left(\begin{bmatrix} \mathbf{X}_{t+1} \\ e_{t+1} \end{bmatrix} \middle| \mathbf{Y}_t^* \right) \sim N \left(\begin{bmatrix} F_t \widehat{\mathbf{X}}_t \\ 0 \end{bmatrix}, \begin{bmatrix} P_{t+1|t}^T & P_{t+1|t} G_{t+1}^T \\ G_{t+1} P_{t+1|t} & H_{t+1} + G_{t+1} P_{t+1|t} G_{t+1}^T \end{bmatrix} \right). \quad (5.34)$$

Aplicando a ida da Proposição 5.2, na expressão 5.34 acima, obtemos a distribuição de $(\mathbf{X}_{t+1} | e_{t+1}, \mathbf{Y}_t^*)$ dada por:

$$(\mathbf{X}_{t+1} | e_{t+1}, \mathbf{Y}_t^*) \sim N(\widehat{\mathbf{X}}_{t+1}, P_{t+1}),$$

em que,

$$\begin{aligned} \widehat{\mathbf{X}}_{t+1} &= F_t \widehat{\mathbf{X}}_t + \underbrace{P_{t+1|t} G_{t+1}^T (H_{t+1} + G_{t+1} P_{t+1|t} G_{t+1}^T)^{-1}}_{K_{t+1}} (e_{t+1} - 0) \\ &= F_t \widehat{\mathbf{X}}_t + K_{t+1} (e_{t+1}) = F_t \widehat{\mathbf{X}}_t + K_{t+1} (Y_{t+1} - \widehat{\mathbf{Y}}_{t+1|t}) \\ &= F_t \widehat{\mathbf{X}}_t + K_{t+1} (Y_{t+1} - G_{t+1} \widehat{\mathbf{X}}_{t+1|t}). \end{aligned}$$

Podemos reescrever essa expressão da seguinte maneira:

$$\widehat{\mathbf{X}}_{t+1} = \widehat{\mathbf{X}}_{t+1|t} + K_{t+1} (Y_{t+1} - G_{t+1} \widehat{\mathbf{X}}_{t+1|t}), \quad (5.35)$$

em que $K_{t+1} = P_{t+1|t} G_{t+1}^T (H_{t+1} + G_{t+1} P_{t+1|t} G_{t+1}^T)^{-1}$.

Também, temos que:

$$\begin{aligned} P_{t+1} &= P_{t+1|t} - \underbrace{P_{t+1|t} G_{t+1}^T (H_{t+1} + G_{t+1} P_{t+1|t} G_{t+1}^T)^{-1} G_{t+1} P_{t+1|t}}_{K_{t+1}} \\ &= P_{t+1|t} - K_{t+1} G_{t+1} P_{t+1|t} \\ &= (I - K_{t+1} G_{t+1}) P_{t+1|t}. \end{aligned}$$

Deste modo, obtemos seguinte equação:

$$P_{t+1} = (I - K_{t+1} G_{t+1}) P_{t+1|t}. \quad (5.36)$$

Substituindo a equação (5.24) na equação(5.36), temos que:

$$P_{t+1} = (I - K_{t+1} G_{t+1}) (F_t P_t F_t^T + R_{t+1} Q_{t+1} R_{t+1}^T), \quad (5.37)$$

Como a distribuição de $(\mathbf{X}_{t+1} | e_{t+1}, \mathbf{Y}_t^*)$ é equivalente a distribuição de $(\mathbf{X}_{t+1} | \mathbf{Y}_{t+1}^*)$, podemos obter a desejada distribuição *à posteriori*, a qual é dada por:

$$(\mathbf{X}_{t+1} | \mathbf{Y}_{t+1}^*) \sim N(\widehat{\mathbf{X}}_{t+1}, P_{t+1}),$$

em que $\widehat{\mathbf{X}}_{t+1}$ e P_{t+1} são expressos, respectivamente, pelas equações (5.35) e (5.37).

As equações (5.35) e (5.37) são usadas para atualizar a média e a matriz de covariância do vetor de estado \mathbf{X}_{t+1} quando uma nova observação Y_{t+1} torna-se avaliada.

A estimativa atualizada $\widehat{\mathbf{X}}_{t+1}$ do vetor de estado é a soma de uma estimativa projetada, usando observações até o tempo t , com o erro de previsão um passo a frente ponderado por uma matriz K_{t+1} , denominada Ganho de Kalman.

Podemos concluir que o filtro de Kalman é um procedimento recursivo que consiste em gerar uma estimativa do vetor estado, a qual será corrigida com base em uma nova observação para a obtenção de uma nova estimativa um passo a frente. O tamanho da correção é determinada pelo quão bem a estimativa preliminar prevê a nova observação.

Observação 5.14.

- i) O Ganho de Kalman, matriz K_{t+1} , pode ser vista como o coeficiente da Regressão de Mínimos Quadrados de \mathbf{X}_{t+1} sobre o erro de previsão e_{t+1} , condicionado às observações \mathbf{Y}_t^* .
- ii) Assim, a média da distribuição *à posteriori* de $(\mathbf{X}_{t+1} | e_{t+1}, \mathbf{Y}_t^*)$, $\widehat{\mathbf{X}}_{t+1}$, é a função de regressão de \mathbf{X}_{t+1} sobre e_{t+1} .

5.3.1 Resumo das recursões do Filtro de Kalman

Dado o sistema de espaço de estados:

$$\begin{aligned} Y_t &= G_t \mathbf{X}_t + \varepsilon_t, & \varepsilon_t &\sim N(0, H_t) \\ \mathbf{X}_{t+1} &= F_t \mathbf{X}_t + R_{t+1} \eta_{t+1}, & \eta_t &\sim N(0, Q_t) \end{aligned}, t \in \mathbb{Z},$$

em que $\mathbb{E}[\eta_t \varepsilon_s^T] = 0, \forall t, s \in \mathbb{Z}$.

Definimos:

$$\begin{aligned} \widehat{\mathbf{X}}_t &= \mathbb{E}[\mathbf{X}_t | Y_1, Y_2, \dots, Y_t] = \widehat{\mathbf{X}}_t = \mathbb{E}[\mathbf{X}_t | \mathbf{Y}_t^*]; \\ P_t &= \text{Var}(\mathbf{X}_t | \mathbf{Y}_t^*) = \mathbb{E}[(\mathbf{X}_t - \widehat{\mathbf{X}}_t)(\mathbf{X}_t - \widehat{\mathbf{X}}_t)^T]. \end{aligned}$$

A recursão inicia a partir da escolha, no tempo zero, das estimativas da média e da variância do vetor de estado inicial \mathbf{X}_0 , respectivamente, denotados por $\widehat{\mathbf{X}}_0$ e P_0 .

As estimativas do vetor de estado e da matriz de covariância, no tempo $t+1$, com observações avaliadas até o tempo t , $\widehat{\mathbf{X}}_{t+1|t}$ e $P_{t+1|t}$, são dadas pelas equações de atualização temporal, que são citadas a seguir.

Equações de Atualização Temporal

$$\begin{aligned} \widehat{\mathbf{X}}_{t+1|t} &= F_t \widehat{\mathbf{X}}_t; \\ P_{t+1|t} &= F_t P_t F_t^T + R_{t+1} Q_{t+1} R_{t+1}^T. \end{aligned}$$

Deste modo, temos que $\mathbb{P}(\mathbf{X}_{t+1} | \mathbf{Y}_t^*) \sim N(\widehat{\mathbf{X}}_{t+1|t}, P_{t+1|t})$,

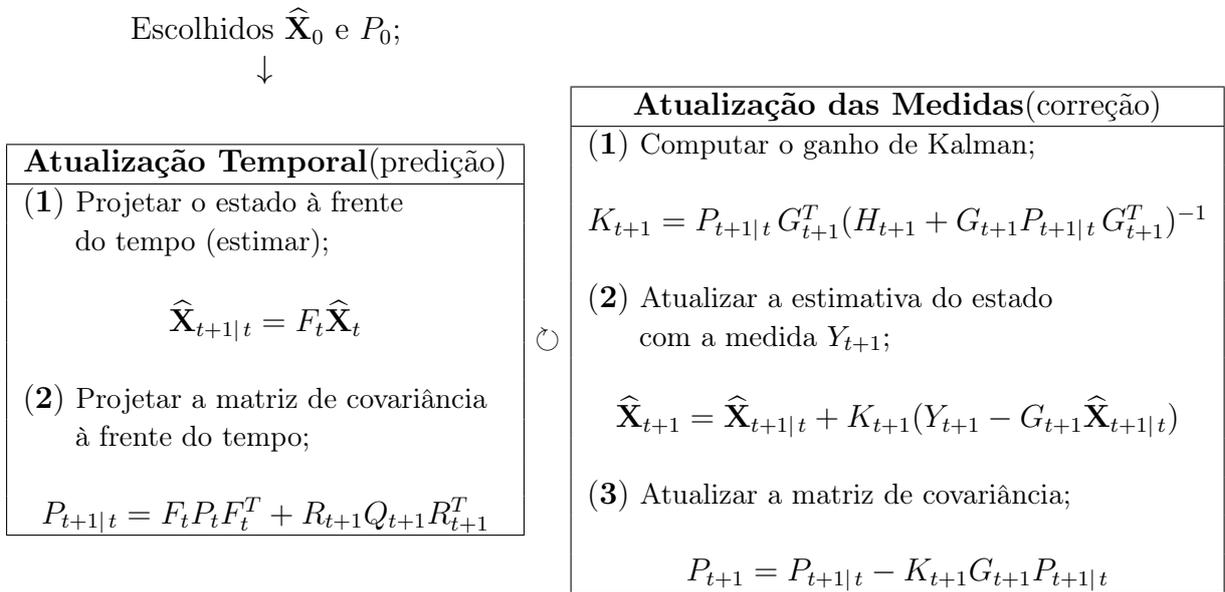
Seja $K_{t+1} = P_{t+1|t} G_{t+1}^T (H_{t+1} + G_{t+1} P_{t+1|t} G_{t+1}^T)^{-1}$ a matriz ganho de Kalman. Caso uma nova observação seja avaliada no tempo $t + 1$, então obtemos $\hat{\mathbf{X}}_{t+1}$ e P_{t+1} , a partir da correção, respectivamente, de $\hat{\mathbf{X}}_{t+1|t}$ e de $P_{t+1|t}$ pelas equações de atualização das medidas, as quais são citadas abaixo.

Equações de Atualização das Medidas

$$\begin{aligned}\hat{\mathbf{X}}_{t+1} &= \hat{\mathbf{X}}_{t+1|t} + K_{t+1}(Y_{t+1} - G_{t+1}\hat{\mathbf{X}}_{t+1|t}); \\ P_{t+1} &= P_{t+1|t} - K_{t+1}G_{t+1}P_{t+1|t}.\end{aligned}$$

Assim, temos que $\mathbb{P}(\mathbf{X}_{t+1} | \mathbf{Y}_{t+1}^*) \sim N(\hat{\mathbf{X}}_{t+1}, P_{t+1})$,

As recursões do filtro de Kalman seguem o esquema abaixo:



Observação 5.15. Em resumo, escolhidos $\hat{\mathbf{X}}_0$ e P_0 , a média e a matriz de covariância do estado inicial \mathbf{X}_0 , o filtro de Kalman, para cada $t \in \mathbb{Z}_{\geq}$, fornece uma estimativa $\hat{\mathbf{X}}_{t+1|t}$ do vetor de estado \mathbf{X}_{t+1} com base nas observações até o tempo t . Em seguida, essa estimativa é corrigida a partir da atual observação Y_{t+1} , fornecendo uma a correção $\hat{\mathbf{X}}_{t+1}$ que será utilizada para a atualização temporal da estimativa $\hat{\mathbf{X}}_{t+1|t}$, isto é, iniciando um novo ciclo com o cálculo da estimativa $\hat{\mathbf{X}}_{t+2|t+1}$ do vetor de estado \mathbf{X}_{t+2} com base nas observações até o tempo $t+1$.

Com a estimativa $\hat{\mathbf{X}}_{t+1|t}$ do vetor de estado \mathbf{X}_{t+1} , podemos obter uma estimativa para a observação Y_{t+1} , com base nas observações até o tempo t , da seguinte maneira:

Pela equação de observação (5.19), uma observação avaliada no tempo $t + 1$, Y_{t+1} , é descrita pela seguinte equação:

$$Y_{t+1} = G_{t+1}\mathbf{X}_{t+1} + \varepsilon_{t+1} \quad (5.38)$$

Agora, tomando a esperança em ambos os lados da equação (5.38) e condicionando a \mathbf{Y}_t^* , isto é, condicionando às observações até o tempo t , temos que:

$$\widehat{Y}_{t+1|t} = G_{t+1}\widehat{\mathbf{X}}_{t+1|t},$$

em que $\widehat{Y}_{t+1|t}$ é uma estimativa para a observação Y_{t+1} , com base nas observações até o tempo t .

Dispondo das n estimativas $(\widehat{Y}_{1|0}, \widehat{Y}_{2|1}, \dots, \widehat{Y}_{n|n-1})$ de uma amostra de observações, geradas pelas recursões do filtro de Kalman, (Y_1, Y_2, \dots, Y_n) , podemos calcular a função de verossimilhança do processo considerado (ver Subseção 5.3.2).

Dessa maneira, poderemos estimar os parâmetros dos processos a partir da maximização da função de Verossimilhança, como exemplo na literatura temos o estimador de Fox e Taquq (1986) que utiliza uma aproximação para a função de autocovariância do processo, proposta por Whittle (1951).

Para o caso de um processo representado por um modelo de espaço de estados gaussiano, dado pelas equações (5.19) e (5.20), queremos obter a função de Máxima Verossimilhança dado um vetor de parâmetros Θ que pode conter como elementos a serem estimados, por exemplo, os parâmetros do processo gerador das observações (Y_1, Y_2, \dots) . Neste trabalho, estamos interessados em estudar a estimação dos parâmetros de um processo k -Factor-GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), assim teríamos que $\Theta = \{\phi_1, \dots, \phi_p, \mathbf{u}, \boldsymbol{\lambda}, \theta_1, \dots, \theta_q\}$, em que $\mathbf{u} = (u_1, \dots, u_k)^T$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)^T$, $\phi(\cdot)$ e $\theta(\cdot)$ são os polinômios de graus p e q , respectivamente, definidos pelas equações dadas em (3.8).

Em seguida, encontramos uma expressão para a função de verossimilhança para um processo linear gaussiano representado em espaço de estados.

5.3.2 Função de Verossimilhança

Dada uma série temporal (Y_0, Y_1, \dots, Y_n) de um processo linear gaussiano, assumimos que os ruídos $\{\eta_t\}_{t \in \mathbb{Z}}$ e $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ são vetores normais independentes e que o vetor de estado inicial tem distribuição $N(\widehat{\mathbf{X}}_0, P_0)$ em que $\widehat{\mathbf{X}}_0$, e P_0 são conhecidos. Então, a função de verossimilhança é dada por:

$$L(\Theta; Y_0, Y_1, \dots, Y_n) = \mathbb{P}(Y_0, Y_1, \dots, Y_n) = \mathbb{P}(Y_0) \prod_{t=1}^n \mathbb{P}(Y_t | \mathbf{Y}_{t-1}^*).$$

Na prática utilizamos o logaritmo da função de verossimilhança:

$$\log L(\Theta; Y_0, Y_1, \dots, Y_n) = \sum_{t=0}^n \mathbb{P}(Y_t | \mathbf{Y}_{t-1}^*),$$

em que $\mathbb{P}(Y_0 | \mathbf{Y}_{-1}^*) = \mathbb{P}(Y_0)$.

Assim, para o modelo gaussiano de espaço de estados, dado pelas equações (5.19) e (5.20), temos que a distribuição de $(Y_t | \mathbf{Y}_{t-1}^*)$ é normal com média dada por:

$$\begin{aligned} \mathbb{E}(Y_t | \mathbf{Y}_{t-1}^*) &= \mathbb{E}(G_t \mathbf{X}_t + \varepsilon_t | \mathbf{Y}_{t-1}^*) \\ &= \mathbb{E}(G_t \mathbf{X}_t | \mathbf{Y}_{t-1}^*) + \mathbb{E}(\varepsilon_t | \mathbf{Y}_{t-1}^*) = G_t \mathbb{E}(\mathbf{X}_t | \mathbf{Y}_{t-1}^*) \\ &= G_t \widehat{\mathbf{X}}_{t|t-1} \end{aligned}$$

Também, temos que a matriz de covariâncias de $(Y_t|\mathbf{Y}_{t-1}^*)$ é dada por:

$$\begin{aligned}
\text{Var}(Y_t|\mathbf{Y}_{t-1}^*) &= \mathbb{E}(Y_t^2|\mathbf{Y}_{t-1}^*) - \mathbb{E}^2(Y_t|\mathbf{Y}_{t-1}^*) \\
&= \mathbb{E}[(G_t\mathbf{X}_t + \varepsilon_t)(G_t\mathbf{X}_t + \varepsilon_t)^T|\mathbf{Y}_{t-1}^*] - (G_t\widehat{\mathbf{X}}_{t|t-1})^2 \\
&= \mathbb{E}[(G_t\mathbf{X}_t(G_t\mathbf{X}_t)^T|\mathbf{Y}_{t-1}^*)] + \mathbb{E}(G_t\mathbf{X}_t\varepsilon_t^T|\mathbf{Y}_{t-1}^*) \\
&\quad + \mathbb{E}[\varepsilon_t(G_t\mathbf{X}_t)^T|\mathbf{Y}_{t-1}^*] + E(\varepsilon_t^2|\mathbf{Y}_{t-1}^*) - (G_t\widehat{\mathbf{X}}_{t|t-1})(G_t\widehat{\mathbf{X}}_{t|t-1})^T \\
&= G_t\mathbb{E}(\mathbf{X}_t^2|\mathbf{Y}_{t-1}^*)G_t^T + G_t\mathbb{E}(\mathbf{X}_t\varepsilon_t^T|\mathbf{Y}_{t-1}^*) \\
&\quad + \mathbb{E}(\varepsilon_t\mathbf{X}_t^T|\mathbf{Y}_{t-1}^*)G_t^T + H_t - G_t\widehat{\mathbf{X}}_{t|t-1}^2G_t^T \\
&= G_t\mathbb{E}(\mathbf{X}_t^2|\mathbf{Y}_{t-1}^*)G_t^T + H_t - G_t\mathbb{E}^2(\mathbf{X}_t|\mathbf{Y}_{t-1}^*)G_t^T \\
&= G_t[\mathbb{E}(\mathbf{X}_t^2|\mathbf{Y}_{t-1}^*) - E^2(\mathbf{X}_t|\mathbf{Y}_{t-1}^*)]G_t^T = G_t\text{Var}(\mathbf{X}_t|\mathbf{Y}_{t-1}^*)G_t^T + H_t \\
&= G_tP_{t|t-1}G_t^T + H_t.
\end{aligned}$$

Logo, $\mathbb{P}(Y_t|\mathbf{Y}_{t-1}^*) \sim N(G_t\widehat{\mathbf{X}}_{t|t-1}, G_tP_{t|t-1}G_t^T + H_t)$.

Portanto, a função densidade de probabilidade de $(Y_t|\mathbf{Y}_{t-1}^*)$ é dada por:

$$\begin{aligned}
\mathbb{P}(Y_t|\mathbf{Y}_{t-1}^*) &= (2\pi)^{-w/2}(\det[\text{Var}(Y_t|\mathbf{Y}_{t-1}^*)])^{-1/2} \exp\left\{-\frac{1}{2}[Y_t - \mathbb{E}(Y_t|\mathbf{Y}_{t-1}^*)]^T\right. \\
&\quad \times \left.(\text{Var}(Y_t|\mathbf{Y}_{t-1}^*))^{-1}[Y_t - \mathbb{E}(Y_t|\mathbf{Y}_{t-1}^*)]\right\}. \tag{5.39}
\end{aligned}$$

Substituindo as respectivas média e variância de $(Y_t|\mathbf{Y}_{t-1}^*)$, na equação (5.39), obtemos:

$$\begin{aligned}
\mathbb{P}(Y_t|\mathbf{Y}_{t-1}^*) &= (2\pi)^{-w/2}(\det[G_tP_{t|t-1}G_t^T + H_t])^{-1/2} \exp\left\{-\frac{1}{2}[Y_t - (G_t\widehat{\mathbf{X}}_{t|t-1})]^T\right\} \\
&\quad \times (G_tP_{t|t-1}G_t^T + H_t)^{-1}[Y_t - (G_t\widehat{\mathbf{X}}_{t|t-1})],
\end{aligned}$$

para $t \in \mathbb{Z}_{\geq}$.

Com isso, temos que a função de verossimilhança do modelo gaussiano de espaço de estados para uma série temporal (Y_0, Y_1, \dots, Y_n) é dada por:

$$\begin{aligned}
L(\Theta; Y_0, Y_1, \dots, Y_n) &= (2\pi)^{\frac{-(n+1)w}{2}} \left(\prod_{t=0}^n \det[G_tP_{t|t-1}G_t^T + H_t] \right)^{-1/2} \\
&\quad \times \exp\left\{-\frac{1}{2} \sum_{t=0}^n [Y_t - (G_t\widehat{\mathbf{X}}_{t|t-1})]^T (G_tP_{t|t-1}G_t^T + H_t)^{-1} [Y_t - (G_t\widehat{\mathbf{X}}_{t|t-1})]\right\}
\end{aligned}$$

O logaritmo da função de verossimilhança é dada por:

$$\begin{aligned}
\log L(\Theta; Y_0, Y_1, \dots, Y_n) &= \frac{-(n+1)w}{2} \log(2\pi) - \frac{1}{2} \sum_{t=0}^n \log(\det[G_tP_{t|t-1}G_t^T + H_t]) \\
&\quad - \frac{1}{2} \sum_{t=0}^n [Y_t - (G_t\widehat{\mathbf{X}}_{t|t-1})]^T (G_tP_{t|t-1}G_t^T + H_t)^{-1} [Y_t - (G_t\widehat{\mathbf{X}}_{t|t-1})],
\end{aligned}$$

para $t \in \mathbb{Z}_{\geq}$.

Observação 5.16. Na literatura, é bastante comum reescrevermos a função de verossimilhança em função do Erro de previsão e_t . Desta maneira, podemos reescrever a equação (5.39) por:

$$\mathbb{P}(Y_t | \mathbf{Y}_{t-1}^*) = (2\pi)^{-w/2} (\det[\text{Var}(Y_t | \mathbf{Y}_{t-1}^*)])^{-1/2} \exp \left\{ -\frac{1}{2} [e_t]^T \times (\text{Var}(Y_t | \mathbf{Y}_{t-1}^*))^{-1} [e_t] \right\}.$$

Isso se deve ao fato de que $Y_t - \mathbb{E}(Y_t | \mathbf{Y}_{t-1}^*) = Y_t - \widehat{Y}_{t|t-1} = e_t$.

Calculando a variância de e_t , obtemos:

$$\begin{aligned} \text{Var}(e_t) &= \text{Var}(Y_t - \widehat{Y}_{t|t-1}) \\ &= \mathbb{E}[Y_t - \widehat{Y}_{t|t-1} - \mathbb{E}(Y_t - \widehat{Y}_{t|t-1})]^2 \\ &= \mathbb{E}[Y_t - \widehat{Y}_{t|t-1}]^2 = \mathbb{E}[Y_t - \mathbb{E}(Y_t | \mathbf{Y}_{t-1}^*)]^2 \\ &= \text{Var}(Y_t | \mathbf{Y}_{t-1}^*). \end{aligned}$$

Deste modo, se denotarmos $\Sigma_t = \text{Var}(e_t)$, podemos reescrever o logaritmo da função de verossimilhança de um modelo gaussiano da seguinte maneira:

$$\log L(\Theta; Y_0, Y_1, \dots, Y_n) = \frac{-(n+1)w}{2} \log(2\pi) - \frac{1}{2} \sum_{t=0}^n [\log(\det \Sigma_t) + e_t^T (\Sigma_t)^{-1} e_t],$$

para $t \in \mathbb{Z}_{\geq}$.

Essa representação foi dada, primeiramente, por Schweppe (1965) e referida como a decomposição do erro de previsão por Harvey (1989).

Podemos notar que Σ_t e e_t são calculados recursivamente pelas equações do filtro de Kalman, assim, podemos computar o logaritmo da função de verossimilhança a partir das saídas desse filtro.

Da mesma forma, a partir das recursões do filtro de Kalman é possível calcular a função de verossimilhança e, a partir delas, implementar um algoritmo otimizador para o vetor de parâmetros desconhecidos Θ do processo. No entanto, implementar esse algoritmo para maximizar a função de verossimilhança para os modelos ARFIMA, GARMA e k -factor-GARMA, com parâmetros de longa dependência, parece ser inviável computacionalmente, visto que suas representações em espaço de estados têm dimensão infinita. No paper de Chan e Palma (1998), é mostrado que apesar de processos de longa dependência terem representações em espaço de estados de dimensão infinita, a função de máxima verossimilhança exata, pelo uso do filtro de Kalman, pode ser computada, recursivamente, em um número finito de passos.

A seguir, enunciamos um Teorema cuja demonstração é encontrada no paper de Chan e Palma (1998) (Teorema 2.2) referente à computação da função de máxima verossimilhança exata de uma amostra finita de um modelo ARFIMA(p, d, q).

Teorema 5.5. Seja $\{Y_0, Y_1, \dots, Y_{n-1}\}$ uma série temporal de n elementos do processo ARFIMA(p, d, q). Se P_0 é a variância do vetor de estado inicial \mathbf{X}_0 para uma representação de dimensão infinita, então a computação da função de verossimilhança depende apenas dos primeiros n componentes das equações de Kalman.

Como consequência dos resultados obtidos por Chan e Palma (1998), podemos concluir que, para os processos (causais) GARMA e k -Factor GARMA com parâmetros de longa dependência, é possível computar, recursivamente em um número finito de passos, a função de máxima verossimilhança exata utilizando o filtro de Kalman. Isto significa que, para a avaliação da função de máxima verossimilhança exata, necessitamos somente dos primeiros n componentes do vetor de estado. Assim, os infinitos componentes restantes desse vetor podem ser omitidos para as computações.

A seguir, estendemos os resultados do Teorema 2.2 de Chan e Palma (1998), para os processos GARMA e k -Factor GARMA.

Proposição 5.3. Seja $\{Y_0, Y_1, \dots, Y_{n-1}\}$ uma série temporal de n elementos do processo GARMA(p, u, λ, q)(causal). Se P_0 é a variância do vetor de estado inicial \mathbf{X}_0 para uma representação de dimensão infinita, então a computação da função de verossimilhança depende apenas dos primeiros n componentes das equações de Kalman.

Da mesma forma, é obtido o seguinte resultado para os processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$):

Proposição 5.4. Seja $\{Y_0, Y_1, \dots, Y_{n-1}\}$ uma série temporal de n elementos do processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$)(causal). Se P_0 é a variância do vetor de estado inicial \mathbf{X}_0 para uma representação de dimensão infinita, então a computação da função de verossimilhança depende apenas dos primeiros n componentes das equações de Kalman.

As demonstrações das Proposições 5.3 e 5.4 são equivalentes a demonstração do Teorema 2.2 de Chan e Palma (1998) e por esse motivo serão omitidas.

Para mais detalhes sobre a função de máxima verossimilhança de um processo gaussiano, ver Brockwell e Davis (2006), Morettin e Toloi (2004), Chan e Palma (1998) e Palma (2007).

Capítulo 6

Conclusões e Trabalhos Futuros

6.1 Conclusões

Em suma, verificamos com este trabalho que os processos fracionários generalizados são uma generalização dos processos ARFIMA(p, d, q). Os processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), em particular, são processos que apresentam k singularidades na função densidade espectral no intervalo $[0, \pi]$, as quais ocorrem não necessariamente na origem. Em contraste, o processo precursor ARFIMA(p, d, q) apresenta apenas uma ilimitação nesse intervalo, a qual ocorre na origem. Deste modo, os processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) podem modelar uma quantidade maior de séries temporais comparado com os modelos antecessores.

Por outro lado, observamos que filtro de Kalman é um procedimento recursivo formado por um conjunto de equações que, a cada iteração, fornece uma estimativa do vetor de estados não observado de um sistema de espaço de estados. Além disso, temos que um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) causal pode ser representado por um sistema de espaço de estados. Deste modo, podemos utilizar as estimativas do vetor de estados produzidas pelo filtro de Kalman para gerar as estimativas das observações de uma série temporal do processo, possibilitando o cálculo da função de verossimilhança dessa série. No Capítulo 4, apresentamos como representar os processos fracionários generalizados causais em sistema de espaço de estados.

Podemos notar que um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) causal, quando apresenta a propriedade de longa dependência, tem representação em sistema de espaço de estados de dimensão infinita. Esse fato, do ponto de vista computacional, acarreta na criação de um algoritmo para a computação da função de verossimilhança que necessita de uma imensa quantidade de cálculos, o que é bastante inviável. Por esse motivo, estendemos o Teorema 2.2 de Chan e Palma (1998) para os processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), que nos permite, com um número finito de passos, computar, recursivamente, a função de verossimilhança exata desse processo com uso do filtro de Kalman.

Portanto, podemos concluir que a partir da representação em espaço de estados de um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) causal podemos obter, recursivamente, a função de verossimilhança desse processo através do filtro de Kalman em um número finito de passos, inclusive quando a representação do processo tiver dimensão infinita. Com isso, podemos maximizar a função de verossimilhança como função do vetor de parâmetros desconhecidos do processo $\Theta = \{\phi_1, \dots, \phi_p, \mathbf{u}, \boldsymbol{\lambda}, \theta_1, \dots, \theta_q\}$, em que $\phi(\cdot)$ e $\theta(\cdot)$ são os polinômios de graus p e q , respectivamente, definidos pelas equações dadas em (3.8), com a finalidade de encontrar estimadores

para estes parâmetros.

6.2 Trabalhos Futuros

Devido à constatação de que é possível obtermos a função de verossimilhança dos processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) causal via modelo de espaço de estados, é de grande interesse a implementação de um algoritmo otimizador dessa função para um trabalho futuro, com a finalidade de calcular os estimadores para os parâmetros do modelo. A seguir, citamos algumas idéias para novos trabalhos:

- i) Estender o estimador de Fox e Taqqu (1986) e Beran (1994) para os processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$).
- ii) Comparar nosso estimador (via espaço de estados e filtro de Kalman) com o estimador de máxima verossimilhança proposto por Fox e Taqqu (1986).
- iii) Através de Simulações de Monte Carlo, comparar os vários estimadores utilizando o vício, o erro quadrático médio e a variância.
- iv) Fazer previsão dos futuros valores de uma série temporal.
- vi) Aplicação da teoria a séries de dados reais.

Bibliografia

- [1] Akaike, H. (1974). *A New Look at the Statistical Model Identification*. IEEE Transactions on Automatic Control. AC 19, 716-723.
- [2] Baillie, R.T. (1996). *Long Memory Processes and Fractional Integration in Econometrics*. Journal of Econometrics. Vol. 73, 5-59.
- [3] Beran, J. (1994). *Statistics for Long-Memory Processes*. New York: Chapman & Hall.
- [4] Bisognin C. (2003). *Estimação e Previsão em Processos de Longa Dependência Sazonal*. Dissertação de Mestrado do Programa de Pós-graduação da UFRGS, Porto Alegre, RS, Brasil.
- [5] Bisognin C. (2007). *Estimação e Previsão em Processos SARFIMA(p, d, q) \times (P, D, Q)_s na Presença de Outliers*. Tese de Doutorado do Programa de Pós-graduação da UFRGS, Porto Alegre, RS, Brasil.
- [6] Bloomfield, P. (1992). *Trends in Global Temperature*. Climatic Change. Vol. 68, 165-176.
- [7] Box, G. E. P. e G. M. Jenkins (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day (Revised edition, 1976).
- [8] Box, G. E. P., G. M. Jenkins e G. C. Reinsel (1994). *Time Series Analysis: Forecasting and Control*. New Jersey: Prentice Hall.
- [9] Brockwell, P. J. e R. A. Davis (2006). *Time Series. Theory and Methods*. 2ª ed.- New York: Springer-Verlag.
- [10] Chan, N. H. e W. Palma (1998). *State Space Modeling de Long-memory Processes*. Ann. Statist. Vol. 22, No. 2, 719-740.
- [11] Cheung, Y.-W. e F. X. Diebold (1994). *On Maximum Likelihood Estimation of the Differencing Parameter of Fractionally-integrated Noise with Unknown Mean*. Journal of Econometrics, Elsevier. Vol. 62(2), 317-350.
- [12] Chung, C-F. (1996). *Estimating a Generalized Long Memory Process*. Journal of Econometrics. Vol. 73, 237-259.
- [13] Collet, J., D. Guégan e P. Valdes-Sosa (2003). *How Shall We Determine the Number and the Location of the Gegenbauer Frequencies? An Empirical Approach*. Note de Recherche IDHE-MORA n° 2003-09.
- [14] Durbin, C. J. e S. J. Koopman (2001). *Time Series Analysis by Estate Space Methods*. New York: Oxford University Press.
- [15] Ferrara, L. e D. Guégan (1999). *Estimation and Applications of Gegenbauer Processes*. Preprint INSEE n° 9927.
- [16] Ferrara, L. e D. Guégan (2001). *Forecasting with k-factor Gegenbauer processes. Theory and Applications*. Journal of Forecasting. Vol. 20, 581-601.

- [17] Ferrara¹, L. e D. Guégan² (2008). *Forecasting with k-Factor Gegenbauer Process. Theory and applications*. Technical Report , ¹Universidade de Paris, ²Universidade de Reims, França.
- [18] Fox, R. e M.S. Taquq (1986). *Large-sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series*. The Annals of Statistics. Vol. 14, 517-532.
- [19] Geweke, J. e S. Porter-Hudak (1983). *The Estimation and Application of Long Memory Time Series Model*. Journal of Time Series Analysis. Vol. 4(4), 221-238.
- [20] Giraitis L. e R. Leipus (1995). *A Generalized Fractionally Differencing Approach in Long-memory Modelling*. From Department of Statistical Science. Southern Methodist University. Dallas, Texas, U.S.A.
- [21] Granger, C. M. G. e R. Joyeux (1980). *An Introduction to Long Memory Time Series Models and Fractional Differencing*. Journal of Time Series Analysis. Vol. 1, 15-29.
- [22] Gray, H. L., N-F. Zhang e W.A. Woodward (1989). *On Generalized Fractional Processes*. Journal of Time Series Analysis, Vol. 10(3), 233-257.
- [23] Hamilton, J. D. (1994). *Time Series Analysis*. New Jersey: Princeton University Press.
- [24] Hannan, E. J. e M. Deistler (1988). *Statistical Theory of Linear Systems*. New York: Wiley.
- [25] Harrison, P. J. e C. F. Stevens (1976). *Bayesian Forecasting (with discussion)*. Journal of the Royal Statistical Society, Serie B. Vol. 38, 205-247.
- [26] Hoffman, K. e R. Kunze (1971). *Linear Algebra*. 2^a ed.- New Jersey: Prentice Hall.
- [27] Hosking, J. R. M. (1981). *Fractional Differencing*. Biometrika. Vol. 68, 165-176.
- [28] Hosking, J. R. M. (1984). *Modelling Persistence in Hydrological Time Series Using Fractional Differencing*. Water Resources Research. Vol. 20(12), 1898-1908.
- [29] Hurst, H. E. (1951). *Long-term Storage Capacity de Reservoirs*. Transactions of the American Society of Civil Engineers. Vol. 16, 770-799.
- [30] Hurst, H. E. (1957). *A Suggested Statistical Model of Time Series that Occur in Nature*. Nature. Vol. 180, 494.
- [31] Hurvich, C.M. e B.K. Ray (1995). *Estimation of the Memory Parameter for Nonstationary or Noninvertible Fractionally Integrated Processes*. Journal of Time Series Analysis. Vol. 16(1), 17-042.
- [32] James, B. R. (2008). *Probabilidade: um curso em nível intermediário*. Rio de Janeiro. Coleção Projeto Euclides, IMPA.
- [33] Jones, R. H. (1984). *Fitting Multivariate Models to Unequally Spaced Data*. In Time Series Analysis of Irregularly Observed Data (E. Parzen, ed.).158-188. Lecture Notes in Statistics. Vol. 25. New York: Springer.
- [34] Kalman, R. E. (1960). *A New Approach to Linear Filtering and Prediction Problems*. Trans. ASME J. Basic Eng.. Vol. 82, 35-45.
- [35] Kalman, R. E. e R. S. Bucy (1961). *New Results in Filtering and Prediction Theory*. Trans. ASME J. Basic Eng.. Vol. 83, 95-108.
- [36] Kitagawa, G. e W. Gersch (1984). *A Smoothness Priors Modeling of Time Series with Trend and Seasonality*. Journal of the American Statistical Association. Vol. 79, 378-389.
- [37] Lopes, S.R.C., B.P. Olbermann e V.A. Reisen (2004). *A Comparison of Estimation in Non-Stationary ARFIMA Processes*. Journal of Statistical Computation and Simulation. Vol. 74(5), 339-347.

- [38] Lopes, S.R.C. e B.V.M. Mendes (2006). *Bandwidth Selection in Classical and Robust Estimation of Long Memory*. International Journal of Statistics and Systems. Vol. 1(2), 167-190.
- [39] Lopes, S.R.C. (2007). *Long-range Dependence in Mean and Volatility. Models, Estimation and Forecasting*. In: Vares M.E. e V. Sidoravicius (eds). *EBP*. Rio de Janeiro: Birkhäuser.
- [40] Magalhães, M. N. (2006). *Probabilidade e Variáveis Aleatórias*. 2ª ed.- São Paulo: Edusp.
- [41] Mandelbrot, B. B. (1965). *Une Classe de Processus Stochastiques Homothétiques à Soi. Application à la Loi Climatologique de H. E. Hurst*. Comptes Rendus (Paris). Vol. 260, 3274-3277.
- [42] Mandelbrot, B. B. e J. R. Wallis (1968). *Noah, Joseph and Operational Hydrology*. Water Resources Research. Vol. 4, 909-918.
- [43] Mandelbrot, B. B. e J.W. Van Ness (1968). *Fractional Brownian Motion, Fractional Noises and Applications*. SIAM Review. Vol. 10, 422-437.
- [44] Mandelbrot, B. B. e J.R. Wallis (1969). *Computer Experiments with Fractional Gaussian Noises. Part One, Averages and Variances*. Water Resources Research. Vol. 5(1), 228-241.
- [45] McLeod, A. I. e K. W. Hipel (1978). *Preservation of the Rescaled Adjusted Range, 1. A Reassessment of the Hurst Phenomenon*. Water Resources Research. Vol. 14, 491-508.
- [46] Meinhold, R. J. e N. D. Singpurwalla (1983). *Understanding the Kalman Filter* The American Statistician. Vol. 37, No. 2, 123-127.
- [47] Morettin, P. A. e C. M. C. Tolo (2004). *Análise de séries Temporais*. São Paulo: Edgard Blücher.
- [48] Olive, D. J.(2008). *A Course in Statistical Theory*. Department of Mathematics, Southern Illinois University.
- [49] Ooms, M (1995). *Flexible Seasonal Long Memory and Economic Time Series*. Preprint of the Econometric Institute. Rotterdam: Erasmus University.
- [50] Palma, W. (2007). *Long-Memory Time Series. Theory and Methods*. New Jersey: Wiley.
- [51] Peiris, M.S. e N. Singh (1996). *Predictors for Seasonal and Nonseasonal Fractionally Integrated Arima Models*. Biometrika. Vol. 38(6), 741-752.
- [52] Reisen, V.A. e S.R.C. Lopes (1999). *Some Simulations and Applications of Forecasting Long-Memory Time Series Models*. Journal of Statistical Planning and Inference. Vol. 80(2), 269-287.
- [53] Robinson, P. M. (1994). *Time Series with Strong Dependence*. In Advances in Econometrics, Sixth World Congress (C. A. Sims, ed.). Vol 1, 47-95. Cambridge Univ. Press.
- [54] Robinson, P.M. (1995). *Log-Periodogram Regression of Time Series with Long Range Dependence*. Annals of Statistics. Vol. 23(3), 1048-1072.
- [55] Rohatgi, V. K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*. New York: John Wiley and Sons.
- [56] Shumway, R. H. e D. S. Stoffer (1982). *An Approach to Time Series Smoothing and Forecasting using the EM Algorithm*. Journal of Time Series Analysis. Vol. 3, 253-264.
- [57] Shumway, R. H. (1985) *Time Series in the Soil Sciences: Is there life after kriging? Soil Spatial Variability*. (J. Boouma and D.R. Nielson, eds.). 35-60. The Netherlands: Pudoc Wageningen.

- [58] Sowell, F. (1992). *Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models*. Journal of Econometrics. Vol. 53, 165-188.
- [59] Wei, W. (2006). *Time Series Analysis. Univariate and Multivariate Methods*. 2^a ed.- New York: Addison Wesley.
- [60] Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. New York: Hafner.
- [61] Woodward, W.A., Q.C. Cheng e H.L. Gray (1998). *A k-Factor GARMA Long-Memory Model*. Journal of Time Series Analysis, Vol. 19(4), 485-504.
- [62] Yajima Y. (1996). *Estimation of the Frequency of Unbounded Spectral Densitie*. Proceedings of the Business and Economic Statistics. Section 4-7. American Statistical Association.