



CLAYTON BACELAR DE LIMA

**FUNDAMENTAÇÃO E APLICAÇÃO DE PROCESSOS
ESTOCÁSTICOS**

Santo André, 2013



UNIVERSIDADE FEDERAL DO ABC

CENTRO DE MATEMÁTICA, COMPUTAÇÃO E COGNIÇÃO

CLAYTON BACELAR DE LIMA

**FUNDAMENTAÇÃO E APLICAÇÃO DE PROCESSOS
ESTOCÁSTICOS**

Orientador: Prof. Dr. André Ricardo Oliveira Fonseca

Dissertação de mestrado apresentada ao Centro de
Matemática, Computação e Cognição para
obtenção do título de Mestre em Matemática

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO
DEFENDIDA PELO ALUNO CLAYTON BACELAR DE LIMA,
E ORIENTADA PELO PROF. DR. ANDRÉ RICARDO OLIVEIRA FONSECA.

SANTO ANDRÉ, 2013

Folha de Aprovação

Substituir pelo pdf da folha de aprovação assinada pela banca.

Dedico essa tese a todos que estiveram e estão ao meu lado durante essa caminhada, principalmente, a minha mãe que já se foi e minha família (meu pai, minha mulher e minhas filhas)

Agradecimentos

Agradeço ao meu orientador, Prof André Fonseca e os professores do Profmat Rafael Grisi, Daniel Miranda, Armando Caputti e João Carlos Ferreira.

Procura ser feliz na eternidade,
que o mundo são brevíssimos instantes.
Alvarenga Peixoto

Resumo

Palavras Chaves: Processos Estocásticos, Ensino Fundamental, Ensino Médio.

Esta dissertação visa apresentar sequências de atividades didáticas explorando os métodos de análise de séries temporais utilizando os conhecimentos estudados na educação básica. Uma das justificativas de estudar esse tema vem das orientações curriculares que afirma: "Ao final do ensino médio, espera-se que os alunos saibam usar a Matemática para resolver problemas práticos do cotidiano; para modelar fenômenos em outras áreas do conhecimento". Assim, o trabalho, sugere atividades que abrangem principalmente os conteúdos de estatística e probabilidade, discute as novas tecnologias na formação do pensamento estocástico dos alunos e seus desdobramentos. Este trabalho é dividido em quatro partes: A primeira seção é constituída de uma revisão bibliográfica sobre séries temporais e processos estocásticos; na segunda parte apresentamos a introdução, objetivos e justificativa; na terceira seção apresentamos um plano de aula com sequências didáticas utilizando alguns softwares. Essas atividades abordam vários conceitos que servem de base para compreender métodos de identificação de séries temporais. Através do método de simulação e interação utilizaremos alguns softwares como o R, Gretl e a planilha eletrônica. Concluimos a quarta parte apresentando uma discussão final sobre o ensino de processos estocásticos por meio da recursos tecnológicos.

Abstract

This dissertation aims to present a sequence of didactic activities exploring time series analysis methods using basic education knowledge. Thus, following curricular orientations, those activities cover mainly statistical and probability contents. This paper is divided into four parts: The first section is composed of a temporal stochastic processes bibliographic review and the second part presents the introduction, objectives and justification; in the third section we present a lesson plan with teaching sequences using some software. Those activities address several concepts related to the basic comprehension of identifying time series. Through the simulation and interaction methods, we have used some software such as R, Gretl and electronic spreadsheet. We conclude the fourth part presenting a final discussion about teaching stochastic processes through technological resources.

Conteúdo

1	FUNDAMENTAÇÃO TEÓRICA	1
1.1	introdução	1
1.2	Definições	3
1.2.1	Variável Aleatória	3
1.2.2	Variância, Covariância e Correlação	3
1.3	Séries temporais	5
1.3.1	Processos estocásticos	6
1.3.2	Especificação de um Processo Estocástico	6
1.3.3	Propriedades da Função de Auto-covariância	8
1.3.4	Processos Lineares Estacionários	9
1.3.5	Processos Não Estacionários	16
1.3.6	Modelagem ARIMA	17
1.4	Resumo	28
2	PRÁTICA	30
2.1	Introdução	30
2.2	Considerações iniciais	31
3	PLANO DE AULA	32
4	CONCLUSÃO	39
4.1	Discussão (outros trabalhos)	39
4.2	Considerações finais	41
	Referência Bibliográfica	43

1 FUNDAMENTAÇÃO TEÓRICA

1.1 introdução

O tema de nosso estudo trabalha com observações de fenômenos físicos, químicos, da medicina e da economia. Serve como exemplo os índices da bolsa de valores de São Paulo, valores mensais de temperatura na cidade de Uberlândia, índice de custo de vida na cidade de Salvador, etc. Esses são exemplos de séries temporais que (Morettin 2005) define como qualquer conjunto de observações ordenadas no tempo. Existem duas perspectivas na análise de séries temporais e nas duas envolve a construção de modelos para séries com propósitos diferentes, no caso do primeiro a análise é realizada no campo do tempo sendo apresentados modelos paramétricos (número finito e parâmetros) já no segundo, o domínio fica no campo da frequência utilizando modelos não paramétricos. Dentro do campo de frequências existe a análise espectral que deve basear-se na decomposição da série em componentes de frequência e a presença do espectro é o atributo essencial. No que se refere aos modelos paramétricos podemos colocar como exemplo os modelos ARIMA. Para facilitar algumas colocações adotaremos uma representação de uma observação de um fenômeno físico por exemplo como no caso da temperatura do mar durante certo período de horas do dia. Chamaremos de $Z(t)$ a temperatura observada no instante t . Se considerarmos que a temperatura foi obtida em três dias diferentes terão trajetórias (curvas) que normalmente não são iguais. Este processo que chamaremos de processo estocástico é o conjunto de trajetórias que conseguimos observar e cada trajetória também é designada de série temporal ou função amostral. Denota-se por $Z^{(1)}(5)$ a temperatura no instante $t=5$, para a primeira trajetória e deste modo para cada trajetória j no instante t teremos um número real $Z^{(j)}(t)$. Para cada t fixo, teremos os valores de uma variável aleatória $Z(t)$, que terá certa distribuição de probabilidades. Vale resaltar que temos situações como no exemplo aqui citado que é possível obter várias trajetórias e outras situações que não há possibilidade de se fazer vários testes e assim só há uma trajetória. Apesar de normalmente t ser mencionado como tempo a série $Z(t)$ poderá ser função de outros parâmetros físicos, como por exemplo espaço e volume e quando isto ocorrer a série é considerada multivariada e multidimensional. Uma vez adquirido as variáveis aleatórias qual o interesse em analisar as séries temporais? Citaremos então os objetivos principais:

- Fazer previsões de valores futuros da série, a curto prazo ou a longo prazo;
- Investigar do mecanismo gerador da série, como o caso de uma série de altura de ondas quanto se pretende investigar como estas foram geradas;
- Descrever o comportamento da série, através de construção de gráficos para, por exemplo, verificar os ciclos.

De maneira geral é comum supor uma série temporal estacionária, ou seja, ela se desenvolve em torno de uma média constante, representando alguma forma de estabilidade.

1 FUNDAMENTAÇÃO TEÓRICA

Porém na prática a maioria das séries refletem uma forma de não-estacionariedade o que nas áreas econômicas e financeiras são chamadas de tendências, o caso mais simples é a tendência linear, onde a série flutua ao redor de uma reta, com inclinação ou não. A classe de modelos ARIMA, já mencionada, tem condição de representar de maneira adequada séries estacionárias e séries não estacionárias que não apresentam comportamento explosivo (homogêneo). Uma transformação muito utilizada quando a série é não estacionária é tomar diferenças sucessivas da série original, até que se consiga uma série estacionária. Assim

$$\Delta Z(t) = Z(t) - Z(t - 1)$$

, é primeira diferença de $Z(t)$ e

$$\Delta^2 Z(t) = \Delta[\Delta Z(t)] = \Delta[Z(t) - z(t - 1)].$$

Deste modo, a n -ésima diferença de $Z(t)$ é

$$\Delta^n Z(t) = \Delta[\Delta^{n-1} Z(t)]$$

Na prática, toma-se uma ou duas vezes diferenças para que a série se torne estacionária. Como já foi citado um dos principais objetivos da análise de séries temporais é realizar previsões e isto conduz a obtenção de um método ou procedimento. Alguns autores como Priestley (1979) sugerem que quase todos os procedimentos estão virtualmente fundamentado na previsão dos mínimos quadrados, os quais são obtidos através de diferentes procedimentos computacionais a partir de combinações lineares de valores passados. Então é necessário obter uma função-perda além do modelo para se chegar a um procedimento. A função-perda mais utilizada é o erro quadrático médio, apesar em algumas situações, outras metodologias possam ser mais adequadas. Para ficar mais claro suponha que obtemos observações de uma série temporal até o instante t e desejamos prever o valor da série no instante $t+h$. Considerando $\hat{Z}_t(h)$ é a previsão de $Z(t+h)$, de origem t e horizonte h . O erro quadrático médio da previsão é

$$E[z(t+h) - \hat{z}_t(h)]^2.$$

Então dado o modelo que retrata a série temporal no instante t e dado que desejamos minimizar o erro quadrático médio, obtemos uma fórmula para $\hat{Z}_t(h)$. Normalmente os procedimentos de previsão variam muito, podendo ser simples e intuitivo, no qual existe nenhuma análise de dados envolvida, ou mais quantitativos e complexos, chegando a uma análise significativa. Em economia existem dois métodos mais utilizados: econométrico e de séries temporais. No primeiro, a previsão é feita com base nas teorias econômicas para construir um modelo, incluindo muitas variáveis, e no segundo o modelo é construído pelos dados estatísticos sem relação com teorias econômicas. Para Ashley e Granger (1979) a abordagem tem ser híbrida, porém para outros autores os modelos econométricos se ajustam melhor aos dados, enquanto que outros modelos como os ARIMA fornecem melhor previsões. Nosso principal objetivo aqui é apresentar a modelagem ARIMA, método de análise e previsão de séries temporais que se destaca pela sua ampla divulgação e utilização, classificado como não automático, o qual exige a intervenção de pessoal especializado para serem aplicados e também pela sua aparente superioridade em várias situações.

1.2 Definições

O objetivo desta seção é apresentar alguns conceitos introdutórios sobre séries temporais apresentando quando possíveis alguns exemplos. Começamos apresentando então:

1.2.1 Variável Aleatória

Suponhamos que a cada ponto de um espaço amostral se atribua um número. Teremos uma função definida no espaço amostral. Esta função é chamada variável aleatória (ou variável estocástica) ou ainda função aleatória (ou função estocástica). Isso corresponde a representar os resultados de um experimento aleatório por meio de números melhorando a tratamento matemático. Denota-se em geral por uma letra maiúscula X, Y, \dots .

EXEMPLO: Jogando-se uma moeda duas vezes, o espaço amostral é $\tau = \{cc, ck, kc, kk\}$ onde c =cara e K =coroa e X é o número de "caras" que aparecem. A cada ponto amostral associamos um número para X , tal como se vê na Tábua 1. Assim, por exemplo, no caso "cc" (2 caras), $X=2$, enquanto que para "kc" (1 cara) $X=1$.

Tabela 1.1: Tábua 1

Ponto Amostral	cc	ck	kc	kk
X	2	1	1	0

Neste espaço amostral, poderíamos definir muitas outras variáveis aleatórias como, por exemplo, o número de "caras" menos o número de "coroas" ou ainda o quadrado de número de "caras". Uma variável aleatória que adota um número infinito ou um número finito enumerável de valores, é chamada variável discreta, ao passo que uma variável aleatória que adota um número infinito (não enumerável) de valores é uma variável aleatória contínua.

1.2.2 Variância, Covariância e Correlação

Antes de apresentarmos algumas definições é importante definirmos o que é média ou esperança. Para uma série de dados $X_1, X_2, X_3, \dots, X_n$ onde n é o número de observações segue que:

$$\text{Média} = E(x) = \mu_x = \frac{\sum_{i=1}^n X_i}{n}$$

Covariância vem de variância conjunta. A variância que é uma medida de dispersão mede em geral a distância dos seus valores em relação à média, ou o quanto seus valores se afastaram de sua média. A variância é calculado primeiro pela soma dos desvios quadrados da média, e dividindo-a pelo número de observações (se os dados representam a população toda) ou por este número, reduzido por um (se os dados representam uma amostra).

$$\text{Variância} = E[(X - \mu)^2] = \sigma_x^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$$

Quando existirem duas séries de dados e se pretende tomar como as duas variáveis se movem juntas a medida mais utilizada é a covariância. Covariância ou variância conjunta é uma medida do grau de inter-relação numérica entre duas variáveis

1 FUNDAMENTAÇÃO TEÓRICA

aleatórias. Pode também ser chamada de medida de dependência linear entre duas variáveis aleatórias o que segue se as variáveis forem independentes a covariância é zero. A covariância fornece uma medida não padronizada do grau no qual elas se movem juntas, e é calculada tomando o produto dos desvios da média para cada variável em cada período.

$$\text{Covariância} = \text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = \sigma_{xy} = \frac{\sum_{i=1}^n (X_i - \mu_x)(Y_i - \mu_y)}{n}$$

Como sinal da covariância indica o tipo de relação entre as duas variáveis então o sinal positivo indica que elas se movem juntas e o sinal negativo indica que elas se movem em direções opostas.

Assim, mesmo que a covariância cresce à medida que o relacionamento aumenta ainda não se pode fazer uma avaliação sobre o poder do relacionamento entre as duas variáveis por ela não ser padronizada. Desse modo utiliza-se a correlação que é uma medida padronizada entre relações de variáveis. A correlação é calculada da covariância:

$$\begin{aligned} \text{Correlação} = \rho_{xy} &= \frac{\text{Cov}(X, Y)}{\text{Dp}(X) \cdot \text{Dp}(Y)} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sqrt{E[(X - \mu_x)^2]} \sqrt{E[(Y - \mu_y)^2]}} \\ &= \frac{\sum_{i=1}^n (X_i - \mu_x)(Y_i - \mu_y)}{\sqrt{\sum_{i=1}^n (X_i - \mu_x)^2} \sqrt{\sum_{i=1}^n (Y_i - \mu_y)^2}} \end{aligned}$$

A correlação são valores que varia de -1 a +1. Uma correlação próxima a zero indica que as duas variáveis não estão relacionadas. Uma correlação positiva indica que as duas variáveis movem juntas, e a relação é forte quanto mais a correlação se aproxima de um. Uma correlação negativa indica que as duas variáveis movem-se em direções opostas, e que a relação também fica mais forte quanto mais próxima de menos 1 a correlação ficar. A autocovariância é a covariância de dois grupos de dados de uma mesma sequência de dados. Supondo o 1º grupo $X(t)$ e o 2º grupo $X(t + \tau)$. Para ilustrar veja o exemplo abaixo:

Tabela 1.2:

	Dados	X_t	X_{t+1}
t_1	3	3	
t_2	2	2	2
t_3	4	4	4
t_4	5	5	5
t_5	6		6

Para calcular a $\gamma(1) = \text{cov}\{X(t), X(t + 1)\}$ da sequência da tabela 1.2 utiliza-se a 1º grupo que são dados de $X(t_1)$ a $X(t_4)$ e a 2º grupo que são dados de $X(t_2)$ a $X(t_5)$. Para facilitar a leitura utilizaremos a seguinte notação: $d_t = X(t) - \mu_t$ e $d_{t+1} = X(t + 1) - \mu(t + 1)$ onde μ é a média conforme citamos acima.

Tabela 1.3:

	X(t)	X(t+1)	d _t	d _{t+1}	d _t ·d _{t+1}
	3	2	-0,5	-2,25	1,125
	2	4	-1,5	-0,25	0,375
	4	5	0,5	0,75	0,375
	5	6	1,5	1,75	2,625
μ	3,5	4,25	∑ d _t d _{t+1} →		1,125
σ ²	1,25	2,188			
Dp	1,118	1,479			

Deste modo então a autocovariância da sequência é:

$$\text{Cov}\{X_t X_{t+1}\} = \frac{(-0,5) \cdot (-2,25) + (-1,5) \cdot (-0,25) + (0,5) \cdot (0,75) + (1,5) \cdot (1,75)}{4}$$

$$\text{Cov}\{X_t X_{t+1}\} = \frac{1,125 + 0,375 + 0,375 + 2,625}{4} = 1,125$$

e para calcular a autocorrelação desta sequência fazemos

$$\rho_1 = \frac{\text{Cov}(X_t, X_{t+1})}{\text{Dp}(X_t) \cdot \text{Dp}(X_{t+1})} = \frac{1,125}{(1,118)(1,479)} = \frac{1,125}{1,65359} = 0,680336$$

1.3 Séries temporais

Segundo (Morettin, Toloí-81) uma série temporal é qualquer conjunto de observações ordenadas no tempo. As séries temporais podem ser discretas citamos com exemplos os valores mensais de vendas televisores no Brasil, valores diários da temperatura da cidade do Rio de Janeiro e a quantidade diária de chuva na cidade de Bauru e, além disso, as séries temporais podem ser contínuas. Aqui utilizamos como exemplo os registros de marés no porto de Paranaguá e convém lembrar que neste caso podemos convertê-la em uma série temporal discreta coletando amostras em intervalo de tempos iguais. Apesar de haver duas maneiras de abordar a análise de séries temporais que conforme (Morettin, Toloí-81) nos dois casos o objetivo é construir modelos para as séries com determinados propósitos. Na primeira abordagem a análise é realizada em função do tempo e os modelos apresentados são os modelos com número finitos de parâmetros (paramétricos). Na segunda abordagem a análise ocorre em função de frequências e os modelos apresentados são os modelos não paramétricos. Dentre os modelos paramétricos existem os modelos de arima que é o nosso principal objeto de estudo. No que se relaciona a análise do domínio de frequências o estudo baseia-se na decomposição de componentes de frequência no qual o espectro é essencialmente importante, ocorrendo uma análise espectral. Esse tipo de análise não será estudado aqui.

1.3.1 Processos estocásticos

Processos Estacionários

Em séries temporais uma das suposições mais elementares é que o processo estocástico gerador da mesma seja estacionário, isto é, que o processo move-se em torno de uma média constante e também de variância constante. Os processos estocásticos estacionários podem ser estritamente (forte) ou fracamente estacionários. Mas antes, definimos formalmente processos estocásticos segundo (Morettin-2006).

Definição: Seja T um conjunto arbitrário. Um processo estocástico é uma família $X(t), t \in \mathcal{T}$, tal que, para cada; $X(t)$ é uma variável aleatória. Como para cada t pertencente a T existe uma variável aleatória $X(t)$ determinada sobre o espaço de probabilidade Ω , isto é, $X(t)$ é uma função de dois argumentos, $X(t, \omega)$, $t \in \mathcal{T}$ e $\omega \in \Omega$. Assim que para cada $t \in \mathcal{T}$, temos uma v.a $X(t, \omega)$, com uma função densidade de probabilidades $f_t(x)$ de outro modo, para cada $\omega \in \Delta$, fixado, adquirimos uma função de t .

Conforme (Morettin-2006) uma maneira de visualizar a distribuição de probabilidades de $X(t, \omega)$, para t fixo é considerar a proporção de trajetórias que passam por uma "janela" de amplitude Δ . Tal proporção será $f_t(x)\Delta$. Esta é a mesma ideia para construir um histograma para distribuição de valores de uma v.a.. Os valores de $X(t)$ são chamados de estados e o conjunto de valores $X(t)$ é chamado de espaço dos estados. Dizemos que se $\mathcal{T} = \mathbb{Z}$ o processo é com parâmetro discreto e se \mathcal{T} estiver num intervalo \mathbb{R} é com parâmetro contínuo.

1.3.2 Especificação de um Processo Estocástico

Sejam $F(x_1, \dots, x_n; t_1, \dots, t_n)$ elementos quaisquer de \mathcal{T} e consideremos

$$F(x_1, \dots, x_n) = P\{X(t_1) \leq x_1, \dots, X(t_n) \leq x_n\}$$

Então, o processo estocástico $X(t)$, $t \in \mathcal{T}$ estará especificado se conhecermos as distribuições finito-dimensionais (2.1), para todo $n \geq 1$. As funções de distribuição devem satisfazer às duas condições seguintes:

(i)(Condição de simetria): para qualquer permutação j_1, \dots, j_n , dos índices $1, 2, \dots, n$, temos

$$F(x_{j_1}, \dots, x_{j_n}; t_{j_1}, \dots, t_{j_n}) = F(x_1, \dots, x_n; t_1, \dots, t_n) \quad (1.1)$$

(ii)(Condição de compatibilidade): para $m < n$,

$$F(x_1, \dots, x_m, \infty, \dots, \infty; t_1, \dots, t_m, \dots, t_n) = F(x_1, \dots, x_m; t_1, \dots, t_m).$$

O lado esquerdo desse da equação acima deve ser entendido como

$$\lim_{x_{m+1}, \dots, x_n \rightarrow \infty} F(x_1, \dots, x_m, x_{m+1}, \dots, x_n; t_1, \dots, t_n).$$

Pode-se demonstrar que qualquer conjunto de funções distribuição finito-dimensionais satisfazendo as condições de simetria e compatibilidade define um processo estocástico $X(t)$ sobre \mathcal{T} . Este resultado é conhecido como o teorema de extensão de Kolmogorov. Na prática o conhecimento de todas as distribuições finito-dimensionais é muito difícil

às vezes até impossível então o que pode ser feito é trabalhar com certas características associadas a essas distribuições e que sejam simples de calcular e interpretar. Assim o que normalmente é feito e trabalhar com momentos de ordem bem pequena. Então nos concentraremos nos chamados processos estacionários de momentos de primeira e segunda ordem. A função média ou simplesmente média de $X(t)$ é dada por

$$\mu(1; t) = \mu(t) = E\{X(t)\} = \int_{-\infty}^{\infty} x dF(x; t),$$

enquanto a função auto-covariância de $X(t)$ é definida como

$$\begin{aligned} \mu(1, 1; t_1, t_2) - \mu(1; t_1)\mu(1; t_2) &= \gamma(t_1, t_2) \\ &= E\{X(t_1)X(t_2)\} - E\{X(t_1)\}E\{X(t_2)\}, \quad t_1, t_2 \in \mathcal{T} \end{aligned}$$

Observamos que $\mu(t)$ é uma função de $t \in \mathcal{T}$ e que $\gamma(t_1, t_2)$ depende de dois argumentos t_1 e t_2 . Note que se $t_1 = t_2 = t$, temos

$$\gamma(t, t) = \text{Var}\{X(t)\} = E\{X^2(t)\} - E^2\{X(t)\},$$

isto é, a função variância do processo $X(t)$ que indicaremos por $\sigma^2(t)$. A função de auto-covariância $\gamma(t_1, t_2)$ dá a covariância entre duas variáveis aleatórias $X(t_1)$ e $X(t_2)$, para quaisquer $t_1, t_2 \in \mathcal{T}$. A função $\mu(t)$ é obtida variando-se t em \mathcal{T} . Imagine um processo estocástico como uma família de variáveis aleatórias que para cada t , temos um conjunto de valores $Z^{(1)}(t), Z^{(2)}(t), \dots$ correspondentes às várias realizações. Calculando, para cada t , a média dos valores de $Z^{(n)}(t)$, média calculada em relação a j , obtemos a função $\mu(t)$. Percebe-se que nos processos os parâmetros mais importantes são a média e a função autocovariância. Utiliza-se como convenção no momento que se fala de um gráfico de $X(t)$ estará falando de um gráfico com todas as trajetórias de $X(t)$, então se utiliza uma trajetória do processo.

Definição: Um processo estocástico $\{X(t), t \in \mathcal{T}\}$ diz-se estritamente estacionário se todas as distribuições finito dimensionais permanecem as mesmas sob translações do tempo, ou seja,

$$F(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau) = F(x_1, \dots, x_n; t_1, \dots, t_n),$$

para quaisquer t_1, \dots, t_n, τ de \mathcal{T} .

Em outras palavras as distribuições uni-dimensionais são invariantes sob translações do tempo, isto é a média $\mu(t)$ e a variância σ^2 são constantes, desse modo

$$E\{X(t)\} = \mu(t) = \mu, \quad \text{para todo } t \in \mathcal{T}$$

O mesmo ocorre com todas as distribuições bi-dimensionais que dependem de diferenças de tempos. Para $t_1, t_2 \in \mathcal{T}$, $\gamma(t_1, t_2) = \gamma(t_1 + t, t_2 + t)$. Substituindo t por $-t_2$, teremos que

$$\gamma(t_1, t_2) = \gamma(t_1 - t_2, 0) = \text{Cov}\{X(t_1 - t_2), X(0)\}.$$

Na prática a covariância é uma função de $|t_1 - t_2|$. Então a função auto-covariância de um processo estacionário forte ou estrito é da seguinte forma

$$\gamma(\tau) = \text{Cov}\{X(t), X(t \pm \tau)\} = \text{Cov}\{X(0), X(\tau)\}$$

1 FUNDAMENTAÇÃO TEÓRICA

Definição: Um processo estocástico $\{X(t), t \in \tau\}$ diz-se fracamente estacionário (ou estacionário de segunda ordem) se e somente se

- (i) $E\{X(t)\} = \mu(t) = \mu$, constante, para todo $t \in \tau$;
- (ii) $E\{X^2(t)\} < \infty$, para todo $t \in \tau$;
- (iii) $\gamma(\tau) = \text{Cov}\{X(t_1), X(t_2)\}$ é uma função apenas de $|t_1 - t_2|$.

Convém salientar que para $X(t)$ ser estritamente estacionário não é necessário que seja fracamente estacionário pois pode não satisfazer a condição (ii) da definição.

Definição: Um processo estocástico real $\{X(t), t \in \tau\}$ diz-se Gaussiano se, para qualquer conjunto t_1, \dots, t_n de τ , v.a. $X(t_1), X(t_2), \dots, X(t_n)$ têm uma distribuição normal n -variada. Veja que um processo Gaussiano com variância finita deve ser calculado pelas médias e covariância, então se ele for estacionário de segunda ordem, será também estritamente estacionário.

1.3.3 Propriedades da Função de Auto-covariância

Seja $\{X(t), t \in Z\}$ um processo estacionário real discreto, de média zero e f.a.c.v. $\gamma = E\{X_t X_{t+\tau}\}$.

Proposição: A f.a.c.v. γ satisfaz as seguintes propriedades:

- (i) $\gamma_0 > 0$,
- (ii) $\gamma_{-\tau} = \gamma_\tau$,
- (iii) $|\gamma_\tau| \leq \gamma_0$,
- (iv) $\gamma_{-\tau}$ é não negativa definida, no sentido que

$$\sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma_{\tau_j - \tau_k} \geq 0,$$

para quaisquer números reais a_1, \dots, a_n , e τ_1, \dots, τ_n de Z . **Prova.** As propriedades (i) e (ii) decorrem imediatamente da definição de γ_τ . A propriedade (iii) segue do fato que

$$E\{X_{t+\tau} \pm X_t\}^2 = E\{X_{t+\tau}^2 \pm 2X_{t+\tau}X_t + X_t^2\} \geq 0$$

Mas o segundo membro é igual a

$$\sigma^2 \pm 2\gamma_\tau \pm \sigma^2 \geq 0$$

ou seja ,

$$2\gamma_0 \pm 2\gamma \geq 0$$

e (iii) fica demonstrada. Quanto a (iv) temos que

$$\sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma_{\tau_j - \tau_k} = \sum_{j=1}^n \sum_{k=1}^n a_j a_k E\{X_{\tau_j} X_{\tau_k}\}$$

$$E \left\{ \sum_{j=1}^n a_j X_{\tau_j} \right\}^2 \geq 0.$$

Definição: A função de auto-correlação (f.a.c.) do processo é definida por

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}, \quad \tau \in Z,$$

e tem as propriedades de γ_τ , exceto que agora $\rho_0 = 1$.

1.3.4 Processos Lineares Estacionários

Apresentaremos aqui alguns exemplos de processos estocásticos estacionários principalmente os processos auto-regressivos (AR), de médias Móveis (MA) e combinação dos dois os processos ARMA que são muito utilizados.

Exemplo 1 Sequência Aleatória

Consideremos $\{X_n, n = 1, 2, \dots\}$ uma sequência de v.a. definidas no mesmo espaço amostral ω . Supomos um processo com parâmetro discreto, ou uma sequência aleatória. Para todo $n \geq 1$, podemos escrever

$$\begin{aligned} P\{X_1 = a_1, \dots, X_n = a_n\} &= P\{X_1 = a_1\} \times P\{X_2 = a_2 | X_1 = a_1\} \\ &\times \dots \times P\{X_n = a_n | X_1 = a_1, \dots, X_{n-1} = a_{n-1}\}. \end{aligned}$$

Se tiver um caso com uma sequência $\{X_n, n \geq 1\}$ de v.a. mutuamente independente então escrevemos

$$P\{X_1 = a_1, \dots, X_n = a_n\} = P\{X_1 = a_1\} \dots P\{X_n = a_n\}.$$

E ainda o caso em que as v.a. todas tem a mesma distribuição configura-se uma sequência de v.a. independentes e identicamente distribuídas (i.i.d) e para o processo X_n ser estacionário será necessário $E\{X_n\} = \mu, \text{Var}\{X_n\} = \sigma^2$, para todo $n \geq 1$, desse modo

$$\gamma_\tau = \text{Cov}\{X_n, X_{n+\tau}\} = \begin{cases} \sigma^2, & \text{se } \tau = 0 \\ 0, & \text{se } \tau \neq 0. \end{cases}$$

Segue-se que $\sigma_\tau = 1$, para $\tau = 0$ e $\sigma_\tau = 0$, caso contrário.

Definição: Dizemos que $\{\varepsilon_t, t \in Z\}$ é ruído branco discreto se as v.a. ε_t são não correlacionadas, isto é, $\text{Cov}\{\varepsilon_t, \varepsilon_s\} = 0, t \neq s$. Uma sequência de v.a. i.i.d., como definida acima, é chamada um processo puramente aleatório e escreveremos

$$\varepsilon_t \sim \text{i.i.d.}(0, \sigma^2).$$

Para representar um ruído branco discreto escreveremos

$$\varepsilon_t \sim \text{RB}(0, \sigma^2).$$

Exemplo 2 Passeio Aleatória

1 FUNDAMENTAÇÃO TEÓRICA

Suponha uma sequência aleatória $\{\varepsilon_t, t \geq 1\}$, de v.a. i.d.d. $\{\mu_\varepsilon, \sigma_\varepsilon^2\}$. Defina a sequência

$$X_t = \varepsilon_1 + \dots + \varepsilon_t.$$

Então $E(X_t) = t\mu_\varepsilon$ e $\text{Var}(X_t) = t\sigma_\varepsilon^2$. Podemos verificar que a função autocovariância é

$$\text{cov}(X_t, X_{t-k}) = \text{Cov}(\varepsilon_1 + \dots + \varepsilon_t - k + \dots + \varepsilon_t, \varepsilon_1 + \dots + \varepsilon_t - k) = (t-k)\sigma_\varepsilon^2$$

e segue função de autocorrelação

$$\rho_t(k) = \frac{t-k}{t}$$

onde fica claro que tanto a média, a variância e as autocovariâncias dependem de t , portanto não é estacionário. Processos como estes são chamados de passeio aleatório ou casual.

Processos Auto-regressivos

Segundo Morettin-2006, dizemos que $\{X_t, t \in Z\}$ é um processo auto-regressivo de ordem p , e escrevemos $X_t \sim \text{AR}(p)$, se satisfizer à equação de diferenças

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \varepsilon_t,$$

onde $\mu, \phi_1, \dots, \phi_p$ são parâmetros reais e $\varepsilon_t \sim \text{RB}(0, \sigma^2)$. Segue-se que $E(X_t) = \mu$ e se escrevermos o processo na forma

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

então

$$\mu = E(X_t) = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p}$$

Definamos o operador retroativo B através de $B^s X_t = X_{t-s}$, $s \geq 1$. Então a equação de um processo AR pode ser escrita

$$\phi(B)\tilde{X}_t = \varepsilon_t$$

, onde $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é o operador auto-regressivo de ordem p e $\tilde{X}_t = X_t - \mu$. Suponha $\mu = 0$ no que segue. Um caso particular importante é o processo $\text{AR}(1)$,

$$X_t = \phi_1 X_{t-1} + \varepsilon_t$$

.

Aqui, $\phi = 1 - \phi B$. Através de substituições sucessivas obtemos

$$X_t = \sum_{j=0}^{r-1} \phi^j \varepsilon_{t-j} + \phi^r X_{t-r}.$$

Se X_t for estacionário, com variância finita ρ_x^2 , então

$$E[X_t - \sum_{j=0}^r \phi^j \varepsilon_{t-j}]^2 = \phi^{2r+2} E[X_{t-r-1}^2] = \phi^{2r+2} \sigma_x^2.$$

Se $|\phi| < 1$, $\phi^{2(r+1)} \rightarrow 0$, quando $r \rightarrow \infty$, portanto sob suposição, podemos escrever

$$X_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j},$$

Onde a convergência é em média quadrática. Logo, a condição $|\phi| < 1$ é suficiente para X_t ser estacionário. Multiplicando ambos os membros do processo AR(1) acima por X_{t-r} tomando a esperança, obtemos

$$\gamma_\tau = \gamma_{\tau-1} = \dots = \phi^\tau \gamma_0$$

lembrando que $\gamma_0 = \sigma_x^2$ e que $X_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$, e substituindo x por X_t então

$$\sigma^2 \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma^2}{1-\phi^2} \Rightarrow \gamma_\tau = \frac{\sigma^2}{1-\phi^2} \phi^\tau, \tau \geq 0$$

Como γ_τ é simétrica, considerando $\tau \in Z$, a função autocovariância (f.a.c.v.) e a função autocorrelação (f.a.c.) de um processo AR(1) é

$$\gamma_\tau = \frac{\sigma^2}{1-\phi^2} \phi^{|\tau|} \Rightarrow \rho_\tau = \frac{\gamma_\tau}{\gamma_0} = \phi^{|\tau|}$$

Podemos ter uma solução de outra forma,

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

Como vimos ,utilizando o operador retroativo podemos definir processo (AR) $\phi(B)\tilde{X}_t = \varepsilon_t$ e desse modo

$$X_t = \phi(B)^{-1} \varepsilon_t = \psi(B) \varepsilon_t,$$

Considerando $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$. Assim como em AR(1), teremos $\sum_j \psi_j^2 < \infty$ para que a solução acima seja estacionária.Podemos obter os coeficientes ψ_j através da relação $\phi(B)\psi(B) = 1$.

Supondo que todas as raízes de $\phi(B) = 0$ estejam fora do círculo unitário para que X_t seja estacionário, assim para $p=1$, $\phi(B) = 1 - \phi B = 0$.Segue que $B = \phi^{-1}$ o que implica $|\phi| < 0$. Supondo o processo estacionário, multiplicando os dois membros da equação

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

por $X_{t-\tau}$ e tomando os valores esperados , para $\tau \geq 0$,

$$\sigma_x^2 = \frac{\sigma^2}{1 - \phi_1 \rho_1 + \dots + \phi_p \rho_p}, \quad \text{para } \tau = 0,$$

$$\gamma_\tau = \phi_1 \gamma_{\tau-1} + \phi_2 \gamma_{\tau-2} + \dots + \phi_p \gamma_{\tau-p}, \quad \text{para } \tau > 0.$$

1 FUNDAMENTAÇÃO TEÓRICA

Note que essas equações pode ser utilizada para obter ρ_τ , bastando dividir esses resultados acima por γ_0 . A solução desta equação é dada por (Miller,1969)

$$\gamma_\tau = A_1 G_1^\tau + A_2 G_2^\tau + \dots + A_p G_p^\tau$$

Onde os G_i 's satisfazem

$$\phi(B) = \prod_{i=1}^p (1 - G_i B).$$

Devemos ter $|G_i| < 1$ para todo $i = 1, \dots, p$ considerando que as raízes de $\phi(B) = 0$ estão fora do círculo unitário. Para obter estimadores dos parâmetros $\phi_j \tilde{Z}_s$, substituindo as f.a.c.v.'s por suas estimativas. Assim faremos $\tau = 1, 2, \dots, p$ na equação $\gamma_\tau = \phi_1 \gamma_{\tau-1} + \phi_2 \gamma_{\tau-2} + \dots + \phi_p \gamma_{\tau-p}$ obtendo assim

$$\Gamma_p \phi_p = \gamma_p,$$

onde $\Gamma_p = [\gamma_{ij}]$, com $\gamma_{ij} = \gamma_{|i-j|}$, $i, j = 1, \dots, p$, $\phi_p = (\phi_1, \dots, \phi_p)'$ e $\gamma_p = (\gamma_1, \dots, \gamma_p)'$

Assim a representação dessa matriz

$$\Gamma_p = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \vdots & \ddots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{bmatrix} \quad \phi_p = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad \gamma_p = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{bmatrix}$$

Esses estimadores são chamados de estimadores de Yule-Walker.

Processos de Médias Móveis

Dizemos que $\{X_t, t \in Z\}$ é um processo de médias móveis de ordem q , se satisfazer à equação de diferenças

$$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q},$$

onde $\mu, \theta_1, \dots, \theta_q$ são constantes reais e $\varepsilon_t \sim \text{RB}(0, \sigma^2)$. Denotaremos por MA(q).

Obtemos a variância do processo, considerando X_t estacionário de média μ e ε_t são não correlacionados,

$$\sigma_x^2 = \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2).$$

Quanto à f.a.c.v., temos

$$\begin{aligned} \gamma_\tau &= E\{X_t X_{t-\tau}\} = \gamma_\varepsilon(\tau) - \sum_{k=1}^q \theta_k \gamma_\varepsilon(k - \tau) \\ &\quad - \sum_{l=1}^q \theta_l \gamma_\varepsilon(\tau + l) + \sum_{k=1}^q \sum_{l=1}^q \theta_k \theta_l \gamma_\varepsilon(\tau + l - k), \end{aligned}$$

Denota-se $\gamma_\varepsilon(\tau)$ como a f.a.c.v de ε_t . Desse modo temos

$$\gamma_\tau = \begin{cases} \sigma^2(-\theta_\tau + \theta_1\theta_{\tau+1} + \dots + \theta_q\theta_{q-\tau}), & \text{se } \tau = 1, \dots, q \\ 0, & \text{se } \tau > q \\ \gamma_{-\tau}, & \text{se } \tau < 0. \end{cases}$$

Da variância e do resultado acima temos que a f.a.c. do processo MA(q):

$$\rho_\tau = \begin{cases} \frac{-\theta_\tau + \theta_1\theta_{\tau+1} + \dots + \theta_q\theta_{q-\tau}}{1 + \theta_1^2 + \dots + \theta_q^2}, & \text{se } \tau = 1, \dots, q \\ 0, & \text{se } \tau > q \\ \rho_{-\tau}, & \text{se } \tau < 0. \end{cases}$$

Para um processo MA(1),

$$X_t = \varepsilon_t - \theta\varepsilon_{t-1},$$

segue que

$$\text{Var}(x_t) = \sigma_x^2 = \sigma^2(1 + \theta^2),$$

$$\rho_\tau = \begin{cases} \frac{-\theta}{1 + \theta^2}, & \text{se } \tau = \pm 1 \\ 0, & \text{se } |\tau| > 1. \end{cases}$$

O operador de médias móveis de ordem q definiremos por

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

Então a equação de diferenças do processo MA(q) podemos escrever assim

$$X_t = \theta(B)\varepsilon_t.$$

Deste modo para um processo MA(1) devemos ter $\theta(B) = 1 - \theta B$, logo

$$X_t = (1 - \theta B)\varepsilon_t$$

assim escrevemos formalmente,

$$\varepsilon_t = (1 - \theta B)^{-1}X_t = (1 + \theta B + \theta^2 B^2 + \dots)X_t,$$

o que segue

$$X_t = -\theta X_{t-1} - \theta^2 X_{t-2} - \dots + \varepsilon_t,$$

Observe que é necessário que $|\theta| < 1$ para que a o lado direito da igualdade convirja e esta equação é escrita como processo auto-regressivo de ordem infinita, isto é, $|\theta| < 1$ é condição de invertibilidade para o processo de M.A.(1). De um modo geral, o processo de M.A. de (q) poderá ser escrito

$$X_t = \sum_{j=1}^{\infty} \pi_j X_{t-j} + \varepsilon_t,$$

Satisfeita a condição de invertibilidade então todas as raízes de $\theta(B) = 0$ devem estar fora do círculo unitário, desse modo a equação acima pode ser escrita

1 FUNDAMENTAÇÃO TEÓRICA

$$\pi(B)X_t = \varepsilon_t,$$

onde $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$, e $\pi(B) = \theta(B)^{-1}$. Assim, podemos obter os coeficientes π_j através da identidade $\theta(B)\pi(B) = 1$.

Processos Auto-regressivos e de Médias Móveis

Um processo auto-regressivo e de médias móveis, de ordem (p,q) , denotado por ARMA (p,q) , é definido por

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q}$$

onde $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$. Considerando que a média do processo é μ e usando os operadores auto-regressivo e de médias móveis, então

$$\phi(B)\tilde{X}_t = \theta(B)\varepsilon_t,$$

onde $\tilde{X}_t = X_t - \mu$. Considerando $\mu = 0$, usaremos o ARMA $(1,1)$, isto é,

$$X_t = \phi X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}.$$

Fazendo substituições sucessivas

$$X_t = \psi(B)\varepsilon_t,$$

onde $\psi_j = \phi^{j-1}(\phi - \theta)$, $j \geq 1$.

As condições de estacionariedade e invertibilidade são as mesmas que para os processos AR (1) e MA (1) , isto é, $|\phi| < 1$ e $|\theta| < 1$ o que nos leva a escrever o processo, $\pi_j = \theta^{j-1}(\phi - \theta)$, $j \geq 1$.

Para a condição de estacionariedade de um processo ARMA (p,q) genérico devemos ter as raízes de $\phi(B) = 0$ fora do círculo unitário, de modo semelhante para condição de invertibilidade devemos ter as raízes de $\theta(B) = 0$ também fora do círculo unitário. Para obtermos as auto-covariância, utilizamos a equação de diferenças definida para o processo ARMA (p,q) , assim consideramos $\mu = 0$ depois multiplicamos por $X_{t-\tau}$ e tomamos a esperança

$$\gamma_\tau = \phi_1\gamma_{\tau-1} + \dots + \phi_q\gamma_{\tau-p} + \gamma X_\varepsilon(\tau)$$

$$- \theta_1\gamma X_\varepsilon(\tau-1) - \dots - \theta_q\gamma X_\varepsilon(\tau-q),$$

onde $\gamma X_\varepsilon(\tau)$ é a covariância cruzada entre X_t e ε_t , definida por

$$\gamma X_\varepsilon(\tau) = E(\varepsilon X_{t-\tau}).$$

onde esta covariância só é diferente de zero para $\tau \leq 0$ logo

$$\gamma_\tau = \phi_1\gamma_{\tau-1} + \dots + \phi_q\gamma_{\tau-p}, \quad \tau > q.$$

Note que as auto-covariâncias de "lags" de 1 a q dependem dos parâmetros de médias móveis, e para "lags" superiores a q ($\tau > q$) as mesmas são influenciadas pelos parâmetros dos modelos AR.

Vejo como exemplo o caso ARMA(1,1) onde temos

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1 - \phi\theta)(\phi - \theta)}{1 + \theta^2 - 2\phi\theta}$$

e no caso de $\tau > 1$,

$$\rho_\tau = \phi\rho_{\tau-1}.$$

Processo Linear Geral

A expressão

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

representa o chamado processo linear geral (PLG), onde $\varepsilon_t \sim RB(0, \sigma^2)$ e ψ_j são constantes. Note que os processos AR, MA e ARMA são casos particulares do (PLG). A condição em que $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ é necessária para que a variância seja finita, então

$$\sigma_x^2 = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2.$$

além disso como $E\{X_t\} = 0$ temos

$$\sigma_\tau = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+\tau},$$

considerando que a série do lado direito da igualdade convirja para um valor finito. Do fato que

$$|E\{X_t X_{t-\tau}\}| \leq [E\{X_t^2\} E\{X_{t-\tau}^2\}]^{1/2} < \infty,$$

e que $\sigma^2 < \infty$, segue que $\gamma_\tau < \infty$ se $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ onde estabelecemos a condição de estacionariedade para o PLG. Sendo assim a f.a.c. de um PLG é

$$\rho = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+\tau}}{\sum_{j=0}^{\infty} \psi_j^2}.$$

Teorema de Wold

Conforme Morettin-2006) segue o teorema abaixo:

Teorema:(Wold) Todo processo estacionário de segunda ordem, puramente não-determinístico, pode ser escrito como

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1,$$

1 FUNDAMENTAÇÃO TEÓRICA

com $\varepsilon_t \sim \text{RB}(0, \sigma^2)$. Um processo diz-se puramente não-determinístico se ele não puder ser previsto exatamente a partir de seu passado.

1.3.5 Processos Não Estacionários

Há duas maneiras de gerar processos não estacionários e que sejam não explosivos

- (a) Acrescentar na fórmula de Wold uma tendência determinística

$$X_t = \beta_0 + \beta_1 t + \psi(B)\varepsilon_t$$

neste caso é um processo "Trend-stationary".

- (b) Supor um PLG com raiz unitária

$$(1 - B)X_t = \delta + \psi(B)\varepsilon_t,$$

Este modelo, com $\psi(1) \neq 0$ descreve variações de X_t e como $\psi(1) = \sum_{j=0}^{\infty} \psi_j \neq 0$, o processo é não-estacionário.

Assim se tivermos um processo como o do item (a) com erro ARMA(p,q) e tendência μ_t um polinômio determinístico de grau d, podemos escrever assim

$$X_t = \sum_{j=0}^d \beta_j t^j + [\phi(B)]^{-1} \theta(B) a_t,$$

onde a_t é um ruído branco. Assim segue

$$\Delta^d X_t = (1 - B)^d X_t = \theta_0 + (1 - B)^d \psi(B) a_t,$$

na qual $\psi(B) = [\phi(B)]^{-1} \theta(B)$ e $\theta_0 = d! \beta_d$, um processo estacionário, mas não invertível.

Definição:

Se $\Delta^d X_t$ for estacionário, dizemos que X_t é integrado de ordem d e escrevemos $X_t \sim I(d)$ E assim

Definição:

Se $\Delta^d X_t \sim \text{ARMA}(p,q)$ dizemos que X_t segue um modelo ARIMA(p,d,q) isto é auto-regressivo integrado e de médias móveis de ordem (p,d,q), ou seja,

$$\phi(B)\Delta^d X_t = \theta_0 + \theta(B)a_t$$

ou, de modo equivalente,

$$\phi(B)W(t) = \theta_0 + \theta(B)a_t, \quad \text{com } W(t) = \Delta^d X_t.$$

Observemos que

$$W(t) = \Delta^d X_t \leftrightarrow X_t = S^d W(T),$$

onde S é o operador soma ou integral

$$S = (1 - B)^{-1} = \Delta^{-1}.$$

ou seja, X_t pode ser obtido somando-se ou integrando-se o processo estacionário $W(t)$ d vezes, donde o nome processo integrado para X_t .

1.3.6 Modelagem ARIMA

Como vimos até agora um modelo ARMA (p,q) é dado pela equação de diferenças

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \theta_t - \sigma_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

onde $\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$. Com os operadores auto-regressivos e de média móvel definidos anteriormente, segue

$$\phi(B)\tilde{X}_t = \theta(B)\varepsilon_t$$

com $\tilde{X}_t = X_t - \mu$. Supondo $\mu \neq 0$, o modelo pode ser escrito

$$X_t = \theta_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q},$$

onde $\theta_0 = \mu(1 - \phi_1 - \dots - \phi_p)$. Assim dizemos que um processo $\{X_t, t \in Z\}$ é um modelo ARIMA (p,d,q) se $\Delta^d X_t$ seguir um modelo ARMA (p,q) , isto é

$$\phi(B)\Delta^d X_t = \theta(B)\varepsilon_t$$

Na determinação de um modelo ARIMA para uma série temporal devemos examinar três estágios:

- (i) Identificação
- (ii) Estimação
- (iii) Diagnóstico

Vale lembrar que os estágio (i)-(ii) devem ser repetidos se o (iii) não atender o que se deseja. Haverá momentos da necessidade de vários modelos para que um determinado critério decida o modelo mais apropriado.

Identificação

Considerada a fase mais crítica do processo iterativo que descrevemos acima, esta escolha deve ser feita utilizando principalmente as auto-correlações e auto-correlações parciais estimadas. Como já foi citado a f.a.c ρ_j é estimada por

$$r_j = \frac{c_j}{c_0}, \quad j = 0, 1, \dots, T-1,$$

onde c_j é a estimativa da f.a.c.v γ_j ,

$$c_j = \frac{1}{T} \sum_{t=1}^{T-j} [(X_t - \bar{X})(X_{t+j} - \bar{X})], \quad j = 0, 1, \dots, T-1,$$

1 FUNDAMENTAÇÃO TEÓRICA

sendo $\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t$ a média amostral. Como as auto-correlações verdadeiras são funções pares, colocamos $c_{-j} = c_j$ e $r_{-j} = r_j$.

Uma expressão aproximada para variância de r_j , para um processo estacionário normal, é dada por

$$\text{Var}(r_j) \simeq \frac{1}{T} \sum_{v=-\infty}^{\infty} [\rho_v^2 + \rho_v + j\rho_{v-j} - 4\rho_j\rho_{v-j} + 2\rho_v^2\rho_j^2]$$

Em um processo em que as auto-correlações são nulas para $v > q$, todos os termos do lado direito da expressão acima se anulam para $j > q$, exceto o primeiro,

$$\text{Var}(r_j) \simeq \frac{1}{T} \left[1 + 2 \sum_{v=1}^q \rho_v^2 \right], \quad j > q.$$

Por não conhecermos as auto-correlações ρ_v , substituímos as mesmas por r_v , chegando a uma estimativa da expressão acima,

$$\widehat{\sigma}^2(r_j) \simeq \frac{1}{T} \left[1 + 2 \sum_{v=1}^q \rho_v^2 \right], \quad j > q.$$

Para T suficientemente grande e considerando que $\rho_j = 0$, para $j > q$, a distribuição de r_j é aproximadamente normal, com média igual a zero e variância dada pela expressão acima. Assim temos um intervalo de confiança aproximado para auto-correlações,

$$r_j \pm t_\gamma \cdot \widehat{\sigma}(r_j),$$

onde t_γ é o valor da estatística t de Student com $T-1$ graus de liberdade, tal que $P(-t_\gamma < t < t_\gamma) = \gamma$. Na prática usa-se $t_\gamma = 2$, correspondendo a $\gamma = 0,95$, aproximadamente, de modo que podemos considerar σ_j como sendo significativamente diferente de zero se

$$|r_j| > 2\widehat{\sigma}(r_j), \quad j > q.$$

Como foi apresentado até aqui, os processos lineares estacionários apresentam f.a.c com características especiais. Assim:

- (i) um processo AR(p) tem f.a.c que decai de acordo com exponenciais e/ou senóides amortecidas, infinita em extensão;
- (ii) um processo MA(q) tem f.a.c finita, no sentido que ela apresenta um corte após o "lag" q;
- (iii) um processo ARMA(p,q) tem f.a.c infinita em extensão, que decai de acordo com exponenciais e/ou senóides amortecidas após o "lag" q-p.

Estas considerações serão importantes no mecanismo de identificação do modelo a ajustar aos dados observados; calculando-se as estimativas das f.a.c., comparando seu comportamento com o descrito acima, para cada modelo, procuramos estabelecer um modelo (ou modelos) que descreva a série analisada. Apesar da f.a.c. ser útil para identificar modelos MA, conforme característica (ii), não é útil para identificar modelos ARMA, visto que f.a.c complexa. Uma outra ferramenta para simplificar esse

mecanismos de identificação Box, Jenkins e Reinsel (1994) recomenda: a função de auto-correlação parcial (f.a.c.p.). Iremos representar por ϕ_{kj} o j -ésimo coeficiente de um modelo AR(k), de tal modo que ϕ_{kk} seja o último coeficiente. Sabemos que

$$\rho_j = \phi_{k1}\rho_j - 1 + \phi_{k2}\rho_j - 2 + \dots + \phi_{kk}\rho_j - k, \quad j = 1, \dots, k,$$

a partir das quais obtemos as equações de Yule-walker

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_2 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}$$

Resolvendo estas equações sucessivamente para $K = 1, 2, 3, \dots$ obtemos

$$\begin{aligned} \phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \end{aligned}$$

e, em geral,

$$\phi_{kk} = \frac{|P_k^*|}{|P_k|},$$

onde P_k é a matriz de auto-correlações e P_k^* é matriz P_k com a última coluna substituída pelo vetor de auto-correlações. Assim a quantidade ϕ_{kk} , encarada como função de K , é chamada função de auto-correlação parcial. Veja que ϕ_{kk} mede a correlação que sobrou entre X_t e X_{t-k} depois de eliminada a influência de $X_{t-1}, \dots, X_{t-k+1}$ pois é possível provar que ϕ_{kk} é igual à correlação parcial entre as variáveis X_t e X_{t-k} ajustadas às variáveis intermediárias $X_{t-1}, \dots, X_{t-k+1}$. Decorrerá indispensável estimar a f.a.c.p. de um processo AR, MA ou ARMA. Um modo baseia-se em estimar, sucessivamente, modelos auto-regressivos de ordens $p = 1, 2, 3, \dots$ por mínimos quadrados e tomar as estimativas do último coeficiente de cada ordem. Um outro caminho fundamenta-se em substituir nas equações de Yule-Walker as f.a.c. ρ_j por estimativas, ou seja, supor

$$r_j = \hat{\phi}_{k1}r_{j-1} + \dots + \hat{\phi}_{kk}r_{j-k}, \quad j = 1, \dots, k,$$

e resolver as equações de Yule-Walker com ρ_j substituída por r_j e ϕ_{kj} substituída por $\hat{\phi}_{kj}$. Quenouille (1949) mostra que, sob a suposição que o processo se AR(p), as

1 FUNDAMENTAÇÃO TEÓRICA

f.a.c.p. estimadas de ordem $p + 1, p + 2, \dots$ são, aproximadamente, independentemente distribuídos, com

$$\text{Var}(\hat{\phi}_{kk}) \approx \frac{1}{T}, \quad K \geq p + 1.$$

Se o número de observações, T , for suficientemente grande, $\hat{\phi}_{kk}$ tem distribuição aproximada normal, o que permite a construção de intervalos de confiança para $\hat{\phi}_{kk}$.

Podemos resolver as equações de Yule-Walker pelo algoritmo de Durbin-Levinson. Mais detalhes em Morettin (1984).

É possível provar (Box, Jenkins e Reinsel, 1994) que, para os processos já explorados, temos:

- (i) um processo AR(p) tem f.a.c.p. $\phi_{kk} \neq 0$, para $K \leq p$ e $\phi_{kk} = 0$, para $K > p$;
- (ii) um processo MA(q) tem f.a.c.p. que se comporta de maneira similar à f.a.c. de um processo AR(p): é dominada por exponenciais e/ou senóides amortecidas;
- (iii) um processo ARMA(p, q) tem f.a.c.p. que se comporta como a f.a.c.p. de um processo MA puro.

Observe que a f.a.c.p. é útil para identificar modelos AR puros, o que não ocorre para identificar modelos MA e ARMA.

Vimos acima que sob a hipótese que o processo seja AR(p),

$$\text{Var}(\hat{\phi}_{jj}) \simeq \frac{1}{T}, \quad j > p,$$

de modo que

$$\hat{\sigma}(\hat{\phi}_{jj}) \simeq \frac{1}{T}, \quad j > p.$$

Além disso, para T grande e sob a hipótese que o processo seja AR(p), $\hat{\phi}_{jj}$ terá distribuição aproximadamente normal, com média zero e variância conforme acima, de modo que consideraremos ϕ_{jj} significativamente diferente de zero se

$$|\hat{\phi}_{jj}| > \frac{2}{\sqrt{T}}, \quad j > P.$$

Procedimento de Identificação A finalidade da identificação é definir os valores de p, d e q do modelo ARIMA(p, d, q). Esse procedimento de identificação possui três fases:

- (a) averiguar se há necessidade de uma transformação na série original, com objetivo de estabilizar sua variância;
- (b) tomar diferenças da série, obtida no item (a), tantas vezes quantas necessárias para se obter uma série estacionária, de modo que o processo $\Delta^d X_t$ seja reduzido a um ARMA(p, q). O número de diferenças, d , necessárias para que o processo se torne estacionário, é alcançado quando a f.a.c. amostral de $W_t = \Delta^d X_t$ decresce rapidamente para zero. Neste estágio, a utilização de um teste para verificar a existência de raízes unitárias no polinômio auto-regressivo, pode ser de grande valia;

- (c) determinar o processo ARMA(p,q) resultante por meio da análise das auto-correlações e auto-correlações parciais estimadas, cujos comportamentos podem reproduzir a ação das respectivas quantidades teóricas, isto é, modelos AR, MA e ARMA.

Para explicar o item(b), do procedimento de identificação seguimos o seguinte raciocínio.

Conforme apresentamos para um modelo ARMA estacionário, as f.a.c. são dadas por

$$\rho_j = A_1 G_1^j + A_2 G_2^j + \dots + A_p G_p^j, \quad j > q - p,$$

considerando raízes distintas. Como $\phi(B) = \prod_{i=1}^p (1 - G_i B)$ e as raízes de $\phi(B) = 0$ devem estar fora do círculo unitário, devemos ter $|G_i| < 1$. Segue-se da equação acima que, se nenhuma raiz estiver muito próxima do círculo unitário, as auto-correlações ρ_j declinarão para zero, para valores moderados de j . Por isso, considere que uma raiz real, G_1 , esteja próxima de um, ou seja, $G_1 = 1 - \varepsilon$, $\varepsilon > 0$ pequeno. Como $G_1^j = (1 - \varepsilon)^j \simeq 1 - j\varepsilon$, vem que $\rho_j \simeq A_1(1 - j\varepsilon)$, o que mostra que a f.a.c. baixará lentamente para zero e de forma aproximadamente linear. A maior dificuldade, neste estágio desse mecanismo é fugir de um exagero de diferenças, por isso usa-se $d=0,1$ ou 2 e é suficiente inspecionar as primeiras 15 ou 20 auto-correlações da série e de suas diferenças. É importante testar se $E(W_t) = \mu_w$ é zero, comparando \overline{W} com seu desvio padrão estimado. A tabela 2 apresenta as variâncias de \overline{W} para alguns modelos usuais. Conforme já foi citado, a f.a.c. e f.a.c.p. são úteis para determinar modelos MA e AR puros. Para identificar modelos ARMA devemos supor modelos simples, como ARMA(1,1), ARMA(1,2), ARMA(2,1) ou ARMA(2,2) e depois da estimação, escolher aquele que melhor atende a alguns critérios.

Formas alternativas de identificação A partir de 1970 foram propostos várias maneiras para identificação de modelos ARMA. O objetivo é escolher as ordens K e l que diminuam a quantidade

$$P(k, l) = \ln \hat{\sigma}_{k,l}^2 + (k + l) \frac{C(T)}{T},$$

em que $\hat{\sigma}_{k,l}^2$ é uma estimativa da variância residual obtida ajustando um modelo ARMA(k,l) às T observações da série e $C(T)$ é uma função do tamanho da série. A quantidade $(k+l)C(T)/T$, denominada termo penalizador, aumenta quando o número de parâmetros aumenta, enquanto que a variância residual $\hat{\sigma}_{k,l}^2$ diminui. Deste modo, ao diminuir a quantidade da expressão acima esta relacionado a identificar as ordens k e l que equilibrem esse comportamento. É possível notar que as ordens escolhidas crescem quando T cresce também. Por isso, Hannan(1982) sugere limites superiores dados por $(\ln T)^\alpha$, $0 < \alpha < \infty$. Veja agora alguns procedimentos de identificação que diminuam as funções penalizadoras particulares.

A-Critério de informação AIC Akaike (1973,1974) recomenda escolher o modelo cujas ordens k e l diminuam o critério

$$AIC(k, l) = \ln \hat{\sigma}_{k,l}^2 + \frac{2(k + l)}{T},$$

1 FUNDAMENTAÇÃO TEÓRICA

Tabela 1.4: Comportamento das f.a.c. e f.a.c.p. de um processo ARIMA(p,d,q)

Ordem	(1,d,0)	(0,d,1)
comportamento de ρ_k	decai exponencialmente	somente $\rho_1 \neq 0$
comportamento de ϕ_{kk}	somente $\phi_{11} \neq 0$	decaimento exponencial dominante
estimativas iniciais	$\phi = \rho_1$	$\rho_1 = -\frac{\theta}{1+\theta^2}$
região de admissibilidade	$-1 < \phi < 1$	$-1 < \theta < 1$
Ordem	(2,d,0)	(0,d,2)
comportamento de ρ_k	mistura exponenciais ou ondas senóides amortecidas	somente $\rho_1 \neq 0$ e $\rho_2 \neq 0$
comportamento de ϕ_{kk}	somente $\phi_{11} \neq 0$ e $\phi_{22} \neq 0$	dominada por mistura de exponenciais ou senóides amortecidas
estimativas iniciais	$\begin{cases} \phi_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2}, \\ \phi_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2} \end{cases}$	$\begin{cases} \rho_1 = \frac{\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}, \\ \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} \end{cases}$
região de admissibilidade	$\begin{cases} -1 < \phi_2 < 1 \\ \phi_2 - \phi_1 < -1 \\ \phi_2 + \phi_1 < 1 \end{cases}$	$\begin{cases} -1 < \theta_2 < 1 \\ \theta_2 - \theta_1 < 1 \\ \theta_2 + \theta_1 < 1 \end{cases}$
Ordem	(1,d,1)	
comportamento de ρ_k	decai exponencialmente após o "lag"1	
comportamento de ϕ_{kk}	dominada por decaimento exponencial após o "lag"1	
estimativas iniciais	$\rho_1 = \frac{(1-\phi\theta)(\phi-\theta)}{1+\theta^2-2\phi\theta}, \quad \rho_2 = \rho_1\phi$	
região de admissibilidade	$-1 < \phi < 1, \quad -1 < \theta < 1$	

Tabela 1.5: Variâncias aproximadas para \bar{W} , onde $W_t = \Delta^d X_t$, $n = T - d$

AR(1)	MA(1)	ARMA(1,1)
$\frac{C_0(1+r_1)}{n(1-r_1)}$	$\frac{C_0(1+2r_1)}{n}$	$\frac{C_0}{n} \left[1 + \frac{2r_1^2}{r_1-r_2} \right]$
AR(2)		MA(2)
$\frac{C_0(1+r_1)(1-2r_1^2+r_2)}{n(1-r_1)(1-r_2)}$		$\frac{C_0(1+2r_1+2r_2)}{n}$

onde $\hat{\sigma}_{k,l}^2$ é o estimador de máxima verossimilhança de σ^2 para um modelo ARMA(k,l). Assim estipula-se valores limites superiores K e L para k e l e calcula-se o resultado de AIC(k,l) para todas as possíveis combinações (k,l) com $0 \leq k \leq K$ e $0 \leq l \leq L$. Em geral, K e L são funções de T, por exemplo, $K=L=\ln T$. Para o caso de modelos AR(p), o critério AIC limita-se a

$$\text{AIC}(k) = \ln \hat{\sigma}_k^2 + \frac{2k}{T}, \quad k \leq K.$$

Shibata (1976) prova que a minimização do AIC produz estimativas incoerentes com a verdadeira ordem do processo AR. Por isso, Hurvich e Tsai (1989) sugere uma correção para o AIC, com o objetivo de diminuir a probabilidade de selecionar uma ordem maior do que a verdadeira, que no caso é

$$\text{AIC}_c(k) = \text{AIC}(k) + \frac{2(k+1)(k+2)}{T-k+2}, \quad k \leq K$$

na prática nota-se que esta correção é adequada quando T é pequeno ou quando K é uma função "moderadamente grande" de T.

B-Critério de informação Bayesiano BIC Akaike(1977), Rissanem (1978) recomendam diminuir o critério de informação Bayesiano, dado por

$$\text{BIC} = -2\ln(\text{FV}) + (\text{número de parâmetros})\ln T,$$

na qual FV é a função de verossimilhança. Para o processo ARMA, essa quantidade é

$$\text{BIC}(k, l) = \ln \hat{\sigma}_{k,l}^2 + (k+l) \frac{\ln T}{T}.$$

Hannan(1980) comprova que, sob determinadas condições, as estimativas k e l que diminuem a expressão acima são fortemente coerentes.

Estimação

Depois de identificado o modelo temporário para a série temporal, a próxima etapa é estimar os parâmetros. Citamos os métodos dos momentos, mínimos quadrados (MQ) e máxima verossimilhança (MV). Os estimadores do método dos momentos quando confrontado com os demais não são muito bons, desse modo são utilizados como valores iniciais para procedimentos mais complexos, como MV.

I-Modelos Auto-regressivos: Consideremos o modelo AR(p)

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

onde $\varepsilon_t \sim \text{RB}(0, \sigma^2)$. Considere $\phi = (\phi_0, \phi_1, \dots, \phi_p)'$ o vetor contendo os coeficientes e seja $\xi = (\phi, \sigma^2)'$. A intenção é estabelecer ϕ e σ^2 . Para exemplificar, suponha o caso $p=1$, ou seja,

$$X_t = \phi_0 + \phi_1 X_{t-1} + \varepsilon_t.$$

1 FUNDAMENTAÇÃO TEÓRICA

considere que tenhamos os dados $\mathbf{X} = (X_1, \dots, X_T)'$. Então, considerando ε_t com distribuição normal e condicional ao valor $X_0 = x_0$, a função de verossimilhança é dada por

$$L(\xi|\mathbf{X}, x_0) = (2\pi\sigma^2)^{-T/2} \exp\left\{\frac{-1}{2\sigma^2} \sum_{t=1}^T (X_t - \phi_0 - \phi_1 X_{t-1})^2\right\}$$

e supondo a log-verossimilhança temos

$$\begin{aligned} L(\xi|\mathbf{X}, x_0) &\approx (-T/2)\log\sigma^2 - \frac{-1}{2\sigma^2} \sum_{t=1}^T (X_t - \phi_0 - \phi_1 X_{t-1})^2 \\ &= (-T/2)\log\sigma^2 - \frac{-1}{2\sigma^2} S(\phi|\mathbf{X}, x_0). \end{aligned}$$

onde $S(\cdot|\cdot)$ é a soma de quadrados condicional. Segue-se que maximizar l , com σ^2 fixo, é equivalente a minimizar S e estimadores de MV de ϕ serão equivalentes a estimadores de MQ.

Os ε são obtidos recursivamente:

$$\varepsilon_1 = X_1 - \phi_0 - \phi_1 X_0 = X_1 - \phi_0 - \phi_1 x_0,$$

$$\varepsilon_2 = X_2 - \phi_0 - \phi_1 X_1, \text{ etc}$$

Um outro caminho é considerar que apenas os dados verdadeiramente observados e iniciar a regressão de X_t sobre 1 e X_{t-1} a partir de $t = 1, 2, \dots, T$.

Em geral, conseguimos adquirir os estimadores dos coeficientes do modelo pela regressão pela equação do modelo AR(p), para $t = p + 1, \dots, T$. Os resíduos serão dados por $\hat{\varepsilon}_t = X_t - \hat{X}_t$, com

$$\hat{X}_t = \hat{\phi}_0 + \hat{\phi}_1 X_{t-1} + \dots + \hat{\phi}_p X_{t-p}.$$

Um estimador de σ^2 é obtido de

$$\hat{\sigma}^2 = \frac{\sum_{t=p+1}^T \hat{\varepsilon}_t^2}{T - 2p - 1}.$$

Além dos estimadores de MQ, podemos utilizar também os estimadores de Yule-Walker, obtidos das equações de Yule-Walker amostrais, que possui boas propriedades e são fáceis de calcular pelo algoritmo de Durbin-Levinson.

II-Modelos de Médias Moveis Para o modelo MA(q)

$$X_t = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q},$$

onde novamente $\varepsilon_t \sim N(0, \sigma^2)$, podemos considerar também estimadores de MV condicionais. Aqui, supomos que $\varepsilon_0 = \varepsilon_{-1} = \dots = 0$ e obtemos recursivamente

$$\varepsilon_1 = X_1 - \theta_0,$$

$$\varepsilon_2 = X_2 - \theta_0 + \varepsilon_1 \theta_1, \quad \text{etc}$$

e a partir desses valores formamos a função de verossimilhança.

Observe esse exemplo. Consideremos um processo ARIMA(0,1,1),

$$\Delta X_t = (1 - \theta B)\varepsilon_t,$$

e suponha que $\theta = 0,8$. Então, podemos escrever, com $W_t = \Delta X_t$,

$$\varepsilon_t = W_t + 0,8\varepsilon_{t-1}.$$

Suponha que utilizemos os dados (hipotéticos) da Tabela 3. Como

$$\varepsilon_1 = W_1 + 0,8\varepsilon_0,$$

iniciamos ε_t especificando $\varepsilon_0 = 0$ e $X_0 = 150$. Então,

$$\varepsilon_1 = -3 + (0,8) \times 0 = -3,$$

$$\varepsilon_2 = W_2 + 0,8\varepsilon_1 = -4 + (0,8)(-3) = -6,4, \quad \text{etc.}$$

Tabela 1.6: Cálculo recursivo de ε_t , $\theta = 0,8$

t	X_t	$W_t = \Delta X_t$	$\varepsilon_t = W_t + 0,8\varepsilon_{t-1}$
0	150		0
1	147	-3	-3,0
2	143	-4	-6,4
3	148	5	-0,12
4	153	5	4,9
5	149	-4	-0,08
6	155	6	5,9
7	162	7	11,7
8	170	8	17,4
9	172	2	15,9

Segue-se que a SQ condicional fica

$$S(0,8|\mathbf{a}_0 = 0) = \sum_{t=1}^9 \varepsilon_t^2(0,8|\mathbf{a}_0 = 0) = 801,26.$$

Calculando-se S para uma grade suficientemente fina de valores de θ no intervalo $(-1,1)$, já que $-1 < \theta < 1$, podemos obter o mínimo aproximado dessa função. Veja que é possível obter estimadores de MV incondicionais (Morettin e Toloi,2004) e também calculamos os estimadores de MV exatos tratando $\varepsilon_0, \varepsilon_{-1}, \dots$ como parâmetros adicionais a serem estimados. Utilizando T suficientemente grande esses métodos permitem resultados próximos.

III-Modelos de Mistos Estimadores condicionais para modelos ARMA(ou ARIMA) são obtidos como no modelo anterior. Veja como neste caso o modelo ARMA(1,1)

$$X_t - \phi X_{t-1} = \varepsilon_t - \theta \varepsilon_{t-1},$$

1 FUNDAMENTAÇÃO TEÓRICA

supondo $\varepsilon_0 = 0$ e $X_0 = x_0$ obtemos, recursivamente,

$$\begin{aligned}\varepsilon_1 &= X_1 - \theta x_0, \\ \varepsilon_2 &= X_2 - \phi X_1 - \theta \varepsilon_1 = X_2 - (\phi - \theta)X_1 - \phi \theta x_0, \quad \text{etc}\end{aligned}$$

Da mesma forma, estimadores de MV incondicionais e exatos são possíveis de alcançar para modelos ARIMA genéricos.

IV-Variâncias dos estimadores Pela precisão dos estimadores existe a necessidade do intervalo de confiança para os parâmetros. Seja $\boldsymbol{\eta} = (\boldsymbol{\phi}, \boldsymbol{\theta})$, de ordem $K \times 1$, onde $k=p+q$. Para n grande, onde $n=T-d$ (d é a ordem d diferenças) os EMV tem uma distribuição assintótica normal, de modo que

$$\hat{\boldsymbol{\eta}} \xrightarrow{D} N_k(\boldsymbol{\eta}, V),$$

$$V = 2\sigma^2 \begin{bmatrix} \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_1^2} & \dots & \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_1 \partial \eta_k} \\ \vdots & & \vdots \\ \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_k \partial \eta_1} & \dots & \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_k^2} \end{bmatrix},$$

onde $S(\boldsymbol{\eta})$ é a soma de quadrados que aparece na log-verossimilhança, por exemplo. É possível demonstrar que o EMV de σ^2 é

$$\widehat{\text{sigma}}^2 = \frac{s(\hat{\boldsymbol{\eta}})}{n}$$

e que, para n grande, $\hat{\sigma}^2$ e $\hat{\boldsymbol{\eta}}$ são não-correlacionados. Substituindo σ^2 na equação de V acima por $\hat{\sigma}^2$ e calculando as derivadas $\frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_i \partial \eta_j}$ numericamente, obtemos estimativas das variâncias dos estimadores e estimativas das covariâncias entre os estimadores. Então com as estimativas das variâncias teremos os intervalos de confiança para os parâmetros η_i , $i = 1, \dots, k$.

A tabela abaixo mostra as variâncias aproximadas dos estimadores dos modelos mais usados.

Diagnóstico

Depois de estimar o modelo devemos verificar se reflete satisfatoriamente os dados. Se houver alguma deficiência pode ser recomendado um modelo alternativo como sendo o mais apropriado. Uma técnica muito comum, se presumirmos que um modelo com mais parâmetros é necessário, é o superajustamento. Estimamos um modelo com parâmetros extras e examinamos se sua inclusão diminui significativamente a variância residual. A verificação pode ser feita observando os resíduos. Considere o modelo

$$\phi(B)W_t = \theta(B)\varepsilon_t,$$

com

$$W_t = \Delta^d X_t$$

Caso este modelo seja o melhor, logo esses erros ε_t formarão um ruído branco.

Tabela 1.7: Variâncias aproximadas para os estimadores dos parâmetros dos modelos mais comuns; $n = T - d$.

Modelo	Variância
AR(1)	$\text{Var}(\hat{\phi}) = \frac{1-\phi^2}{n}$
AR(2)	$\text{Var}(\hat{\phi}_1) = \text{Var}(\hat{\phi}_2) \simeq \frac{1-\phi_2^2}{n}$
MA(1)	$\text{Var}(\hat{\theta}) \simeq \frac{1-\theta^2}{n}$
MA(2)	$\text{Var}(\hat{\theta}_1) = \text{Var}(\hat{\theta}_2) \simeq \frac{1-\theta_2^2}{n}$
ARMA(1,1)	$\text{Var}(\hat{\phi}) \simeq \frac{(1-\phi^2)(1-\phi\theta)^2}{n(\phi-\theta)^2}$ $\text{Var}(\hat{\theta}) \simeq \frac{(1-\theta^2)(1-\phi\theta)^2}{n(\phi-\theta)^2}$

Teste de auto-correlação residual Estimamos ϕ e θ , as quantidades

$$\hat{\varepsilon}_t = \hat{\theta}_{-1}(B)\hat{\phi}(B)W_t$$

são chamadas resíduos estimados ou simplesmente resíduos. Se o modelo for apropriado, os $\hat{\varepsilon}_t$ deverão estar próximo dos ε_t e, portanto, deverão ser aproximadamente não-correlacionados. Se indicarmos por \hat{r}_k as auto-correlações dos resíduos $\hat{\varepsilon}_t$, então deveríamos ter $\hat{r}_k \simeq 0$. Assim, deveríamos ter, aproximadamente,

$$\hat{r}_k \sim N\left(0, \frac{1}{n}\right),$$

onde $n=T-d$ é o número efetivo de observações e sempre sob a suposição que o modelo ajustado seja apropriado. As auto-correlações \hat{r}_k são calculadas por

$$\hat{r}_k = \frac{\sum_{t=k+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=k+1}^n \hat{\varepsilon}_t^2}$$

Contudo, o desvio padrão de \hat{r}_k pode ser bem menor que $\frac{1}{\sqrt{n}}$, especialmente para pequenos valores k , como mostrou Durbin(1970). Ele demonstrou que para AR(1), $\text{Var}(\hat{r}_k) \simeq \frac{\phi^2}{n}$, que pode ser consideravelmente menor que $1/n$. Box, Jenkins e Reinsel (1994) demonstraram que, para o modelo AR(1), escrevemos

$$\text{Var}(\hat{r}_k) \simeq \frac{1}{n}[1 - \phi^{2(k-1)}(1 - \phi^2)]$$

$$\text{Cov}(\hat{r}_i, \hat{r}_j) \simeq \frac{1}{n}\{\delta_{ij} - \phi^{i+j-2}(1 - \phi^2)\},$$

onde δ_{ij} é o delta de Kronecker. Assim, quando K é de moderada para grande a variância de \hat{r}_k é aproximadamente, $1/n$, e as auto-correlações são não-correlacionadas. Assim a comparação de \hat{r}_k com os limites $\pm \frac{2}{\sqrt{n}}$ nos aponta de modo geral uma quebra

1 FUNDAMENTAÇÃO TEÓRICA

do procedimento de ruído branco em ε_t , deste modo vale lembrar que para valores pequenos de k , este limites desprezará a relevância de qualquer divergência.

V-Teste de Box-Pierce-Ljung Conforme Box e Pierce(1970) existe um teste para as auto-correlações dos resíduos estimados, que, apesar de não detectar quebras específicas no comportamento de ruído branco, pode mostrar se esses valores são muito altos. Ljung e Box (1978) sugere uma modificação neste teste que é

$$Q(K) = n(n+2) \sum_{k=1}^k \frac{\hat{r}_k^2}{(n-k)}$$

terá uma distribuição χ^2 com $k-p-q$ graus de liberdade. Em geral não se utiliza para valores grande de $Q(\cdot)$, isto é, para as 15 ou 20 primeiras \hat{r}_k .

V-Teste de Box-Pierce-Ljung Desejamos prever X_{T+h} , obtendo-se observações até o instante T , usando um modelo ARIMA. Seja

$$\varphi(B) = \phi(B)\Delta^d = (1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_{p+d} B^{p+d})$$

Chamemos de $\hat{X}_T(h)$ a previsão de X_{T+h} de origem T e horizonte h . É possível provar que a previsão de EQMM (erro quadrático médio mínimo) é dada pela esperança condicional de X_{T+h} dado o passado X_T, X_{T-1}, \dots , ou seja,

$$\begin{aligned} \hat{X}_T(h) = & E(\varphi_1 X_{T+h-1} + \dots + \varphi_{p+d} X_{T+h-p-d} \\ & + \theta_0 + a_{T+h} - \theta_1 a_{T+h-1} - \dots - \theta_q a_{T+h-q} | X_T, X_{T-1}, \dots) \end{aligned}$$

Para calcular a previsões usamos os fatos:

$$\begin{aligned} \text{(a)} E(X_{T+j} | X_T, X_{T-1}, \dots) &= \begin{cases} X_{T+j}, & \text{se } j \leq 0 \\ \hat{X}_T(j), & \text{se } j > 0 \end{cases} \\ \text{(b)} E(\varepsilon_{T+j} | X_T, X_{T-1}, \dots) &= \begin{cases} \varepsilon_{T+j}, & \text{se } j \leq 0 \\ 0, & \text{se } j > 0 \end{cases} \end{aligned}$$

Logo, para calcular previsões temos que:

- (a) substituir esperanças passadas ($j \leq 0$) por valores conhecidos, X_{T+j} e ε_{T+j} ;
- (b) substituir esperanças futuras ($j > 0$) por previsões $\hat{X}_T(j)$ e 0.

1.4 Resumo

Este trabalho de conclusão de curso visa apresentar uma sequências de atividades explorando os métodos de análise de séries temporais utilizando os conhecimento estudados na educação básica. Desse modo, sugere atividades que abrangem principalmente os conteúdos de estatística e probabilidade discute as novas tecnologias na formação do pensamento estocástico dos alunos e seus desdobramentos. Este trabalho é dividido em

três partes: As primeiras seções é constituída de uma revisão bibliográfica sobre series temporais e processos estocásticos e a segunda apresenta um plano de aula com sequências didáticas utilizando alguns software. Essas atividades abordam vários conceitos que servem de base para compreender métodos de identificação de séries temporais. Através do método de simulação e interação utilizaremos alguns softwares como o R, Gretl e a planilha eletrônica . A terceira parte apresenta uma discussão sobre o ensino de estocástica por meio da tecnologia.

2 PRÁTICA

2.1 Introdução

Há muito tempo temos vivenciado a abordagem do ensino de probabilidade e estatística e percebemos que nos últimos anos é cada vez mais evidente a importância dada a essas disciplinas. Até meados dos anos 90 os livros de matemática apresentavam os assuntos probabilidade e estatísticas não dando o devido destaque. Principalmente estatística era deixado como último assunto a ser estudado. O estudo desse tema é fundamental na sociedade atual em que informação é muito utilizada numa análise antes das decisões tomadas. Não é por acaso que o tema tem sido alvo de pesquisa em algumas partes do mundo, havendo muitas publicações procurando justificar a sua importância. Vale destacar a implantação da educação estocástica na escola básica. O termo "estocástica" refere-se à interligação entre as áreas de probabilidade e estatística. Termo muito utilizado por vários pesquisadores de ensino, aprendizagem e avaliação de probabilidade e estatística. No Brasil destacamos as pesquisas de Lopes (1998,2003) além dos grupos de trabalho PME-stochastic WorkingGroup e do CERME-Developing Stochastic Thinking que tratam do assunto. Nesta perspectiva, a estocástica "desempenha um papel fundamental na formação para a cidadania" (Ponte e Fonseca, 2000, p.179). Como o exercício de uma cidadania crítica, reflexiva e participativa deve ser desenvolvido desde cedo, pois as crianças também convivem com da dos estocásticos, "desenvolver o pensamento estatístico e probabilístico ao longo da escolaridade constitui um aspecto importante da formação que a escola deve proporcionar" (Abrantes, Serrazina e Oliveira, 1999, p.94). Temos que considerar a ótica da educação estocástica no ensino fundamental pois o pensamento estocástico é uma linha de raciocínio matemático o que faz parte de nosso currículo, bem como a probabilidade e estatística são conceitos fundamentais da estocástica. O ensino da matemática tem como herança a exatidão e o determinismo e o cálculo, sendo contrário à exploração de situações que envolvam aproximação, aleatoriedade e estimação, as quais podem limitar a visão matemática que o aluno poderá desenvolver. Dificultando suas possibilidades de estabelecimento de estratégias para a resolução de problemas que podem aparecer ao longo de sua vida. Atividades com estatística e probabilidade torna-se interessante ao possibilitar ao estudante desenvolver a capacidade de coletar, organizar, interpretar e comparar dados para obter e fundamentar conclusões, muito importante no desempenho de uma atitude científica. Temas importantes na educação para a cidadania, uma vez que possibilitam o desenvolvimento de uma análise crítica sob diferentes aspectos científicos, tecnológicos e/ou sociais. Ainda hoje percebemos a carência de pesquisa em torno do tema estocástica. Talvez, por não ser, por razões históricas, um assunto muito abordado nas salas de aula de escola básica, como parte do currículo da matemática: ou talvez, por não ter atenção merecida dentro dos cursos de licenciatura. Pode ser apontado quatro aspectos que tem retardado o ensino efetivo de estocástica: "(a) O papel da probabilidade e da estatística no currículo; (b) relações entre pesquisa e instrução; (c)

a preparação do professores de matemática; (d) a maneira pela qual a aprendizagem está sendo conduzida" (SHAUGHNESSY, 1992, p. 466).

Objetivo e justificativa do trabalho

O objetivo de trabalho é apresentar uma proposta de ensino que aborda os conteúdos de probabilidade e estatística estudando os métodos ou modelos de ARIMA através de softwares estatísticos. Para justificar os objetivos de nosso trabalho partimos da seguinte pergunta: Pode-se obter informação de fenômenos aleatórios? Considerando que um fenômeno ou experimento aleatório é aquele que repetido varias vezes não produz o mesmo resultado, mesmo que as condições sejam as mesmas. Ou ainda é aquele fenômeno que não sabemos, com certeza, a priori qual será o resultado. Citamos alguns exemplos como a quantidade de chuva que irá cair em determinado local ou quem será o próximo campeão mundial? Tais fenômenos precisam de um modelo matemático diferente para seu estudo. São o que chamaremos de modelos não-determinísticos ou probabilísticos. E como os modelos probabilísticos ou estocásticos são objetos de estudo da análise de séries temporais é importante conhecermos este assunto tão importante para varias áreas como ciências física, engenharia, ciências biológicas e humanas . O estudo de análise de fenômenos probabilísticos tem vários objetivos deste a descrição dos comportamento até as projeções de valores futuros. Este trabalho também encontra justificativas nas orientações curriculares do ensino médio- (Ministério da Educação, Secretaria de Educação Básica, 2006-Orientações curriculares para o ensino médio ; volume 2), quando afirmam: Ao final do ensino médio, espera-se que os alunos saibam usar a Matemática para resolver problemas práticos do cotidiano; **para modelar fenômenos em outras áreas do conhecimento**; compreendam que a Matemática é uma ciência com características próprias, que se organiza via teoremas e demonstrações; percebam a Matemática como um conhecimento social e historicamente construído; saibam apreciar a importância da Matemática no desenvolvimento científico e tecnológico. (p. 25, grifo nosso).

2.2 Considerações iniciais

Nosso objetivo é desenvolver um plano de aula que seja possível aplicar a teoria citada no capítulo 1 através dos conhecimentos de probabilidade e estatística estudados no Ensino Médio.

3 PLANO DE AULA

Preparamos uma aula com algumas atividades em que exploramos alguns conceitos de probabilidade e estatística que são utilizados como ferramenta na análise de séries temporais. Vale lembrar que são sugestões e podem ser modificadas conforme o interesse do professor. Para cada situação utilizaremos um software específico.

- **Objetivos**

- **Objetivos gerais:**

- O principal objetivo é apresentar noções de identificação dos modelos para séries temporais (ARIMA) aplicando conceitos estudados nas aulas de probabilidade e estatística.

- **objetivos específicos:**

- Compreender o cálculo de covariância, autocovariância e autocorrelação;
 - Conhecer e identificar alguns fenômenos aleatórios realizando simulações como o lançamento de moeda comparando o conceito de frequência com o conceito clássico através da observação da quantidade de resultados possíveis para cada possibilidade;
 - Compreender como os gráficos de função autocorrelação são utilizados para identificar modelos de processos AR, MA e ARMA.

- **metodologia**

- A aula será realizada na sala de informática onde uma parte da aula expomos alguns conceitos sobre o assunto, porém o resto da aula os alunos desenvolvem atividades cada um sua máquina utilizando aplicativos onde realizam simulações. Desse modo estamos utilizando a metodologia da interação e simulação.

- **Ação didática ou roteiro de aula**

- **1º momento**

- Apresentar conceitos sobre fenômenos aleatórios. Ao simular alguns fenômenos orientamos a manipulação de alguns comandos na planilha eletrônica relativos a aleatoriedade. Utilizando o comando **ALEATÓRIO()** junto com a função lógica **SE(teste lógico; valor se verdadeiro; valor se falso)**. Observando uma atividade que conforme a planilha abaixo realiza-se o comando **SE(ALEATÓRIO() > 0,5; "Cara"; "Coroa")** ou, para facilitar, então colocar **Cara = 1 e Coroa = 0**. Deste modo podemos apresentar o conceito de frequência de probabilidade e simular o lançamento de uma moeda, por exemplo, 500 vezes e então calcular as frequências relativas e comparar o

resultado com a probabilidade obtida pelo conceito clássico. Observe a figura abaixo que mostra o experimento. Um conceito que pode acarretar dúvidas, pois envolve passagem ao limite e desta forma não permite visualização e também não é totalmente intuitivo. Solicitar que faça experimentos com roletas ou cartas de baralho e propor relato das conclusões sobre as observações.

Tabela 3.1: Conceito de frequência e probabilidade

Lançamentos	Cara/Coroa	Freq. acumulada	Freq. relativa	Porcentagem
1	0	0	0	0%
2	1	1	0,50	50%
3	1	2	0,67	67%
4	0	2	0,50	50%
5	1	3	0,60	60%
6	1	4	0,67	67%
7	0	4	0,57	57%
8	1	5	0,63	63%

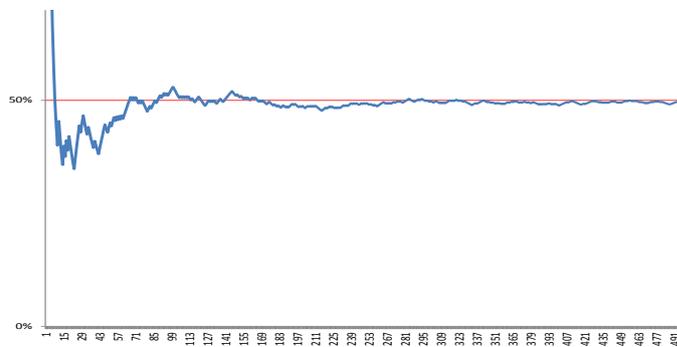


Figura 3.1: Figura 5

2º momento

Apresentar definição e o modo de calcular a covariância, autocovariância e autocorrelação. Solicitar realização de cálculos de variância, autocovariância e autocorrelação na planilha. Os cálculos devem ser feitos de duas formas: primeiro de forma que todas as operações são escritas nas células da planilha eletrônica e no segundo caso utilizando os comandos de funções. Assim como na tabela abaixo o cálculo da diferença entre o elemento da sequência é a média da sequência sendo assim $X_n - \bar{X}$, onde \bar{X} é a média. Consideramos $\sum (Dif 1) \cdot (Dif 2)$ a soma do produto da diferença 1 e diferença 2. O cálculo do produto dos desvios padrões fazemos $(Dp1) \cdot (Dp 2)$. Para o cálculo da autocovariância por exemplo a covariância defasagem 1 fazemos $\gamma(1) =$

3 PLANO DE AULA

$cov\{X_t, X_{t+1}\} = \frac{\sum(Dif1).(Dif2)}{n} = \frac{13,40}{7} = 1,91$. Para calcular o autocorrelação fazemos $\sigma(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{Autocovariância}{(Dp1).(Dp2)} = \frac{1,91}{2,68} = 0,71$. Depois os cálculos são realizados utilizando as funções COVAR(Sequência 1;Sequência 2) e CORREL(Sequência 1;Sequência 2).

Tabela 3.2:

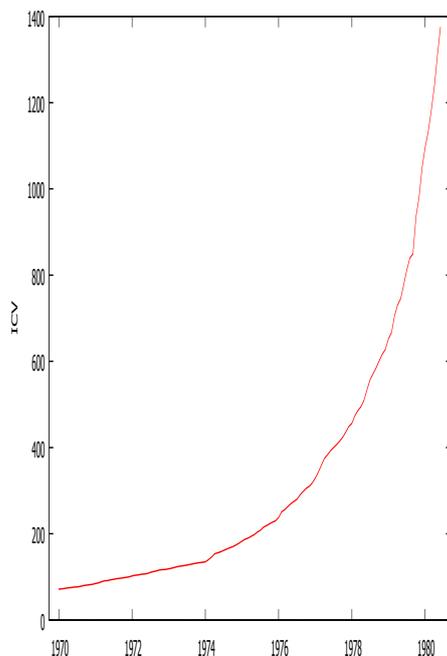
X_n	Sequência 1	Sequência 2	Dif 1	Dif 2	(Dif 1).(Dif 2)
42	42	38,8	2,89	0,66	1,90
38,8	38,8	39,5	-0,31	1,36	-0,43
39,5	39,5	39,7	0,39	1,56	0,60
39,7	39,7	39,5	0,59	1,36	0,79
39,5	39,5	38,1	0,39	-0,04	-0,02
38,1	38,1	36,2	-1,01	-1,94	1,97
36,2	36,2	35,2	-2,91	-2,94	8,58
35,2					
	Média 1	Média 2		$\sum(Dif 1).(Dif 2)$	13,40
	39,11	38,14		Autocovariância	1,91
	Desvio padrão 1	Desvio padrão 2		AutoCorrelação	0,71
	1,63	1,65			
	(Dp 1).(Dp 2)				
	2,68				

3º momento

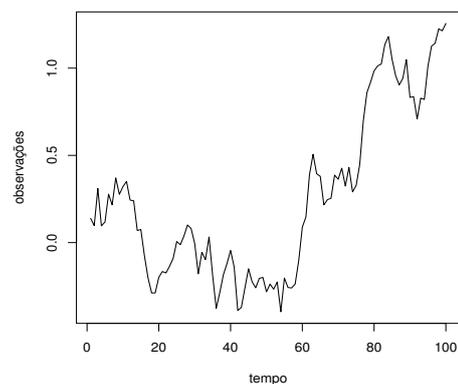
Tabela 3.3: Índice de custo de vida no Município de São Paulo

Ano	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
1970	71,6	72,5	73,5	74,5	75,2	76,3	76,9	78,1	80	80,9	81,7	82,9
1971	84,7	86,3	88,8	90,9	91,5	93,4	94,6	95,9	96,7	97,8	99,1	100
1972	102	104	105	106	107	108	110	112	114	116	117	118
1973	119	120	122	124	125	126	128	129	131	132	133	134
1974	135	140	146	153	156	158	162	165	167	170	174	178
1975	183	188	190	194	198	204	208	215	219	223	227	230
1976	238	251	256	263	270	275	280	290	298	305	310	318
1977	329	343	359	375	383	393	400	407	415	424	436	449
1978	456	474	486	495	510	535	558	572	586	602	617	628
1979	653	667	707	731	746	778	812	840	849	936	980	1049
1980	1096	1133	1182	1237	1309	1374						

Neste momento o objetivo é apresentar conceitos e noções de séries temporais e processos estocásticos estacionários. Através de dados coletados de fenômenos aleatórios podemos apresentar alguns exemplos de processos estocásticos não-estacionários e estacionários. Na tabela acima apresentamos as observações do custo de vida no município de São Paulo (ICV-SP). Com estes dados geramos o gráfico abaixo de um processo estocástico no qual fica bem claro que tanto média quanto variância não são constantes.



(a) Gráfico do índice de custo de vida de São Paulo - 1970 a 1980



(b) Gráfico de um passeio aleatório

Figura 3.2: Gráfico de processos não-estacionários (a) e (b)

Um outro exemplo muito importante e muito usado de série não estacionária é o passeio aleatório ou passeio de bêbado. Aqui é importante salientar que $\{\varepsilon_t, t \in \mathbb{Z}\}$ é ruído branco discreto se as v.a. ε_t são não correlacionadas. Define-se um passeio aleatório a partir do ruído branco gerando a sequência $X_t = \varepsilon_1 + \dots + \varepsilon_t$. Assim $E(X_t) = t\mu_\varepsilon$ e $\text{Var}(X_t) = t\sigma_\varepsilon^2$, ou seja, a média, a variância e as autocovariâncias dependem de t , portanto o processo não é estacionário conforme observamos no gráfico anterior. Mas se observarmos o gráfico da diferença do passeio aleatório podemos perceber que a média e a variância são constantes e assim o processo é estacionário. Observe o gráfico abaixo:

3 PLANO DE AULA

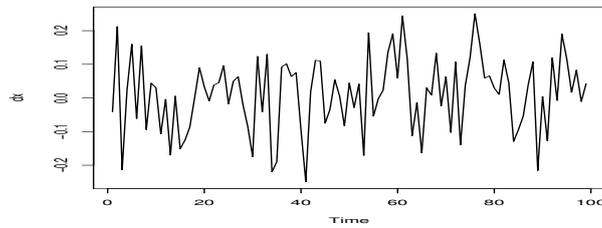


Figura 3.3: Gráfico de exemplo de processo estacionário

4º momento

Apresentar a função autocorrelação como ferramenta para identificar os processos estocásticos. Uma característica importante nos processos estocásticos não estacionários é o decaimento das autocorrelações de maneira muito lenta. Veja abaixo a função autocorrelação de um passeio aleatório (Correlograma):

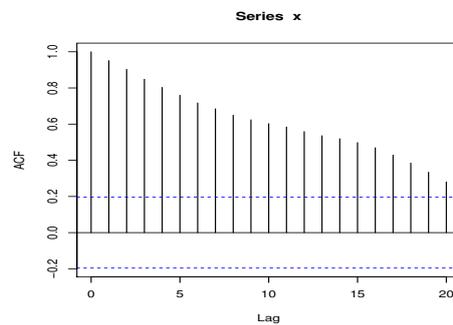
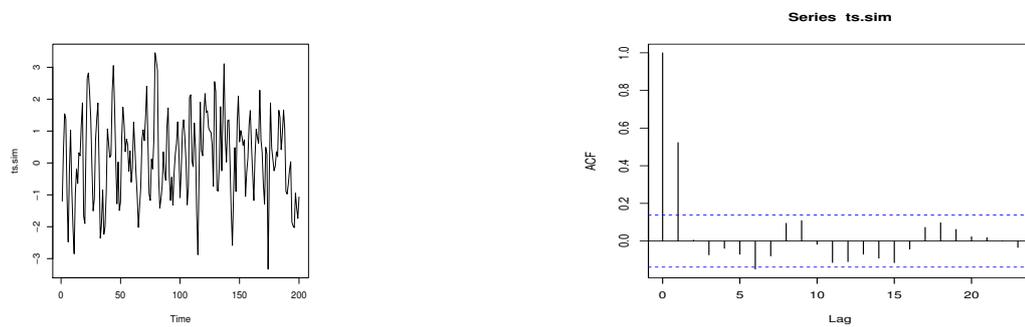


Figura 3.4: Gráfico de processo estacionário

Pretende-se mostrar que num processos $AR(p)$ tem f.a.c. que decai de acordo com exponenciais e/ou senóides amortecidas, infinita em extensão. E um processo $MA(q)$ tem f.a.c. finita, no sentido que apresenta um corte após o " lag " q . Essas características são apresentadas nos correlogramas dos processos $MA(1)$:



(a) Gráfico do processo $MA(1,1)$

(b) F.a.c. do processo $MA(1,1)$

Figura 3.5: Gráfico e correlograma de processos MA (a) e (b)

No processo $ARMA(p,q)$ devemos mostrar que a f.a.c é infinita em extensão, que decai de acordo com exponenciais e/ou senóides amortecidas após " lag " $q-p$.

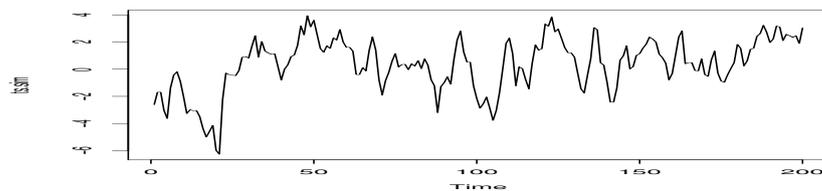


Figura 3.6: Gráfico do processo $ARMA(1,1)$

3 PLANO DE AULA

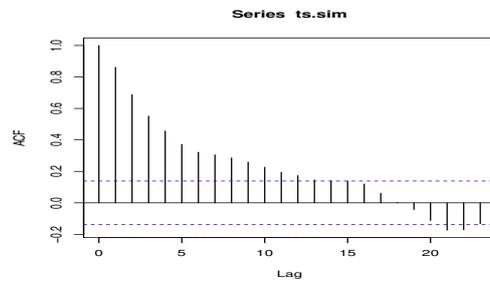


Figura 3.7: Função autocorrelação do processo ARMA(1,1)

4 CONCLUSÃO

4.1 Discussão (outros trabalhos)

A presente discussão pretende responder a seguinte questão: É possível estudar processos estocásticos na Educação Básica?

Primeiro vamos tentar entender o que são os processos estocásticos. Segundo Morettin(2006) os modelos utilizados para descrever séries temporais são processos estocásticos, isto é, processos controlados por leis probabilísticas. E série temporal é qualquer conjunto de observações ordenadas no tempo. São exemplos de séries temporais:

- registro de marés no porto de Santos
- valores mensais de temperatura na cidade de São Paulo
- índice diário da Bolsa de valores de São Paulo

Hoje os educadores e pesquisadores na área da educação tem usado o termo "Estocástica" fazendo referência à interface entre a estatística e a probabilidade entendendo que tais assuntos quando estudados na Educação básica possibilitam o desenvolvimento de formas particulares de pensamento e raciocínio, compreendendo fenômenos aleatórios, interpretação de amostras e elaboração de inferências.

Sendo assim, a estocástica é um meio importantíssimo para o desenvolvimento das capacidades de análise e resolução de situações problemáticas reais, de raciocínio e comunicação, bem como da autoconfiança na aplicação de conhecimentos adquiridos.(NUNES, 2008) Além disso, através das novas tecnologias, é possível com o uso do computador a construção e simulação de modelos que permite avaliar o comportamento de problemas dinâmicos - aqueles que sofrem alterações ao longo do tempo. Será que é possível simular alguns processos estocásticos através de alguns aplicativos específicos e criar situações de aprendizado para alunos da Educação Básica?

Recorremos ao artigo "Ensino de estocástica por meio de Simulação virtual" de Leandro de Oliveira Souza e Celi Spasadin Lopes(2011). O artigo é parte de uma pesquisa de mestrado de natureza qualitativa com análise interpretativa segundo categorias emergente: simulação, interação e resolução de problemas que investigou as contribuições da inserção da tecnologia para a educação estocástica. O processo de análise caminhou para elaboração de estudo de caso sobre grupo formado por quatro alunas com as quais se desenvolveram as atividades e se investigou o seguinte problema: com os recursos tecnológicos podem contribuir para a construção de novos conhecimentos da Estocástica no Ensino Fundamental? Para responder essa questão, utilizou-se a perspectiva Vygotskiana, considerando que, através de atividades de ensino, se possa promover a aprendizagem por meio da internalização. Comprovou-se que tais recursos geram conhecimento mais amplo e precisos, destacando a importância da simulação e do processo de interação na educação. Segundo os autores foram desenvolvidos quatro atividades: duas de estatística, aplicadas a duas turmas de aproximadamente 40 alunos, em uma

4 CONCLUSÃO

escola Municipal de São José dos Campos; e mais duas de Probabilidade, aplicadas a um grupo de quatro alunas voluntárias dessas turmas. Neste artigo será descrito as atividades que focalizam a Probabilidade. Os autores aplicaram a um grupo reduzido de alunos devido a possuírem apenas uma licença do software Fathom, no qual eles reconheceram um grande potencial para o ensino de Estatística e Probabilidade, o que nos levou a utiliza-lo como simulador de eventos aleatórios na pesquisa.

Os autores optaram por fazer gravações da discussão em áudio e trabalhar com um número pequeno de alunos, para que pudessem levantar dados, participando do processo de ensino aprendizagem e intermediando-os, dar condições para analisara esses dados com maior precisão. Segundo os autores embora o objetivo central fosse investigar as contribuições dos recursos tecnológicos à aquisição de conhecimentos probabilísticos, também puderam ressaltar considerações relacionadas à pratica docente e ao processo de interação dos aluno durante a construção do conhecimento.

O uso de simulações no ensino da estocástica

Em consonância com as diretrizes postuladas por Chance e Rossman (2006), neste estudo, os autores buscaram evidenciar que o processo de ensino e aprendizagem da Estocástica no Ensino Fundamental pode ser ampliado e tornar-se mais significativo com o uso dos computadores: estes podem auxiliar os estudantes a adquirir conhecimento sobre coleta de dados e sobre organização, descrição, análise e interpretação; e a utilizá-los para tomada de decisões.

Era esperado que as quatro alunas acima mencionadas fizessem comparações do jogo da roleta com um outro jogo, anteriormente desenvolvido em Souza (2009), que levava os alunos a estimar a probabilidade da soma dos resultados obtidos pelo lançamento indicado nas faces superiores de dois dados. Pretendiam, através da comparação dos jogos, que as estudantes descobrissem que existiam injustiças na organização dos jogos e que elas podem ser resolvidas, conhecendo as chances reais de cada jogador.

Considerações finais

Os autores relatam que no decorrer de toda essa atividade, ficou nítido que a ferramenta computacional foi uma grande aliada para a aquisição e a construção dos novos conhecimentos. Porém é fundamental que as atividades estejam interligadas com a prática e com simulações reais e que sejam feitas constantes intervenções pelos professores. Sem estas e sem um bom planejamento de aula, seria impossível para o aluno autoestimular-se a utilizar jogos como esses e buscar neles informações pertinentes ao estudo da Probabilidade.

É notável durante todo o desenvolvimento das atividades, as alunas enfrentaram o dilema: Seriam os jogos justos ou não?; ganhar seria pura sorte? Isso causou desconfiança, o que estimulou o aprendizado.

É importante destacar aqui que J. Bordier (apud COUTINHO,1994) elaborou um questionário seguindo algumas recomendações que julgamos importantes para o desenvolvimento de atividades que permitem a aquisição das noções de probabilidade; são elas:

- As questões devem ser direcionadas para as representações espontâneas da probabilidade e não sobre problemas qu exigem habilidades lógicas ou matemáticas particulares. Elas devem propor ao aluno situações que o levem a fazer uma escolha entre diversas possibilidades.
- Deve-se pedir que o aluno comente suas respostas indicando o que levou a fazer determinadas escolha, sem necessidade de justificações matemáticas para tal.

(COUTINHO,1994, p. 34).

Os dois softwares utilizados ajudaram muito: com eles foi feito as simulações de várias amostras, o que economizou um tempo precioso de aula e auxiliou as alunas a praticar na máquina aquilo que seria impossível fazer sem auxílio desta.

Na pesquisa aqui referida, durante todo o projeto, as aulas apresentaram significativas habilidades relativas à resolução de problemas. Destacaram-se em relação às categorias emergentes consideradas para a análise: simulação; interação; interação e resolução de problemas; capacidade para criar métodos de coleta de dados, aplica-los, tabula-los e interpreta-los, usando recursos tecnológicos. O autor destaca que provavelmente tais atividades pudessem ter sido desenvolvidas sem o uso desses recursos, de forma que as alunas construíssem os gráficos e simulassem as atividades manualmente, porém perder-se-ia a oportunidade da construção dos diferentes tipos de gráficos e o processo de iteração.

Fica claro que através de quatro atividades desenvolvidas foi possível levar os alunos a construção e à simulação de dados, e também desenvolver o trabalho com estatística e probabilidade, em que prevaleceu análise crítica.

Para Batanero et e al. (2005) as simulações são ferramentas fundamentais no início do ensino fundamental, pois permitem que os alunos construam modelos e experimentem com fenômenos aleatórios e previsões a longo prazo, dando a real interpretação para a frequência dos fenômenos probabilísticos, o que seria praticamente impossível sem a rapidez dos simuladores tecnológicos. Respondendo a questão do primeiro parágrafo podemos através dos resultados do estudo de campo do trabalho de Souza e Lopes concluímos que através de um bom planejamento utilizando simuladores ou softwares adequados podemos simular vários processos de modo que o alunos possam obter informações pertinentes, analisar, obter suas conclusões levando a melhora de suas habilidades probabilísticas e estatísticas e o desenvolvendo de seu pensamento estocástico.

Atualmente, o uso das tecnologias de informação e comunicação, através de software específico de Estatístico ou recorrendo à internet, no ensino e na aprendizagem da Estatística é amplamente preconizado por educadores e investigadores, mais frequentemente no tema de Estatística do que em qualquer outro tema de Matemática. Segundo Branco e Martins (2002) o uso das tecnologias “é, hoje em dia, um aspecto fundamental da prática da Estatística” (p. 13) e Jolliffe (2007) advoga que as maiores alterações do ensino da Estatística são consequência da revolução tecnológica. A matemática escolar deve propiciar cada vez mais a investigação, a reflexão e a criatividade, rompendo com o determinismo que geralmente inibe a ideia de movimento e transformação, tão importante na elaboração do conhecimento. Isso, quando o enfoque do processo educativo não é a aquisição de informações, mas sim a percepção das relações contidas nos temas investigados (LIPMAN, 1995). Como diz Agostinho da Silva "O importante é educar as pessoas para perguntar, não para responder.

4.2 Considerações finais

Iniciamos nosso estudo sobre processos estocásticos nos questionando sobre a possibilidade de desenvolver uma situação de ensino aprendizagem envolvendo esse assunto na educação básica. Tópico muito estudado nas áreas de engenharia, ciências físicas,

4 CONCLUSÃO

ciências biológicas e humanas. Ao iniciar este trabalho tínhamos um determinado conhecimento sob importância da probabilidade e estatística dentro da Matemática. Hoje passado alguns meses vejo o estudo dessas disciplinas de modo bem diferente. Nosso objetivo era desenvolver uma sequência didática para trabalhar processos estocásticos na escola Básica e através de estudo de vários trabalhos e da fundamentação teórica concluímos nosso objetivo desenvolvendo uma sequência didática para trabalhar o tema. Esperamos contribuir para formação dos nossos jovens e na prática de ensino ajudando os colegas a conhecer outros caminhos entre tantos que a matemática nos oferece.

Bibliografia

- [1] MORETTIN, P.A.; *Econometria Financeira: Um Curso de Séries Temporais Financeiras*, São Paulo, 2006.
- [2] COUTINHO, S. Q. S.A.; *Introdução ao Conceito de Probabilidade por uma visão Frequentista. Estudo Epistemológico de Didático*, São Paulo, 1994. Dissertação de Mestrado em Educação Matemática. PUC/SP. São Paulo.
- [3] Chance, B.; and Rossman, A. ; *Investigating Statistical Concepts, Applications, and Methods*, Belmont, CA: Duxbury, 2006.
- [4] NUNES, A. R.S. ; *Ensino da estocástica no 6.º ano de escolaridade: opções metodológicas e dificuldades sentidas pelos professores*, Especialização em Ensino e Aprendizagem da Matemática, Universidade do Minho, 2008.
- [5] SOUZA, L. O. ; LOPES, C. E; *O Ensino de Estocástica por meio de Simulação Virtual*, In: CONFERÊNCIA INTERAMERICANA DE EDUCAÇÃO MATEMÁTICA, XIII, 2011, Recife.
- [6] BATANERO, C; GODINO; CAÑIZARES, m.j.; *Using simulation to bridge teacher's content and pedagogical Knowledge in probability*. Proceedings of the fifteenth International Commission on Mathematics Instruction Study (pp. 1-6). Águas de Lindóia, Brazil: ICMI, 2005.
- [7] LIPMAN, Matthew; *O pensar na Educação*. Petrópolis: Vozes, 1995
- [8] SILVA, I. A.; *Probabilidade: a visão laplaciana ou visão frequentista*, PUC/SP - 2002 - Dissertação (Mestrado em educação profissional) - São Paulo
- [9] BRANCO, J. & Martins, M. E. G.; *Literacia estatística. Educação e Matemática*, 69, 9-13 2002.
- [10] JOLLIFE, F.; *The changing brave new world of statistics assessment*; The Proceedings of the ISI/IASE Satellite on Assessing Student Learning in Statistics, Voorburg: International Statistical Institute (2007).
- [11] PONTE, J. P. e FONSECA, H.; *A estatística no currículo do ensino básico e secundário*, Lisboa - Sociedade Portuguesa de estatística (2000).
- [12] ABRANTES, P., SERRANI e OLIVEIRA, I; *A matemática na educação básica* Lisboa, 1999.
- [13] VYGOTSKY, L.S.; *A formação social da mente.*, 4ª ed . São Paulo: Martins Fontes, 1991.

Bibliografia

- [14] SHAUGHNESSY, J. M.: *Research in probability and statistics: reflections and directions*. In: GROUWS, D. A. (Ed.). Handbook of Research on Mathematics Teaching and Learning. USA: NCTM, 1992. p. 465-494.