

**UNIVERSIDADE DE SÃO PAULO**

**Métodos Estocásticos para Previsão de Carga  
em Sistemas de Energia Elétrica**

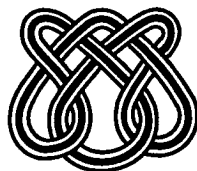
Marinho G. Andrade

N<sup>o</sup> 49

---

**NOTAS DIDÁTICAS**

---



**Instituto de Ciências Matemáticas de São Carlos**

UNIVERSIDADE DE SÃO PAULO  
Instituto de Ciências Matemáticas e de Computação  
ISSN 0103-2585

**Métodos Estocásticos para Previsão de Carga  
em Sistemas de Energia Elétrica**

Marinho G. Andrade

N<sup>o</sup> 49

NOTAS DIDÁTICAS



São Carlos  
Jun./2001

# Métodos Estocásticos para Previsão de Carga em Sistemas de Energia Elétrica

**Marinho G. Andrade**

Instituto de Ciências Matemáticas e Computação - ICMC  
Departamento de Computação e Estatística - SCE  
Universidade de São Paulo - USP

São Carlos -SP  
2001



Quero agradecer a **Selene Loibel**  
pela paciência e dedicação  
na leitura dos manuscritos originais



# Sumário

	iii
<b>Prefácio</b>	<b>ix</b>
<b>Introdução</b>	<b>xi</b>
<b>I Revisão de Probabilidade e Estatística</b>	<b>1</b>
<b>1 Conceitos de Probabilidade e Estatística</b>	<b>3</b>
1.1 Estatística Descritiva . . . . .	3
1.2 Axiomas de Probabilidade . . . . .	10
1.3 Probabilidade Condicional . . . . .	11
1.4 Função Distribuição de Probabilidade . . . . .	14
1.5 Valor Esperado e Variância . . . . .	20
1.6 Distribuição de Probabilidade Conjunta . . . . .	20
1.7 Soma e Produto de Variáveis Aleatórias . . . . .	25
1.8 Covariância e Correlação . . . . .	27
1.9 Densidade de Probabilidade Condicional . . . . .	29
1.10 Valor Esperado e Variância Condicional . . . . .	31
1.11 Técnica de Regressão . . . . .	35
<b>2 Tópicos de Inferência Estatística</b>	<b>43</b>
2.1 Inferência Estatística . . . . .	43
2.2 Estimativa de Máxima Verossimilhança (MLE) . . . . .	46
2.2.1 Qualidade do Estimador . . . . .	48
2.3 Inferência Bayesiana . . . . .	49
2.3.1 Algoritmo Amostrador de Gibbs . . . . .	51

<b>3</b>	<b>Tópicos de Processos Estocástico</b>	<b>53</b>
3.1	Estacionariedade . . . . .	53
3.2	Função de Autocorrelação (FAC) . . . . .	56
<b>II</b>	<b>Modelos para Séries Temporais</b>	<b>59</b>
<b>4</b>	<b>Modelos com Tendência e Sazonalidade</b>	<b>61</b>
4.1	Introdução . . . . .	61
4.2	Modelo com Tendência . . . . .	62
4.2.1	Métodos de regressão para tendência . . . . .	62
4.2.2	Diferenciação discreta . . . . .	63
4.3	Modelos com Sazonalidade . . . . .	66
4.3.1	Método dos momentos sazonais . . . . .	67
4.3.2	Método de regressão (série de Fourier) . . . . .	70
4.3.3	Diferenciação discreta sazonal . . . . .	74
<b>5</b>	<b>Modelos Lineares Estacionários</b>	<b>75</b>
5.1	Processo $AR(p)$ . . . . .	76
5.1.1	Função de Autocorrelação do $AR(p)$ . . . . .	77
5.1.2	Função de Autocorrelação Parcial do $AR(p)$ . . . . .	78
5.2	Processo $MA(q)$ . . . . .	81
5.2.1	Função de Autocorrelação do $MA(q)$ . . . . .	81
5.2.2	Função de Autocorrelação Parcial do $MA(q)$ . . . . .	82
5.3	Processo $ARMA(p, q)$ . . . . .	84
5.3.1	Função de Autocorrelação do $ARMA(p, q)$ . . . . .	84
5.3.2	Função de Autocorrelação Parcial do $ARMA(p, q)$ . . . . .	85
5.4	Ajuste de Modelo Linear Estacionário . . . . .	86
5.4.1	Ajustando um modelo $AR(p)$ . . . . .	87
5.5	Critério de Seleção de Modelo . . . . .	88
5.6	Previsão . . . . .	91
5.6.1	Previsão com modelo $ARMA(p, q)$ . . . . .	92
5.6.2	Previsão com modelo $AR(p)$ . . . . .	93
5.7	As Etapas do Ajuste de um modelo linear . . . . .	95
<b>6</b>	<b>Um Exemplo de Previsão de Carga.</b>	<b>97</b>
6.1	Cosumo Industrial de Energia Elétrica . . . . .	97
6.1.1	A Tendência . . . . .	99



6.1.2	A Sazonalidade . . . . .	101
6.1.3	A Componente Estocástica . . . . .	102
6.1.4	A Previsão . . . . .	105
6.1.5	Conclusão . . . . .	108



# Prefácio

O objetivo deste trabalho é familiarizar o engenheiro do setor elétrico com técnicas estatísticas e modelos de séries temporais utilizadas para identificação, estimação, seleção de modelos para previsão de picos de demanda e consumo de energia elétrica. O texto pretende passar uma visão crítica dos procedimentos de análise e seleção de modelos para previsão de carga, mostrando que não há rigorosamente um procedimento padrão para se fazer previsão de carga e sim uma análise de cada caso individualmente, buscando as variáveis mais importantes para explicar o comportamento da série de carga de acordo com o horizonte de previsão.

O texto é técnico o suficiente para capacitar um profissional da área a analisar e escolher o melhor modelo para o seu problema de previsão de carga. No entanto o texto contém uma revisão dos principais tópicos de probabilidade, estatística e processos estocásticos visando minimizar o esforço do leitor na busca de literatura adicional, se bem que leituras adicionais serão necessárias para se aprofundar em qualquer uma das técnicas apresentadas neste trabalho.

O texto inicia com uma introdução sobre o problema de previsão de carga e em seguida é dividido em duas partes; na primeira temos três capítulos de revisão. O Capítulo 1 faz uma revisão de probabilidade. O Capítulo 2 faz uma revisão de alguns tópicos de inferência e o Capítulo 3 apresenta alguns tópicos de processos estocásticos. Mesmo sendo estes capítulos de revisão, eles ilustram os conceitos com exemplos voltados ao problema de previsão de carga. Na segunda parte do texto encontram-se os capítulos voltados para a análise e previsão de séries temporais. O Capítulo 4 apresenta modelos para tendência e sazonalidade de séries temporais. O Capítulo 5 apresenta os modelos lineares para processos estocásticos estacionários (Modelos de Box & Jenkins) e por fim no Capítulo 6 temos alguns exemplos de problemas de previsão de carga ilustrando a aplicação do conteúdo do texto.

No médio prazo estão presentes, para alguns seguimento do mercado consumidor de energia, uma sazonalidade. Por exemplo no setor rural esta sazonalidade é atribuída a períodos de colheitas, no setor comercial a sazonalidade é atribuída aos hábitos da sociedade, etc. Assim, técnicas tais como suavização e decomposição espectral (série de Fourier), usadas na previsão de curto prazo, são aplicadas para descrever a sazonalidade anual da carga.

- 3) Modelos ARIMA

Esta técnica, como já dissemos, é utilizada nas previsões de médio e longo prazo descrevendo a correlação temporal entre as observações mensais ou anuais. Também são utilizadas para descrever a correlação temporal da componente estocástica da carga que se soma a tendência (ajustada por polinômios) e à sazonalidade (ajustada por série de Fourier).

- 4) Modelos multivariados (análise de regressão)

O uso desta técnica para previsão de médio prazo visa descrever o consumo de energia em função de variáveis climáticas tais como temperatura máxima do verão e mínima do inverno e variáveis socio-econômicas tais como renda per capita da região, potência média dos eletrodomésticos residenciais, densidade demográfica, etc. Em geral esses modelos são de difícil ajuste devido à falta de uniformização nas diferentes fontes de informação dos dados.

- Técnicas de previsão de longo prazo

Essas técnicas geralmente consideram modelos macroeconômicos e política de desenvolvimento, tais como benefícios para implantações de indústrias em uma região, política habitacional, desenvolvimento agrário promovido por irrigações em larga escala e disponibilidades de fontes alternativas de energia. Em geral o desenvolvimento socioeconômico de uma região está atrelado ao crescimento no consumo de energia elétrica, de tal forma que o consumo per capita é um dos indicadores de desenvolvimento social. Assim todas as previsões de crescimento social são indicadores de crescimento no consumo de energia elétrica a longo prazo.

- 3) Previsão de longo prazo

Consiste em previsões feitas para um horizonte de dez a vinte anos em base mensal, semestral ou anual. Neste horizonte as variáveis a serem previstas são os consumo de energia das diferentes classes de consumidor (residencial, comercial, industrial, etc). As ferramentas mais comuns para estas previsões são os modelos econométricos e neste caso as previsões obtidas com modelos matemáticos devem ser confrontadas com uma análise econômica para diferentes cenários de desenvolvimento. A previsão de longo prazo é necessária principalmente para avaliação e planejamento da expansão da capacidade do parque gerador de energia elétrica, uma revisão desses modelos é apresentada em [9]. É com base nesse planejamento que se decide sobre a construção de uma nova usina hidroelétrica ou uma termoelétrica para assegurar um baixo risco de déficit no atendimento do mercado consumidor.

Podemos notar portanto que a classificação das técnicas por horizonte de previsão não é perfeita pois os objetivos destas previsões dentro de um mesmo horizonte pode levar à técnicas bem distintas e impossíveis de serem comparadas, mesmo assim apresentamos uma tentativa de classificação das técnicas até então aplicadas para previsão de carga segundo o horizonte de previsão, isso não quer dizer que uma técnica usada na previsão de curto prazo seja totalmente inadequada para previsões de médio e longo prazo.

- Técnicas de previsão de curto prazo

- 1) Decomposição espectral (ou série de Fourier)

Esta abordagem descreve séries periódicas como a soma de funções trigonométricas. Essa técnica é adequada para previsões de carga horária porque descreve bem a sazonalidade desta carga ao longo das 24 horas de um dia.

- 2) Métodos de suavização exponencial (alisamento)

Esta técnica visa eliminar o efeito de ruídos aleatórios dos dados observados resultando em um conjunto de pontos médios que descrevem períodos de séries sazonais, como é o caso de carga horária. Mas este é um exemplo típico de técnica que pode ser usada para determinar sazonalidade e tendência em séries de observações mensais visando previsões de médio e longo prazo.

esse aumento tornará disponível uma energia de  $100MW \times 365dia \times 24hora$  ou seja  $876.000MW-hora$  de energia. Mas, se a previsão do consumo de energia é de um consumo somente de  $200.000 MW-hora$ , então significa que deve ser instalado uma unidade de pico (ou ponta), ou seja uma unidade para operar somente nas horas em que a demanda instantânea precisa ser atendida em vez de uma unidade de base. A diferença entre essas duas é geralmente substancial.

A Previsão de carga é necessária também para se estabelecer programas de operação de sistemas geradores, onde se deseja minimizar o custo com combustível e avaliar o risco de falha no atendimento do mercado consumidor. Em resumo, uma boa previsão, refletindo as tendências e sazonalidades futuras do mercado com um bom julgamento por parte do previsor é a chave de um bom planejamento.

## • Classificação das Técnicas de Previsão

Ao se estudar as técnicas de previsão de carga um dos primeiros objetivos é classificá-las visando encontrar um método de escolha da técnica mais apropriada para resolver o problema de previsão abordado. Existem várias maneiras de se classificar os modelos de previsão de carga, uma delas é classificá-los segundo a variável que se deseja prever, assim temos:

- Classificação de acordo com a variável a ser prevista
  - 1) Modelos para previsão de pico de demanda (potência  $MW$ )
  - 2) Modelos para previsão de consumo de energia (energia  $MW-hora$ )
  - 3) Modelo para previsão de curva de carga horária (potência  $MW$ )

No entanto essa não é a forma mais adequada de classificação porque as técnicas variam muito de acordo com o horizonte de previsão para uma mesma variável. Por exemplo várias técnicas completamente diferentes podem ser usadas para determinar o pico de demanda anual de uma empresa como mostra [4]. Além disso a escolha da técnica tem que levar em conta o horizonte de previsão necessário. Não seria possível fazer uma comparação entre as técnicas de previsão de pico de demanda diário (com base horária) com modelos de previsão de pico de demanda anual (em base mensal). No primeiro caso é fundamental que o modelo considere variáveis climáticas tais como: temperatura, umidade relativa do ar, luminosidade, etc, [5]. Enquanto

# Introdução

A previsão de consumo de energia elétrica e demanda de potência são cruciais para o planejamento e operação de sistemas geradores e distribuidores de energia elétrica. Neste trabalho as palavras “*carga*”, “*energia*” e “*demanda*” são usadas com frequência. O significado dessas palavras aqui são os seguintes: Carga é um termo geral que pode significar potência demandada ou energia elétrica consumida. A palavra Demanda tem significado aqui de taxa de variação de energia no tempo, ou seja é a potência instântanea ou pico de potência e sua unidade de medida é *Watts(W)* ou *Mega-Watts(MW)* se referindo a  $1 \times 10^6W$ , enquanto que a energia é medida em *MW-hora*.

O termo “*previsão*” refere-se a projeção das necessidades de carga para o futuro. Tais projeções são feitas, geralmente, com técnicas estatísticas aprimoradas e bom senso do previsor. A precisão de uma previsão é fundamental para qualquer empresa fornecedora de energia elétrica, pois é com base nessa previsão que a empresa tomará decisões sobre a política de expansão da sua capacidade para atender ao mercado consumidor e sobre a política de operação do sistema gerador de energia. Por exemplo, uma previsão que erra para menos, prevendo uma demanda inferior a que de fato ocorre pode resultar em perdas de rendimentos, pois o sistema opera despreocupado com a escassez de recursos e pode ser surpreendido por uma demanda maior do que sua capacidade. Por outro lado uma previsão acima do real pode resultar em problemas financeiros com investimentos que não produzem retorno suficiente por falta de demanda e criando capacidade ociosa no sistema.

A previsão da demanda (potência) é necessária para se planejar a expansão da capacidade de geração, transmissão e distribuição de sistemas de energia elétrica e a previsão de consumo de energia visa apoiar o decisor quanto ao tipo de equipamento ou instalação necessária. Por exemplo se a previsão de demanda prever para o futuro a necessidade de um aumento da capacidade do sistema para atender uma demanda de  $100MW$ , isto significa que

para previsão de pico de demanda anual essas variáveis, quando são consideradas, ficam reduzidas à temperatura máxima do verão e mínima do inverno além disso é fundamental que se considere as tendências de crescimento anual do mercado assim como variáveis relacionadas à política econômica [10]. Portanto as técnicas devem ser distintas quando temos objetivos distintos com a previsão de carga. Assim uma sugestão de classificação das técnicas de previsão é classifica-las segundo o horizonte de previsão.

- Classificação segundo o horizonte de previsão

Para fazer uma classificação das técnicas segundo o horizonte de previsão, vamos definir primeiro quais os horizontes mais comuns nos problemas de previsão de carga.

- 1) Previsão de curto prazo

Consiste em previsões feitas para o horizonte de um dia (24 horas) ou uma semana em base horária ou fração de uma hora [3], [6]. Neste tipo de problema a variável a ser prevista é a demanda de potência para toda empresa ou por subestação. O objetivo destas previsões geralmente estão relacionados com problemas operacionais da rede elétrica tais como programação de manutenção, segurança, manobras, etc. Estas previsões são portanto subsídio para cálculo de confiabilidade do sistema. É comum os modelos de previsão a curto prazo levar em consideração variáveis climáticas tais como temperatura, luminosidade, etc.

- 2) Previsão de médio prazo

Consiste em previsões feitas para um horizonte de um a cinco anos, essas previsões são geralmente feitas em base semanal ou mensal [7]. No médio prazo são necessárias previsões de demanda de potência, para planejamento da expansão da capacidade do sistema para evitar corte de carga nas horas de pico [11]. Neste horizonte também são necessárias as previsões de consumo de energia para planejamento da operação energética do sistema visando minimizar custo com combustível gasto para manter termoeletricas ligadas por um longo período. Nas previsões de consumo de energia no médio prazo geralmente o mercado consumidor é dividido em classes tais como, residencial, industrial, comercial, rural, etc com objetivo de se analisar tendências e sazonalidade de cada classe separadamente [1].



- 3) Modelos ARIMA

Esta técnica relaciona as observações da carga em um dado instante de tempo (hora) com observações de instantes anteriores somados à combinações de ruídos aleatórios. Em geral essas técnicas são utilizadas para considerar correlações entre os resíduos que não são modelados por funções que descrevem o comportamento periódico da série de carga horária. Mas aqui também temos uma técnica que é potencialmente utilizada nas previsões de médio e longo prazo descrevendo a correlação temporal entre as observações mensais ou anuais.

- 4) Modelos em espaço de estado (Filtros Lineares)

Estas técnicas são usadas na previsão de curto prazo principalmente por causa de sua adaptatividade, que permite a atualização de parâmetros do modelo automaticamente (tempo real) e recursividade, a medida que novas observações são adquiridas, possivelmente a cada minuto, a recursividade desta técnica permite que as previsões sejam refeitas sem a necessidade de considerar todo histórico de dados já utilizados na estimação dos parâmetros, essa propriedade torna essa técnica bastante atraente para o problema de operação on-line de sistemas.

- 5) Modelos multivariados (análise de regressão)

Estes modelos levam em conta variáveis climáticas tais como temperatura, nebulosidade, umidade relativa do ar, etc e busca estabelecer uma relação matemática entre essas variáveis e a carga horária.

- Técnicas de previsão de médio prazo

- 1) Técnica de extrapolação de tendência

Nesta técnica são ajustados polinômios para descrever a tendência de crescimento na carga devido ao aumento de população e desenvolvimento regional. Esta técnica em geral não é sensível à mudanças na política econômica que podem afetar o crescimento de uma região resultando em graves erros de previsão, por isso essa técnica deve sempre ser acompanhada de uma análise econômica conjuntural

- 2) Técnicas de ajuste de sazonalidade

**Parte I**

**Revisão de Probabilidade e  
Estatística**

# Capítulo 1

## Conceitos de Probabilidade e Estatística

### 1.1 Estatística Descritiva

Porque a abordagem probabilística é necessária em problemas de engenharia? Uma abordagem probabilística pode ser usada para considerar nossa incapacidade de modelar perfeitamente alguns aspectos de um problema. Outra razão é a incapacidade de se fazer medidas perfeitas de certas variáveis do problema. Ou, finalmente, porque certos fenômenos na natureza são aleatórios e só podemos descrevê-los de forma probabilística.

Probabilidade e Estatística são importantes em muitas aplicações de engenharia, só para exemplificar, algumas aplicações são listadas a seguir:

- Previsão de valores futuros de variáveis aleatórias.
- Caracterização da confiabilidade de sistemas.
- Descrição estatística (média, desvio-padrão, moda, mediana, percentil) de variáveis aleatórias.
- Tomada de decisão com observação parcial de dados.
- Análise do efeito de ruídos em componentes.

**Exemplo 1** *Neste exemplo vamos analisar estatisticamente medidas de corrente na saída de um transformador de corrente. Supondo que o transfor-*

#### 4 CAPÍTULO 1. CONCEITOS DE PROBABILIDADE E ESTATÍSTICA

mador foi projetado para produzir uma corrente de 5 Amperes quando alimentado corretamente. Um experimento planejado em laboratório resultou em uma amostra de 30 medidas de corrente apresentadas na Tabela 1.

*Tabela 1 Medidas de corrente em Ampere*

5.0000	5.8956	5.5689	4.7660	5.6232
5.2379	4.6821	5.7310	4.7444	5.1184
5.7990	3.9922	5.4282	5.6771	3.5249
6.0950	5.5779	4.6225	5.3148	5.9409
4.2580	6.0823	4.6490	5.2120	4.8685
3.1260	5.0403	4.7041	6.4435	4.0079

Para analisar esses dados estatisticamente, primeiramente podemos fazer um gráfico das medidas para verificar a variabilidade existente. Para isso é necessário escolher adequadamente a escala do gráfico, uma escala muito ampla em relação a variabilidade das medidas pode levar erroneamente a conclusão de que as medidas apresentam baixa variabilidade como mostra a Figura 1a.

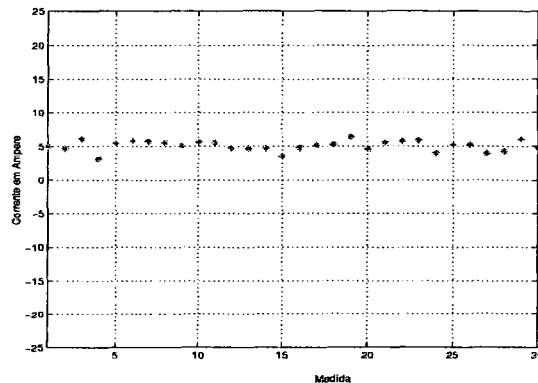


Figura 1a Gráfico com escala ampliada

Por outro lado uma escala reduzida pode levar a uma supervalorização

da variabilidade como mostra a Figura 1b.

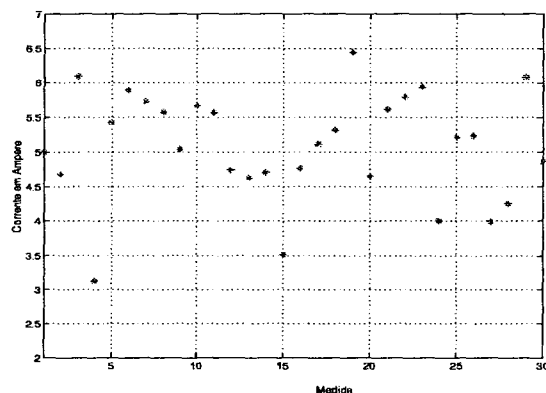


Figura 1b Gráfico com escala reduzida

Para analisar estatisticamente os dados da Tabela 1, uma primeira abordagem é construir um gráfico que mostre como as observações podem ser classificadas de acordo com sua magnitude uma alternativa é a construção de um Histograma.

- Histograma ou Distribuição de Freqüências

O Histograma é uma representação gráfica da distribuição de uma amostra em classes construídas de forma a incluir toda faixa de variação dos dados observados. A construção de um histograma segue os passos:

1. Observar os valores mínimo  $X_{\min}$  e máximo  $X_{\max}$  da amostra.
2. Escolher o número de classes  $NC$  ( $NC \leq 2 \ln(N)$ ).
3. Calcular a amplitude de cada classe:  $\Delta = (X_{\max} - X_{\min})/NC$ .
4. Calcular as classes  $[X_k, X_{k+1})$ ,  $k = 1, \dots, NC$ , fazendo:

$$X_1 \leq X_{\min}, \quad X_{k+1} = X_k + \Delta \quad \text{e} \quad X_{NC} \geq X_{\max}$$

Para o Exemplo 1 temos:  $X_{\min} = 3.1260$ ,  $X_{\max} = 6.4435$ ,  $NC = 5$ ,  $\Delta = 0.6635$ .

Tabela 2 Distribuição de frequência do Exemplo 1

Classe	Faixa	Média	Frequência
1	[3.1260 , 3.7895)	3.45775	2
2	[3.7895 , 4.4530)	4.12125	3
3	[4.4530 , 5.1165)	4.78475	9
4	[5.1165 , 5.7800)	5.44825	10
5	[5.7800 , 6.4435]	6.11175	6

## 6 CAPÍTULO 1. CONCEITOS DE PROBABILIDADE E ESTATÍSTICA

O Histograma referente a distribuição de frequência dada na Tabela 2 é:

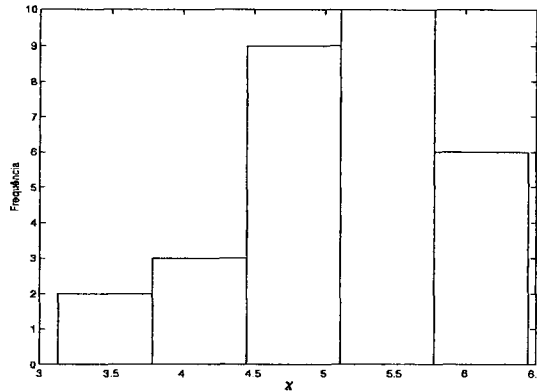


Figura 2 Histograma dos dados do Exemplo 1

Surge então a necessidade de se estimar algumas quantidades estatísticas para descrever o comportamento aleatório da variável a ser analisada.

As estatísticas descritivas normalmente utilizadas são:

- Média Amostral

Denotando por  $X_i, i = 1 \dots n$  as medidas observadas, a média amostral representa uma medida central em torno da qual as observações estão variando. A média amostral é calculada por:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1.1)$$

- Variância Amostral

A variância amostral é uma medida da dispersão das observações em torno da média amostral. A variância amostral é calculada por:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) \quad (1.2)$$

Uma medida de dispersão com ordem de grandeza mais adequada é o desvio-padrão amostral, calculado como:

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} = \sqrt{\frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)} \quad (1.3)$$

As medidas de dispersão dadas em (1.2) e (1.3) são úteis para avaliar se uma amostra de uma variável aleatória está muito dispersa em relação a sua média amostral. No entanto essas estatísticas dependem da ordem de grandeza da variável analisada, por exemplo: Sucessivas medidas do diâmetro de uma peça com um micrômetro acusaram uma média amostral de 5.49mm e um desvio-padrão de 0.012mm e várias medidas de uma corda efetuadas com outro micrômetro acusaram medidas de de 0.75in e desvio-padrão de 0.003in. Qual dos dois micrômetro é o mais preciso? Para responder essa questão calculamos os coeficientes de variação, definidos por:

- Coeficiente de Variação:

Expressa o desvio-padrão como percentagem da média.

$$CV = \frac{S}{\bar{X}} 100\% \quad (1.4)$$

Assim, para o exemplo dos micrômetros temos:

$$CV_{\text{diâmetro}} = \frac{0.012}{5.49} 100\% \approx 0.22\% \quad CV_{\text{corda}} = \frac{0.003}{0.75} 100\% \approx 0.40\%$$

Temos que  $CV_{\text{diâmetro}} < CV_{\text{corda}}$ , portanto o primeiro micrômetro é mais preciso que o segundo.

O coeficiente de variação permite fazer uma avaliação da dispersão de uma variável aleatória em relação a sua média. Mas, essa estatística nada diz sobre a forma como essa dispersão ocorre, ou seja se a dispersão ocorre de forma simétrica ou não em torno da média. Para analisar melhor essa dispersão são necessárias outras estatísticas descritivas:

- Mediana Amostral:

É o valor observado que divide a amostra em dois grupos com mesma quantidade de observações. A mediana amostral pode ser estimada como:

$$\bar{X}_m = \begin{cases} X_{(n+1)/2}, & \text{se } n \text{ é ímpar} \\ \frac{1}{2}(X_{n/2} + X_{1+n/2}), & \text{se } n \text{ é par} \end{cases} \quad (1.5)$$

Quando a mediana coincide com a média, temos uma indicação de simetria dos dados em torno da média.

## 8 CAPÍTULO 1. CONCEITOS DE PROBABILIDADE E ESTATÍSTICA

- Coeficientes de Assimetria

O coeficiente de assimetria de Pearson pode estimado por:

$$CK = \frac{(\bar{X} - \bar{X}_m)}{S} \quad (1.6)$$

Para uma distribuição perfeitamente simétrica temos  $CK = 0$ , de modo geral o valor de  $CK$  situa-se entre 1 e  $-1$  como mostra a Figura 3 e a divisão pelo desvio-padrão torna os valores de  $CK$  independente da ordem de grandeza da variável amostrada.

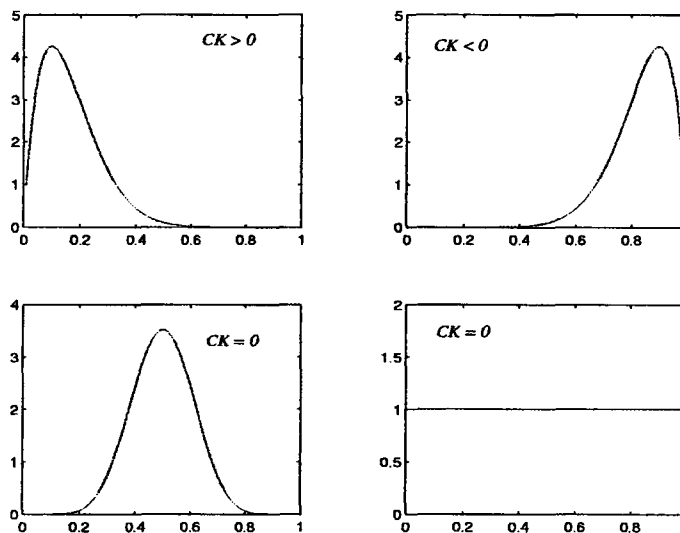


Figura 3 Exemplos de distribuições assimétricas

- Moda

É definida como o valor que ocorre com maior frequência. A moda é também utilizada para descrever o meio ou o centro de um conjunto de dados e tem a vantagem de poder ser calculada também para dados não numéricos. Quando os dados tem distribuição simétrica em torno da média, então a moda coincide com a média e a mediana.

- Percentil

São valores que divide a amostra, ordenados em ordem crescente, dentro de um conjunto de 100 partes iguais. Assim o percentil  $P_k$  pode ser



definido como o valor tal que  $k\%$  dos dados são menores e  $(100 - k)\%$  são maiores ou iguais a  $P_k$ . Podemos aproximar o  $k$ -ésimo percentil,  $P_k$ , a partir dos dados de uma amostra por:

$$P_k = \text{o valor do } \left( \frac{kn}{100} \right) \text{-ésimo termo da amostra ordenada} \quad (1.7)$$

Assim para a amostra dada na Tabela 1, temos o percentil 80 (ou 80%) dado por  $P_{80} = 5.731$ , ou seja o percentil 80 corresponde ao valor  $X_{24}$  da amostra. Quando o número calculado por (1.7) não é inteiro, aproxima-se o percentil pela média dos valores vizinhos, por exemplo, para calcular o percentil 62 da amostra da Tabela 1 temos  $(62 \cdot 30/100) = 18.6$ , que não é inteiro, então usa-se como aproximação  $P_{62} = (X_{18} + X_{19})/2 = 5.3715$ . Examinando a amostra vamos ver que 60% dos valores estão abaixo de  $P_{62}$  e não 62% como queríamos, esse é um erro de aproximação ao qual se está sujeito quando se utiliza amostras para fazer estimativas.

### Exemplo 2 Cálculo de estatísticas descritiva da mostra do Exemplo 1

Tabela 3 Estatísticas descritiva

Média	5.0910
Mediana	5.1652
Moda	4.6188
Desvio-Padrão	0.7834
Coef. de Variação	15.38 %
Coef. de Assimetria	$-9.4642 \times 10^{-2}$

Um dos objetivos da análise estatística de uma variável aleatória é avaliar qual a chance de que uma variável aleatória assuma valores superior a certo número ( $a \leq X$ ), ou qual a chance da variável não ultrapassar um valor máximo desejado ( $X \leq b$ ), ou mesmo qual a chance da variável permanecer em um intervalo  $(a, b]$  pré-estabelecido ( $a < X \leq b$ ). Com base na análise dos dados de uma amostra essa estimativa só pode ser feita sujeita a erros. Os cálculos de probabilidade envolvendo variáveis aleatória só podem ser feitos de forma mais preciso quando se conhece a expressão matemática de sua distribuição de probabilidade, ou seja, quando temos um modelo probabilístico que represente o comportamento aleatório da variável em questão.

## 1.2 Axiomas de Probabilidade

A abordagem necessária para se construir um modelo probabilístico para uma variável aleatória ( $X$ ) começa pela definição do espaço amostral ( $\Omega$ ), ou seja o espaço onde a variável aleatória pode assumir todos seus possíveis valores. O espaço amostral de uma variável aleatória é definido de acordo com a natureza da variável: contínua, discreta, limitada, ilimitada, etc Os exemplos seguintes ilustram alguns tipos de espaço amostral.

**Exemplo 3** *Espaços amostrais comumente usados*

- 1)  $X = \{\text{número de falhas de um equipamento}\}$ .  
 $\Omega = \{0, 1, \dots\}$  é o conjunto dos números inteiros  $N$ .
- 2)  $X = \{\text{variação no nível de tensão na saída de um transformador}\}$ .  
 $\Omega = \{-\infty < x < \infty\}$  é o conjunto dos números reais  $R$ .
- 3)  $X = \{\text{consumo de energia elétrica em uma residência}\}$ .  
 $\Omega = \{x \in R : x \geq 0\}$  é a semi reta positiva.

O segundo conceito importante na construção de um modelo probabilístico é o conceito de *evento*. Um *evento* é definido como qualquer subconjunto do espaço amostral. Aqui os *eventos* são denotados por letras maiúscula  $A, B, C, \dots$ , para os exemplos de espaço amostral dados no exemplo anterior podemos definir os seguintes exemplos de eventos:

**Exemplo 4** *Alguns eventos relacionados ao espaço amostral do Exemplo anterior*

- 1)  $A = \{\text{pelo menos 3 falhas}\}$ ,  $A = \{x \in N : x \geq 3\}$ .
- 2)  $B = \{\text{uma variação inferior a 5KV}\}$ ,  $B = \{x \in R : |x| < 5KV\}$ .

**Definição 5** *Alguns eventos de particular interesse são:*

- 1) O evento certo  $\Omega$  (espaço amostral)
- 2) O evento complementar  $A^c$ , de um evento  $A$ :  $A^c = \{x \in \Omega : x \notin A\}$
- 3) O evento impossível (o conjunto vazio)  $\emptyset$ .
- 4) União de eventos:  $A \cup B = \{x \in \Omega : x \in A \text{ ou } x \in B\}$ .
- 5) Intersecção de eventos:  $A \cap B = \{x \in \Omega : x \in A \text{ e } x \in B\}$ .

A análise de um problema cuja saída é uma variável aleatória  $X$ , que assume valores em um espaço amostral  $\Omega$ , implica diretamente na definição de alguns *eventos*,  $A, B, \dots$ , de interesse para o analista. Portanto para

que essa análise seja matematicamente bem formulada é fácil imaginar que devemos ter uma coleção (ou um conjunto)  $\mathcal{A}$  de todos os possíveis eventos. Esse conjunto de eventos denominamos  $\sigma$ -Algebra de eventos.

**Definio 6** Uma coleção  $\mathcal{A}$  de eventos de  $\Omega$  é uma  $\sigma$ -Algebra se:

- 1)  $\Omega \in \mathcal{A}$
- 1) Se  $A \in \mathcal{A} \implies A^c \in \mathcal{A}$
- 2) Se  $A_1, A_2, \dots \in \mathcal{A}$  e  $A_i \cap A_j = \emptyset$  para  $i \neq j \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Com a definição de espaço amostral e coleção de eventos ( $\sigma$ -Algebra), o modelo probabilístico se completa associando-se uma medida de probabilidade a cada *evento* de  $\mathcal{A}$ . A probabilidade de um evento qualquer  $A \in \mathcal{A}$ , é um número que deve satisfazer aos seguintes axiomas:

**Axioma 7** :Axiomas de Probabilidade

- 1)  $P(A) \geq 0$
- 2)  $P(\Omega) = 1$
- 3) Se  $A_1, A_2, \dots \in \mathcal{A}$  com  $A_i \cap A_j = \emptyset$  para  $i \neq j$ , então:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

como consequência dos axiomas de probabilidade pode ser mostrado que:

- 1)  $P(\emptyset) = 0$ .
- 2)  $P(A^c) = 1 - P(A)$ .
- 3) Se  $B \subset A \implies P(B) \leq P(A)$ .
- 4)  $0 \leq P(A) \leq 1$  para qualquer evento  $A \in \mathcal{A}$ .

## 1.3 Probabilidade Condicional

Em muitos problemas de engenharia há interesse em avaliar como a ocorrência de um certo evento afeta a probabilidade de outro, por exemplo: Qual a probabilidade de se observar uma variação da tensão na saída de um transformador superior a 5KV, dado que esse transformador já falhou mais de três vezes? Para responder essa pergunta vamos definir os eventos:

- a)  $A = \{\text{uma variação superior a } 5KV\} \Leftrightarrow A = \{x \in R : |x| > 5KV\}$ .
- a)  $B = \{\text{ocorreu pelo menos 3 falhas}\} \Leftrightarrow B = \{x \in N : x \geq 3\}$ .

Queremos calcular a probabilidade  $P(A|B)$  que pode ser calculada intuitivamente para uma situação particular em que observamos o comportamento de  $N$  transformadores idênticos dos quais  $N_B$  falharam mais de 3 vezes e entre estes formam observados conjuntamente a ocorrência dos eventos  $A$  e  $B$ ,  $N_{AB}$  vezes. Então podemos deduzir, quando  $N$  é suficientemente grande, que:

$$P(A|B) = \frac{N_{AB}}{N_B} = \frac{N_{AB}/N}{N_B/N}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B)P(B) \quad (1.8)$$

Os eventos  $A$  e  $B$  são considerados independentes quando a ocorrência de um não afeta a probabilidade de ocorrência do outro. Neste caso  $P(A|B) = P(A)$  e  $P(A \cap B) = P(A)P(B)$ .

#### • Probabilidade Total

Uma aplicação direta do conceito de probabilidade condicional é o cálculo da probabilidade de um evento  $B$  que pode ser causado por vários outros eventos disjuntos  $A_1, A_2, \dots, A_n$  que constituem uma partição do espaço amostral, ou seja  $A_i \cap A_j = \emptyset$  e  $\bigcup_{i=1}^n A_i = \Omega$ . Então:

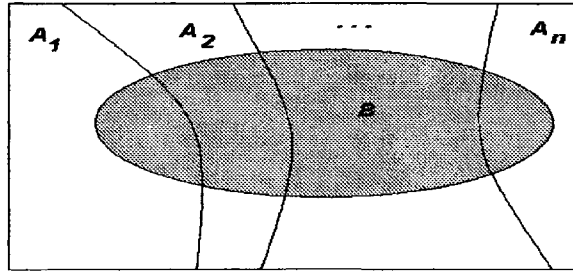


Figura 4 Partição de um espaço amostral

A probabilidade do evento  $B$  pode ser calculada conhecendo-se as probabilidades condicionadas  $P(B|A_i)$  e  $P(A_i)$  para  $i = 1, \dots, n$ , por:

$$P(B) = P\left(\bigcup_{i=1}^n B \cap A_i\right) = \sum_{i=1}^n P(B \cap A_i)$$

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i) \quad (1.9)$$

**Exemplo 8** Para exemplificar o uso da equação (1.9) vamos imaginar uma situação onde o fornecimento de energia elétrica para uma região pode falhar devido a três possíveis causas:

$$A_1 = \{\text{Sobregarga}\} \quad A_2 = \{\text{Falha na rede}\} \quad A_3 = \{\text{Falha humana}\}$$

$$P(A_1) = 0.35 \quad P(A_2) = 0.50 \quad P(A_3) = 0.15$$

Denotando por  $B = \{\text{ocorrência de falha no fornecimento}\}$  e sabendo que:

$$P(B|A_1) = 0.25 \quad P(B|A_2) = 0.45 \quad P(B|A_3) = 0.30$$

Podemos calcular a probabilidade de haver falha no fornecimento como:

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)$$

$$P(B) = 0.25 \times 0.35 + 0.45 \times 0.50 + 0.30 \times 0.15 = 0.3575$$

#### • Teorema de Bayes

Uma outra relação muito utilizada que faz uso da probabilidade condicionada é o *Teorema de Bayes*. Esse teorema afirma que:

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{P(B)} \quad (1.10)$$

usando a equação (1.9) podemos escrever o teorema de Bayes como:

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

No contexto do exemplo anterior, esse teorema permite calcular a probabilidade do evento ( $A_k$ ) ser a causa quando a falha ( $B$ ) ocorreu.

**Exemplo 9** Considerando os dados do exemplo anterior, e dado que ocorreu uma falha no fornecimento de energia. Qual a causa mais provável?

Para responder essa questão aplicamos (1.10) para os três eventos,  $A_1, A_2, A_3$ :

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B)} = \frac{0.25 \times 0.35}{0.3575} = 0.2447$$

$$P(A_2|B) = \frac{P(B|A_2)P(A_2)}{P(B)} = \frac{0.45 \times 0.50}{0.3575} = 0.6294$$

$$P(A_3|B) = \frac{P(B|A_3)P(A_3)}{P(B)} = \frac{0.30 \times 0.15}{0.3575} = 0.1259$$

Portanto a causa mais provável da falha no fornecimento é o evento  $A_2 = \{\text{Falha na rede}\}$ .

Nos exemplos anteriores os valores das probabilidades dos *eventos* envolvidos são considerados conhecidos. No entanto, na realidade, determinar a probabilidade da demanda de energia elétrica de uma região ser superior a certo nível crítico não é uma tarefa simples. Um caminho para se calcular as probabilidades dos *eventos* envolvidos no problema é descobrir uma função matemática adequada, definida no espaço amostral da varável aleatória em questão, que permita calcular as probabilidade dos *eventos* que interessam para a solução do problema. Essa função recebe o nome de Função Distribuição de Probabilidade.

## 1.4 Função Distribuição de Probabilidade

Vamos considerar que estamos lidando com uma situação que exige o conhecimento do comportamento de uma variável aleatória  $X$ . Para resolver esse problema é necessário conhecer a Função Distribuição de Probabilidade desta variável.

**Definio 10** Se  $X$  é uma varável aleatória contínua, assumindo valores em toda reta, existe uma função  $F_X(x)$ , não decrescente, contínua a direita, assumindo valores em  $[0, 1]$ . Tal que:

$$F_X(a) = P(X \leq a) \quad (1.11)$$

e com as propriedades limites:  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  e  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ .

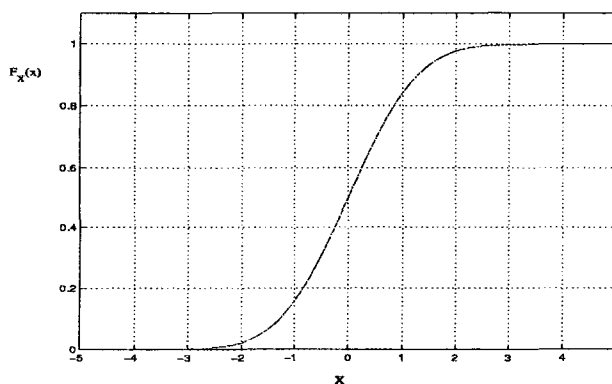


Figura 5 Exemplo de Função Distribuição de Probabilidade

O Cálculo de probabilidade com função  $F_X(x)$  pode ser feito usando a definição (1.11) e os axiomas de probabilidade (1)-(3), como mostram os seguintes exemplos:

**Exemplo 11** *Cálculo da probabilidade  $P(a < X \leq b)$*

$$\{X \leq b\} = \{x \leq a\} \cup \{a < X \leq b\} \Rightarrow P(\{X \leq b\}) = P(\{x \leq a\} \cup \{a < X \leq b\})$$

Usando o axioma 3, e a definição de  $F_X(x)$  podemos escrever:

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

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

**Exemplo 12** *Cálculo da probabilidade  $P(X > b)$*

$$\Omega = \{x \leq b\} \cup \{X > b\} \Rightarrow P(\Omega) = P(\{x \leq b\} \cup \{X > b\}) = 1$$

Usando os axiomas 2 e 3 e a definição de  $F_X(x)$ , temos:

$$P(X > b) = 1 - P(X \leq b) = 1 - F_X(b)$$

**Exercício 13** *Mostre que a função:*

$$F_X(x) = \frac{1}{\pi} \left( 1 + \frac{2}{\pi} (\tan^{-1}(x)) \right)$$

*é uma Função Distribuição de Probabilidade.*

Se  $X$  é uma variável aleatória contínua, muitas vezes é mais conveniente adotar como modelo probabilístico a **função densidade de probabilidade** em vez da função distribuição. A função densidade de probabilidade é definida como:

**Definição 14** *Se  $X$  é uma variável aleatória contínua com Função Distribuição de Probabilidade  $F_X(x)$ , então existe uma função  $f_X(x)$  definida como função densidade de probabilidade, dada por:*

$$f_X(x) = \frac{dF_X(x)}{dx} \quad (1.12)$$

O teorema fundamental do cálculo permite calcular  $F_X$  em função de  $X$  por:

$$F_X(x) = \int_{-\infty}^x f_X(s) ds \quad (1.13)$$

A monotonicidade e as propriedades limites da função  $F_X(x)$  levam às seguintes propriedades para a função densidade:

$$f_X(x) \geq 0 \quad (1.14)$$

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1 \quad (1.15)$$

As probabilidades calculadas nos exemplos anteriores, usando  $F_X(x)$  podem ser feitas com a função densidade de probabilidade como mostram os exemplos seguintes:

**Exemplo 15** *Cálculo da probabilidade  $P(a < X \leq b)$*

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

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

**Exemplo 16** *Cálculo da probabilidade  $P(X > b)$*

$$P(X > b) = 1 - F_X(b) = 1 - \int_{-\infty}^b f_X(x) dx = \int_b^{+\infty} f_X(x) dx$$

Qualquer função satisfazendo as propriedades (1.14) e (1.15) é uma função densidade de probabilidade. Uma das funções densidade de probabilidade mais importantes é a função Normal ou Gaussiana.

**Definio 17** *Uma variável aleatória é dita normalmente distribuída com parâmetros  $\mu$  e  $\sigma^2$  se sua função densidade de probabilidade é:*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right\}, \quad -\infty \leq x \leq +\infty \quad (1.16)$$



Na densidade normal,  $\mu$  é o parâmetro de locação e  $\sigma$  é o parâmetro de forma, como mostram a Figura 6a e 6b respectivamente.

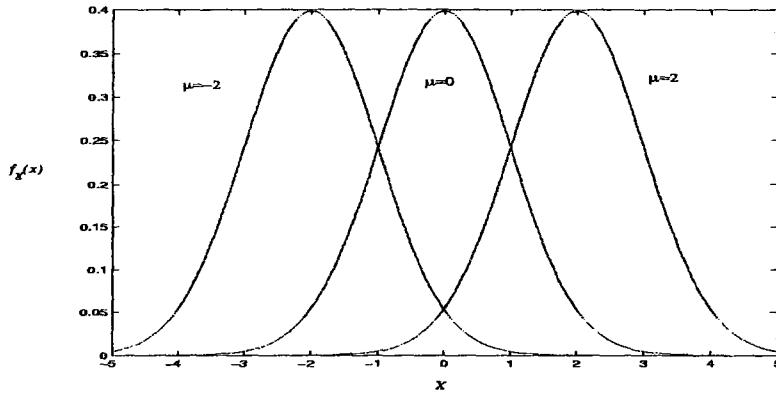


Figura 6a Variação da distribuição Normal em função dos parâmetros  $\mu$

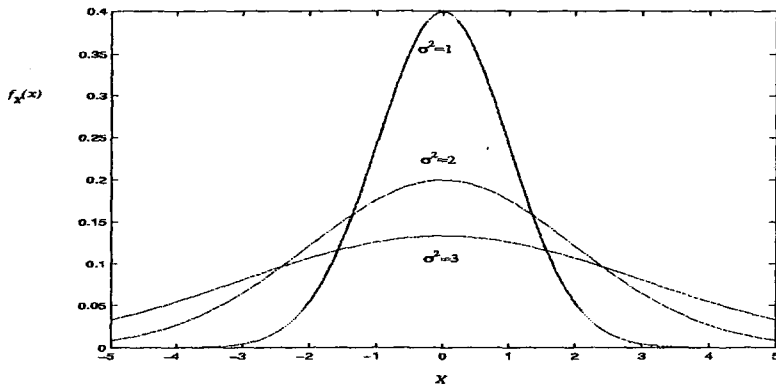


Figura 6b Variação da distribuição Normal em função dos parâmetros  $\sigma$

**Definio 18** Uma variável aleatória  $Z$  normal é definida como padrão quando ela tem distribuição Normal com  $\mu = 0$  e  $\sigma^2 = 1$ . Neste caso sua função densidade é dada por:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{z^2}{2} \right\}, \quad -\infty \leq z \leq +\infty \quad (1.17)$$

O cálculo de probabilidade envolvendo uma variável aleatória normal com qualquer valor de  $\mu$  e  $\sigma$  pode ser feito através da variável Normal Padrão, usando a relação:

$$Z = \frac{X - \mu}{\sigma} \Leftrightarrow X = \mu + \sigma Z \quad (1.18)$$

Assim para calcular a probabilidade:

$$P(a < X \leq b) = P\left(\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = P(z_1 < Z \leq z_2)$$

onde:

$$z_1 = \frac{a - \mu}{\sigma} \quad z_2 = \frac{b - \mu}{\sigma}$$

Denotando por  $\Phi(z_1) = P(Z \leq z_1)$ , podemos escrever:

$$P(a < X \leq b) = \Phi(z_2) - \Phi(z_1)$$

Os valores da função  $\Phi(z)$  são encontrados em tabelas (ver anexo 1). O exemplo a seguir ilustra o uso desta tabela.

**Exemplo 19** *Suponha que a corrente elétrica do Exemplo 1 é modelada como uma variável aleatória Normal com  $\mu = 5$  e  $\sigma^2 = 1$ . Queremos calcular a probabilidade  $P(4.5A < X \leq 5.5A)$ . Usando a definição da variável normal padrão, temos:*

$$z_1 = \frac{a - \mu}{\sigma} = \frac{4.5 - 5}{1} = -0.5$$

$$z_2 = \frac{b - \mu}{\sigma} = \frac{5.5 - 5}{1} = +0.5$$

Então,

$$P(4.5 < X \leq 5.5) = \Phi(0.5) - \Phi(-0.5)$$

$$P(4.5 < X \leq 5.5) = 0.6915 - 0.3085 = 0.3829$$

*Note que se usarmos os dados da Tabela 1, verificamos que existem 5/30 valores abaixo de 4.5 e 11/30 valores acima de 5.5, portanto existem 14/30 valores entre 4.5 e 5.5. A estimativa (não paramétrica) da probabilidade procurada é portanto  $\hat{p} = 14/30 = 0.4666$ , ou seja bem maior que o valor verdadeiro se a suposição de normalidade  $N(5,1)$  for correta. A Figura 7 mostra a relação entre os percentis e os parâmetros da densidade Normal*

padrão.

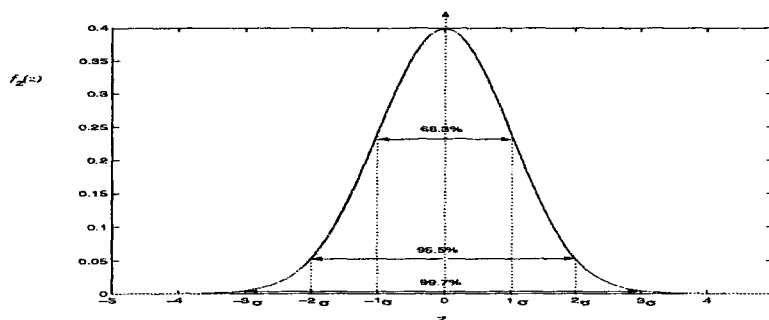


Figura 7 Percentis da densidade de probabilidade Normal padrão

Algumas funções densidade muito utilizadas são dadas na Tabela 4

Tabela 4 Funções densidade de probabilidade usuais

Nome	Densidade $f_X(x)$
<i>Binomial</i>	$\binom{n}{x} p^x (1-p)^{n-x} I_{[0,1,\dots,n]}, 0 \leq p \leq 1$
<i>Poisson</i>	$\frac{\exp(-\lambda)\lambda^x}{x!} I_{[0,1,\dots]}, \lambda > 0$
<i>Uniforme</i>	$\frac{1}{b-a} I[a, b], -\infty < a < b < \infty$
<i>Exponencial</i>	$\alpha \exp(-\alpha x) I_{(0,\infty)}, \alpha > 0$
<i>Weibull</i>	$abx^{b-1} \exp(-ax^b) I_{(0,\infty)}, a > 0, b > 0$
<i>Gama</i>	$\frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) I_{(0,\infty)}, a > 0, b > 0$
<i>Normal</i>	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \sigma > 0, -\infty < \mu < \infty$
<i>LogNormal</i>	$\frac{1}{x\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right\} I_{(0,\infty)}, \sigma > 0, -\infty < \mu < \infty$
<i>Gambel</i>	$\frac{1}{b} \exp\left\{-\left[\frac{x-a}{b} + \exp\left(-\frac{x-a}{b}\right)\right]\right\}, -\infty < a < \infty, b > 0$
<i>t-Student</i>	$\frac{\Gamma[(k+1)/2]}{\sqrt{k\pi}\Gamma(k/2)} \left[1 + \left(\frac{x-\mu}{\sigma}\right)^2\right]^{-(k+1)/2}, k > 1, -\infty < \mu < \infty, \sigma > 0$

## 1.5 Valor Esperado e Variância

Assumindo-se um modelo probabilístico para uma variável aleatória, é razoável que se expresse os valores de média e variância em função dos parâmetros da função distribuição (ou densidade) de probabilidade da variável.

**Definio 20** *Se  $X$  é uma variável aleatória contínua com função densidade de probabilidade  $f_X(x)$ . Defina-se:*

1) *O valor esperado de  $X$  como:*

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx \quad (1.19)$$

2) *A variância de  $X$  como:*

$$V(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2 \quad (1.20)$$

O valor esperado é uma medida do valor central em torno do qual a variável aleatória vai esta distribuída e a variância é uma medida de dispersão, medido o quanto a variável se distancia do valor esperado. Portanto ao coletar uma amostra de uma variável, esperamos encontrar uma média amostral  $\bar{X}$  e uma variância amostral  $S^2$ , próxima do valor esperado  $E(X)$  e da variância  $V(X)$ .

**Exemplo 21** *Se  $X$  é uma variável aleatória normalmente distribuída com parâmetros  $\mu$  e  $\sigma$ , então pode ser mostrado que:*

$$E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x \exp \left\{ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right\} dx = \mu \quad (1.21)$$

$$V(X) = E(X^2) - \mu^2 = \sigma^2 \quad (1.22)$$

Este exemplo permite interpretar o fato de que maiores valores de  $\sigma^2$  acarretam em maior abertura na curva de  $f_X(x)$ , como mostra a Figura 6b.

## 1.6 Distribuição de Probabilidade Conjunta

Os modelos matemáticos (probabilísticos) para muitos problemas de engenharia, geralmente envolvem mais de uma variável aleatória. Assim quando

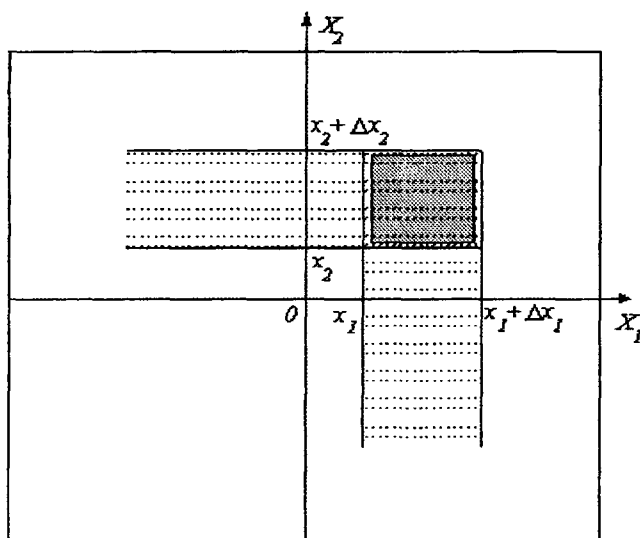
se tem duas ou mais variáveis representamos estas por um vetor aleatório  $\mathbf{X} = \{X_1, \dots, X_n\}$ . Vamos iniciar considerando modelos com duas variáveis aleatórias  $\mathbf{X} = \{X_1, X_2\}$ . Nestes problemas, geralmente é necessário calcular a **probabilidade conjunta** do evento  $\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\}$ . Denotamos essa probabilidade conjunta por  $P[X_1 \leq x_1, X_2 \leq x_2]$ . O uso das funções de distribuições individuais  $F_{X_1}(x_1)$  e  $F_{X_2}(x_2)$  para calcular essa probabilidade conjunta só pode ser feito em casos particulares de independência entre as variáveis. No caso geral, é imprescindível que se conheça a **Função Distribuição de Probabilidade Conjunta**  $F_{X_1, X_2}(x_1, x_2)$  que é definida como:

$$F_{X_1, X_2}(x_1, x_2) = P[X_1 \leq x_1, X_2 \leq x_2] \quad (1.23)$$

• **Propriedades da Distribuição de Probabilidade Conjunta**

- 1)  $F_{X_1, X_2}(x_1, \infty) = P[X_1 \leq x_1, X_2 \leq \infty] = P[X_1 \leq x_1] = F_{X_1}(x_1)$
- 2)  $F_{X_1, X_2}(\infty, x_2) = P[X_1 \leq \infty, X_2 \leq x_2] = P[X_2 \leq x_2] = F_{X_2}(x_2)$
- 3)  $F_{X_1, X_2}(\infty, \infty) = P[X_1 \leq \infty, X_2 \leq \infty] = 1$
- 4)  $P[X_1 \leq x_1 + \Delta x_1, X_2 \leq x_2] = F_{X_1, X_2}(x_1 + \Delta x_1, x_2) - F_{X_1, X_2}(x_1, x_2)$
- 5)  $P[X_1 \leq x_1, X_2 \leq x_2 + \Delta x_2] = F_{X_1, X_2}(x_1, x_2 + \Delta x_2) - F_{X_1, X_2}(x_1, x_2)$
- 6)  $P[x_1 < X_1 \leq x_1 + \Delta x_1, x_2 < X_2 \leq x_2 + \Delta x_2] =$   
 $= F_{X_1, X_2}(x_1 + \Delta x_1, x_2 + \Delta x_2) - F_{X_1, X_2}(x_1, x_2 + \Delta x_2) -$   
 $F_{X_1, X_2}(x_1 + \Delta x_1, x_2) + F_{X_1, X_2}(x_1, x_2)$

Uma ilustração gráfica das propriedades 4 a 6 é apresentada na Figura 8 que se segue:

Figura 8 - Cálculo da probabilidade conjunta de  $(X_1, X_2)$ 

**Definio 22** *Sejam  $X_1$  e  $X_2$  variáveis aleatórias contínuas com Função Distribuição de Probabilidade Conjunta  $F_{X_1, X_2}(x_1, x_2)$ . Então a função densidade de probabilidade conjunta, é definida por:*

$$f_{X_1, X_2}(x_1, x_2) = \frac{\partial^2 F_{X_1, X_2}(x_1, x_2)}{\partial x_1 \partial x_2} \quad (1.24)$$

Da equação (1.24), podemos deduzir que:

$$F_{X_1, X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1, X_2}(s_1, s_2) ds_1 ds_2 \quad (1.25)$$

• **Propriedades da função densidade conjunta**

$$1) f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_2$$

Da equação (1.25) e da propriedade (1) da Função Distribuição Conjunta,  $F_{X_1, X_2}(x_1, \infty) = F_{X_1}(x_1)$ , então:

$$F_{X_1, X_2}(x_1, \infty) = F_{X_1}(x_1) = \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f_{X_1, X_2}(s_1, s_2) ds_1 ds_2$$

Sabendo que  $F_{X_1}(x_1) = \int_{-\infty}^{x_1} f_{X_1}(x_1)dx_1$ , concluímos a demonstração:

$$2) f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2)dx_1$$

A demonstração é idêntica à da propriedade (1)

$$3) F_{X_1, X_2}(\infty, \infty) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1, X_2}(s_1, s_2)ds_1 ds_2 = 1$$

Essa propriedade é o resultado direto da propriedade (3) da função distribuição conjunta

**Exemplo 23** *Cálculo de Probabilidade usando a densidade conjunta:*

$$\begin{aligned} P(x_1 < X_1 \leq x_1 + \Delta x_1, x_2 < X_2 \leq x_2 + \Delta x_2) &= \\ &= \int_{x_1}^{x_1 + \Delta x_1} \int_{x_2}^{x_2 + \Delta x_2} f_{X_1, X_2}(x_1, x_2)dx_1 dx_2 \end{aligned}$$

Uma generalização para a função distribuição de probabilidade e função de densidade conjunta pode ser feita considerando um vetor aleatório  $n$ -dimensional  $\mathbf{X} = [X_1, \dots, X_n]$  e a relação entre as funções distribuição e densidade conjunta.

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \frac{\partial^n F_{X_1, \dots, X_n}(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n} \quad (1.26)$$

O cálculo de probabilidade neste caso pode ser feito considerando um subconjunto  $B \subset R^n$  e temos:

$$P(\mathbf{X} \in B) = \int_B f_{X_1, \dots, X_n}(x_1, \dots, x_n)dx_1 \dots dx_n \quad (1.27)$$

A integral  $\int_B$  em (1.27) é uma integral múltipla  $n$ -dimensional.

• **Variáveis aleatórias independentes e função densidade conjunta**

Se  $(X_1, X_2)$  são independentes então, temos:  $P[X_1 \leq x_1, X_2 \leq x_2] = P[X_1 \leq x_1]P[X_2 \leq x_2]$ , conseqüentemente, temos:

$$\begin{aligned} F_{X_1, X_2}(x_1, x_2) &= F_{X_1}(x_1)F_{X_2}(x_2) \\ f_{X_1, X_2}(x_1, x_2) &= f_{X_1}(x_1)f_{X_2}(x_2) \end{aligned}$$

ou seja a lei de lei de probabilidade para eventos independentes aplica-se também às funções distribuição e densidade conjunta.

**Exemplo 24** Um vetor aleatório  $\mathbf{X} = (X_1, X_2)'$  Normal bidimensional, tem distribuição conjunta dada por:

$$f_{X_1, X_2}(x_1, x_2) = C \exp \left\{ -\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \right\} \quad (1.28)$$

onde  $\boldsymbol{\mu} = (\mu_1, \mu_2)'$  é um vetor das médias com  $E(X_1) = \mu_1$  e  $E(X_2) = \mu_2$ . A constante  $C$  e a matriz  $\boldsymbol{\Sigma}_{2 \times 2}$  são dadas por:

$$C = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \text{ e } \boldsymbol{\Sigma} = \begin{matrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{matrix}$$

onde  $V(X_1) = \sigma_1^2$ ,  $V(X_2) = \sigma_2^2$  e  $\rho$  é um número em  $(-1, 1)$  dado pela relação  $\rho\sigma_1\sigma_2 = E(X_1X_2) - \mu_1\mu_2$ , que será interpretado mais adiante neste curso. Pode ser mostrado, com algum esforço de cálculo, as propriedades (1) e (2) da função densidade conjunta, ou seja:

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} C \exp \left\{ -\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \right\} dx_2 \\ f_{X_1}(x_1) &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp \left\{ -\frac{1}{2} \left( \frac{x - \mu_1}{\sigma_1} \right)^2 \right\} \\ f_{X_2}(x_2) &= \int_{-\infty}^{\infty} C \exp \left\{ -\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \right\} dx_1 \\ f_{X_2}(x_2) &= \frac{1}{\sqrt{2\pi}\sigma_2} \exp \left\{ -\frac{1}{2} \left( \frac{x - \mu_2}{\sigma_2} \right)^2 \right\} \end{aligned}$$

**Exemplo 25** No Exercício anterior, se  $\rho = 0$ , então  $X_1, X_2$  se comportam como variáveis aleatórias normais independentes, ou seja:

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= \frac{1}{\sqrt{2\pi}\sigma_1\sigma_2} \exp \left\{ -\frac{1}{2} \left[ \left( \frac{x - \mu_1}{\sigma_1} \right)^2 + \left( \frac{x - \mu_2}{\sigma_2} \right)^2 \right] \right\} \\ f_{X_1, X_2}(x_1, x_2) &= f_{X_1}(x_1) f_{X_2}(x_2) \end{aligned}$$

Esse exercício mostra que o parâmetro  $\rho$  está associado à dependência entre as variáveis normais. e esse parâmetro é denominado de coeficiente de correlação.



## 1.7 Soma e Produto de Variáveis Aleatórias

Operações matemáticas envolvendo variáveis aleatórias são fundamentais para análise de alguns problemas reais. Para isso vamos apresentar aqui algumas propriedades da soma e do produto de variáveis aleatórias.

### • 1) Valor Esperado da Soma de Duas Variáveis Aleatórias

Se  $(X_1, X_2)$  é um vetor aleatório contínuo com densidade conjunta  $f_{X_1, X_2}(x_1, x_2)$ , então  $E(X_1 + X_2) = E(X_1) + E(X_2)$

$$\begin{aligned}
 E(X_1 + X_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 + x_2) f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \\
 E(X_1 + X_2) &= \int_{-\infty}^{\infty} x_1 \left( \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_2 \right) dx_1 + \\
 &\quad + \int_{-\infty}^{\infty} x_2 \left( \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_1 \right) dx_2 \\
 E(X_1 + X_2) &= \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 + \int_{-\infty}^{\infty} x_2 f_{X_2}(x_2) dx_2 \\
 E(X_1 + X_2) &= E(X_1) + E(X_2)
 \end{aligned}$$

### • 2) Valor Esperado da Combinação Linear de Duas Variáveis Aleatórias

Se  $(X_1, X_2)$  é um vetor aleatório contínuo com densidade conjunta  $f_{X_1, X_2}(x_1, x_2)$ , e  $a_1, a_2$  são constantes, então

$$E(a_1 X_1 + a_2 X_2) = a_1 E(X_1) + a_2 E(X_2)$$

A generalização dessa propriedade para  $n$ -variáveis aleatórias  $X_1, \dots, X_n$  com densidade conjunta  $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ , leva ao seguinte resultado:

$$E \left( \sum_{i=1}^n a_i X_i \right) = \sum_{i=1}^n a_i E(X_i)$$

### • 3) Valor Esperado do Produto de Duas Variáveis Aleatórias (Independentes)

Se  $(X_1, X_2)$  é um vetor aleatório contínuo com densidade conjunta  $f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$ , então  $E(X_1 X_2) = E(X_1)E(X_2)$

$$E(X_1 X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 x_2) f_{X_1, X_2}(x_1, x_2) dx_1 dx_2$$

$$E(X_1 X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 x_2) f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2$$

$$E(X_1 X_2) = \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) \left( \int_{-\infty}^{\infty} f_{X_2}(x_2) dx_2 \right) dx_1 \times \\ \times \int_{-\infty}^{\infty} x_2 f_{X_2}(x_2) \left( \int_{-\infty}^{\infty} f_{X_1}(x_1) dx_1 \right) dx_2$$

$$E(X_1 X_2) = \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 \times \int_{-\infty}^{\infty} x_2 f_{X_2}(x_2) dx_2$$

$$E(X_1 X_2) = E(X_1)E(X_2)$$

• **4) Variância da Soma de Duas Variáveis Aleatórias (Independentes)**

Se  $(X_1, X_2)$  é um vetor aleatório contínuo com densidade conjunta  $f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$ , então  $V(X_1 + X_2) = V(X_1) + V(X_2)$

$$V(X_1 + X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 + x_2)^2 f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 - (\mu_1 + \mu_2)^2$$

$$V(X_1 + X_2) = \int_{-\infty}^{\infty} x_1^2 \left( \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_2 \right) dx_1 + \\ + 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2 \\ + \int_{-\infty}^{\infty} x_2^2 \left( \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_1 \right) dx_2 - (\mu_1 + \mu_2)^2$$

$$V(X_1 + X_2) = \int_{-\infty}^{\infty} x_1^2 f_{X_1}(x_1) dx_1 + \\ 2 \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 \int_{-\infty}^{\infty} x_2 f_{X_2}(x_2) dx_2 + \\ + \int_{-\infty}^{\infty} x_2^2 f_{X_2}(x_2) dx_2 - (\mu_1 + \mu_2)^2$$

$$\begin{aligned} V(X_1 + X_2) &= E(X_1^2) + 2\mu_1\mu_2 + E(X_2^2) - (\mu_1 + \mu_2)^2 \\ V(X_1 + X_2) &= E(X_1^2) - \mu_1^2 + E(X_2^2) - \mu_2^2 = V(X_1) + V(X_2) \end{aligned}$$

### 5) Variância do Produto de Uma Constante Por Uma Variável Aleatória

Se  $X_1$  é uma variável aleatória contínua com densidade  $f_{X_1}(x_1)$  e  $a_1$  é uma constante, então  $V(a_1X_1) = a_1^2V(X_1)$

$$\begin{aligned} V(a_1X_1) &= \int_{-\infty}^{\infty} (a_1x_1)^2 f_{X_1}(x_1) dx_1 - (a_1\mu_1)^2 \\ V(a_1X_1) &= a_1^2 \left( \int_{-\infty}^{\infty} x_1^2 f_{X_1}(x_1) dx_1 - \mu_1^2 \right) = a_1^2 V(X_1) \end{aligned}$$

Combinado-se as propriedades (4) e (5) temos a generalização dessa propriedade para  $n$ -variáveis aleatórias  $X_1, \dots, X_n$  independentes:

$$V \left( \sum_{i=1}^n a_i X_i \right) = \sum_{i=1}^n a_i^2 V(X_i)$$

## 1.8 Covariância e Correlação

A covariância é uma medida do grau de dependência linear entre duas variáveis aleatórias, esta medida é calculada como:

$$Cov(X_1, X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)] \quad (1.29)$$

Na equação (1.29) estamos usando a notação que  $E(X_1) = \mu_1$  e  $E(X_2) = \mu_2$ , desenvolvendo o produto em (1.29), podemos mostrar que:

$$Cov(X_1, X_2) = E(X_1 X_2) - \mu_1 \mu_2 \quad (1.30)$$

Para interpretar a covariância entre duas variáveis vamos supor que  $\mu_1 = \mu_2 = 0$ . Neste caso temos  $Cov(X_1, X_2) = E(X_1 X_2)$  e podemos dizer que se  $X_1 X_2 > 0 \Rightarrow X_1$  e  $X_2$  com mesmo sinal, então  $Cov(X_1, X_2) > 0$ , caso contrário temos  $Cov(X_1, X_2) < 0$ . A Figura 9 apresenta uma interpretação gráfica do sinal da covariância.

Um estimador para a covariância, baseado em uma amostra de um vetor aleatório é dado por:

$$C_{X_1 X_2} = \frac{1}{n-1} \sum_{i=1}^n (X_{i,1} - \bar{X}_1)(X_{i,2} - \bar{X}_2) \quad (1.31)$$

$$C_{X_1 X_2} = \frac{1}{n-1} \sum_{i=1}^n X_{i,1} X_{i,2} - \frac{n}{(n-1)} \bar{X}_1 \bar{X}_2 \quad (1.32)$$

A dificuldade na interpretação da covariância é o fato desta depender da ordem de grandeza das variáveis envolvidas, por isso é mais prático calcular uma medida normalizada entre  $-1$  e  $1$ . Um coeficiente que preserva as mesmas propriedades do sinal da  $Cov(X_1, X_2)$  e só assume valores em  $[-1, 1]$  é denominado coeficiente de correlação linear.

**Definio 26** Se  $X_1$  e  $X_2$  são duas variáveis aleatórias com variâncias  $\sigma_1^2$  e  $\sigma_2^2$  respectivamente, então define-se o coeficiente de correlação linear entre  $X_1$  e  $X_2$  por:

$$\rho = \frac{Cov(X_1, X_2)}{\sigma_1 \sigma_2} = \frac{E(X_1 X_2) - \mu_1 \mu_2}{\sigma_1 \sigma_2} \quad (1.33)$$

- **Propriedades do Coeficiente de Correlação Linear**

1)  $|\rho| \leq 1$

2) Se  $X_1$  e  $X_2$  são independentes então  $\rho = 0$

3) Se  $X_1 = X_2 \Rightarrow \rho = 1$

Se  $X_1$  e  $X_2$  são independentes  $\Rightarrow E(X_1 X_2) = E(X_1)E(X_2) = \mu_1 \mu_2$ ,  
então  $E(X_1 X_2) - \mu_1 \mu_2 = 0$

4) Se  $(X_1, X_2)$  é um vetor aleatório normal bivariado e  $\rho = 0$ , então  $X_1$  e  $X_2$  são independentes.

- **Estimador de  $\rho$**

Um estimador para o coeficiente de correlação, baseado em uma amostra de um vetor aleatório é dado por:

$$\hat{\rho} = \frac{C_{X_1 X_2}}{S_1 S_2} = \frac{\frac{1}{n-1} \sum_{i=1}^n X_{i,1} X_{i,2} - \frac{n}{(n-1)} \bar{X}_1 \bar{X}_2}{S_1 S_2} \quad (1.34)$$

onde  $\bar{X}_1$  e  $\bar{X}_2$  são as médias amostrais calculadas por (1.1),  $S_1$  e  $S_2$  são os desvios-padrões amostrais calculados por (1.3) A Figura 9 apresenta uma interpretação gráfica do sinal da covariância e das propriedades do

coeficiente de correlação.

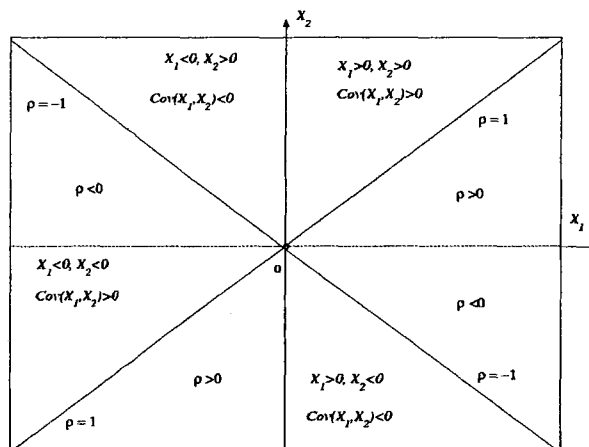


Figura 9 Interpretação do sinal da covariância

## 1.9 Densidade de Probabilidade Condicional

Quando o problema a ser analisado apresenta duas ou mais variáveis aleatórias é importante saber como estas variáveis estão relacionadas. Por exemplo, supondo que uma das variáveis assume um certo valor, como isto afeta a probabilidade de ocorrência das outras variáveis? Esta é uma das questões que só pode ser respondida conhecendo-se a **Função Distribuição de Probabilidade Condicionada**. Para definir essa função distribuição condicional vamos considerar os eventos  $\{X_1 \leq x_1\}$  e  $\{X_2 \in E\}$ . Queremos calcular a probabilidade condicional:

$$P(X_1 \leq x_1 | X_2 \in E) = \frac{P(X_1 \leq x_1, X_2 \in E)}{P(X_2 \in E)} \tag{1.35}$$

Note que  $P(X_1 \leq x_1 | X_2 \in E)$  é uma lei de probabilidade para a variável  $X_1$  pois a segunda variável  $X_2$  é assumida conhecida. Considerando os eventos:

$$A = \{x_1 < X_1 \leq x_1 + \Delta x_1\}$$

$$B = \{x_2 < X_2 \leq x_2 + \Delta x_2\}$$

podemos escrever:

$$P(X_1 \in A | X_2 \in B) = \frac{P(x_1 < X_1 \leq x_1 + \Delta x_1, x_2 < X_2 \leq x_2 + \Delta x_2)}{P(x_2 < X_2 \leq x_2 + \Delta x_2)}$$

Com a propriedade (6) da função distribuição de probabilidade conjunta, podemos escrever que:

$$\frac{\partial^2 F_{X_1, X_2}(x_1, x_2)}{\partial x_1 \partial x_2} = \lim_{\Delta x_1, \Delta x_2 \rightarrow 0} \frac{P[x_1 < X_1 \leq x_1 + \Delta x_1, x_2 < X_2 \leq x_2 + \Delta x_2]}{\Delta x_1 \Delta x_2}$$

Usando a relação (1.24), podemos escrever que:

$$P[x_1 < X_1 \leq x_1 + \Delta x_1, x_2 < X_2 \leq x_2 + \Delta x_2] \approx f_{X_1, X_2}(x_1, x_2) \Delta x_1 \Delta x_2 \quad (1.36)$$

Desenvolvendo a mesma relação para o caso univariado, temos:

$$\frac{dF_{X_2}(x_2)}{dx_2} = \lim_{\Delta x_2 \rightarrow 0} \frac{P[x_2 < X_2 \leq x_2 + \Delta x_2]}{\Delta x_2}$$

Assim temos:

$$P[x_2 < X_2 \leq x_2 + \Delta x_2] \approx f_{X_2}(x_2) \Delta x_2 \quad (1.37)$$

As equações (1.36)-(1.37) permitem deduzir que:

$$P[x_1 < X_1 \leq x_1 + \Delta x_1 | x_2 < X_2 \leq x_2 + \Delta x_2] = \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)} \Delta x_1 \quad (1.38)$$

definindo a função densidade de probabilidade conjunta por:

$$f_{X_1 | X_2}(x_1 | x_2) = \lim_{\Delta x_2 \rightarrow 0} \frac{P[x_1 < X_1 \leq x_1 + \Delta x_1 | x_2 < X_2 \leq x_2 + \Delta x_2]}{\Delta x_1} \quad (1.39)$$

Com as equações (1.38) e (1.39), podemos definir formalmente a função densidade de probabilidade conjunta como:

**Definio 27** Se  $X_1$  e  $X_2$  são duas variáveis aleatórias contínuas com função densidade de probabilidade conjunta  $f_{X_1, X_2}(x_1, x_2)$  e  $X_2$  tem densidade marginal  $f_{X_2}(x_2)$ , então a função densidade de probabilidade de  $X_1$  condicionada a  $X_2$ , é definida como:

$$f_{X_1 | X_2}(x_1 | x_2) = \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_1, x_2)} \quad (1.40)$$

**Exemplo 28** Um vetor aleatório  $\mathbf{X} = (X_1, X_2)'$  Normal bidimensional, tem distribuição conjunta dada por (1.28):

$$f_{X_1, X_2}(x_1, x_2) = C \exp \left\{ -\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \right\} \quad (1.41)$$

A densidade condicional  $f_{X_1|X_2}(x_1|x_2)$  pode ser calculada usando (1.40), e chega-se a seguinte expressão:

$$f_{X_1|X_2}(x_1|x_2) = D \exp \left\{ -\frac{1}{2\sigma_1^2(1-\rho)} \left[ (x_1 - \mu_1 - \frac{\rho\sigma_1}{\sigma_2}(x_2 - \mu_2)) \right]^2 \right\} \quad (1.42)$$

com a constante  $D = 1/\sqrt{2\pi}\sigma_1(1-\rho^2)$ . Notamos portanto que denotando por  $\mu_{1|2} = \mu_1 - (\rho\sigma_1^2/\sigma_2^2)(x_2 - \mu_2)$  e por  $\sigma_{1|2}^2 = \sigma_1^2(1-\rho^2)$  podemos escrever  $f_{X_1|X_2}(x_1|x_2)$  como uma densidade normal  $N(\mu_{1|2}, \sigma_{1|2}^2)$ .

## 1.10 Valor Esperado e Variância Condicional

Quando o cálculo do valor esperado de uma variável aleatória é feito com a função densidade condicional em lugar da função densidade marginal, esse valor esperado é denominado valor esperado condicional e o mesmo se aplica para a variância. O valor esperado e a variância condicional são definidos como:

**Definio 29** Se  $\mathbf{X} = (X_1, X_2)$  é um vetor aleatório contínuo com função densidade de probabilidade condicional  $f_{X_1|X_2}(x_1|x_2)$ . Defina-se:

1) O valor esperado condicional de  $X_1$  dado  $X_2$  como

$$E(X_1|X_2) = \int_{-\infty}^{\infty} x_1 f_{X_1|X_2}(x_1|x_2) dx_1 \quad (1.43)$$

2) A variância condicional de  $X_1$  dado  $X_2$  é:

$$V(X_1|X_2) = E(X_1^2|X_2) - [E(X_1|X_2)]^2 \quad (1.44)$$

**Exemplo 30** Considerando o exemplo anterior, temos:

$$E(X_1|X_2 = x_2) = \mu_1 + (\rho\sigma_1/\sigma_2)(x_2 - \mu_2)$$

$$V(X_1|X_2 = x_2) = \sigma_1^2(1 - \rho^2)$$

**Exemplo 31** Considere os dados do consumo médio mensal de energia elétrica  $X$  (KW-hora) de uma residência de classe média de uma cidade tropical e os dados da temperatura média mensal desta cidade  $T$  ( $^{\circ}\text{C}$ ) para os últimos cinco anos, dados na Tabela 5. Queremos analisar estatisticamente essas duas variáveis aleatórias e verificar se há uma relação entre elas.

Tabela 5 Consumo de energia elétrica  $X$  e Temperatura  $T$ 

$t$	$X$	$T$	$t$	$X$	$T$	$t$	$X$	$T$	$t$	$X$	$T$	$t$	$X$	$T$
1	105	19	13	101	27	25	98	27	37	100	22	49	101	25
2	105	18	14	96	26	26	101	30	38	104	28	50	100	30
3	101	24	15	100	28	27	101	31	39	105	36	51	101	31
4	102	21	16	104	29	28	104	31	40	100	27	52	101	22
5	99	27	17	98	27	29	101	26	41	103	29	53	102	29
6	105	33	18	105	27	30	99	25	42	100	26	54	99	24
7	103	29	19	100	27	31	101	26	43	99	25	55	99	22
8	99	23	20	101	29	32	105	20	44	101	22	56	99	22
9	99	24	21	100	28	33	106	20	45	99	23	57	98	24
10	100	23	22	100	24	34	100	26	46	103	30	58	106	35
11	100	28	23	101	23	35	98	22	47	104	30	59	97	23
12	99	26	24	102	23	36	102	27	48	102	28	60	104	30

A análise dos dados tabelados começa com um gráfico de  $X \times T$ , dado na Figura 10. Este gráfico mostra uma relação não linear entre as variáveis, de forma que a estimativa do coeficiente de correlação para esses dados é baixo  $\hat{\rho} = 0.25$ .

	$X$ (Energia)	$T$ (Temperatura)
média	$\bar{X} = 101.13$	$\bar{T} = 26.12$
dp	$S_X = 2.3826$	$S_T = 3.7736$

$$\hat{\rho} = \frac{\frac{1}{n-1} \sum_{i=1}^n X_{i,1} X_{i,2} - \frac{n}{(n-1)} \bar{X}_1 \bar{X}_2}{S_1 S_2}$$

$$\hat{\rho} = \frac{2643.48 - (60/59)101.13 \times 26.12}{2.3826 \times 3.7736} = 0.25$$



No entanto uma análise mais cuidadosa do gráfico da Figura 10 identifica duas relações lineares: uma para  $T \leq 25^\circ\text{C}$  e outra para  $T > 25^\circ\text{C}$ .

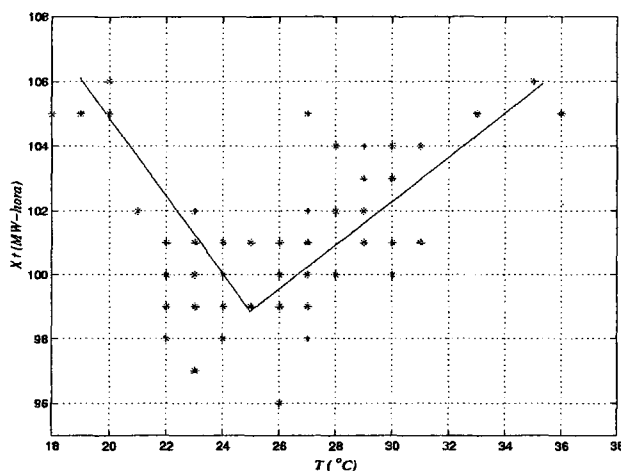


Figura 10 Relação consumo de energia  $\times$  temperatura

Separando os dados da Tabela 5 em dois conjuntos, um com os valores de  $T < 25^\circ\text{C}$ , e o segundo com o restante, temos dois conjuntos de dados. A Tabela 6 apresenta o conjunto de dados para  $T < 25^\circ\text{C}$ .

Tabela 6 Consumo de energia para  $T < 25^\circ\text{C}$

$t$	$X$	$T$	$t$	$X$	$T$
1	105	19	33	106	20
2	105	18	35	98	22
3	101	24	37	100	22
4	102	21	44	101	22
8	99	23	45	99	23
9	99	24	52	101	22
10	100	23	54	99	24
22	100	24	55	99	22
23	101	23	56	99	22
24	102	23	57	98	24
32	105	20	59	97	23

A estimativa do coeficiente de correlação para esses dados é  $\hat{\rho} = -0.78$ , portanto bem maior que a estimativa calculada com todos os dados.  $\circ$

34CAPÍTULO 1. CONCEITOS DE PROBABILIDADE E ESTATÍSTICA

gráfico da Figura 11 mostra a relação linear entre essas duas variáveis.

	$X$ (Energia)	$T$ (Temperatura)
média	$\bar{X} = 100.72$	$\bar{T} = 22.18$
dp	$S_X = 2.5295$	$S_T = 1.6800$

$$\hat{\rho} = \frac{\frac{1}{n-1} \sum_{i=1}^n X_{i,1} X_{i,2} - \frac{n}{(n-1)} \bar{X}_1 \bar{X}_2}{S_1 S_2}$$

$$\hat{\rho} = \frac{2231.13 - (22/21)100.72 \times 22.18}{2.5295 \times 1.6800} = -0.78$$

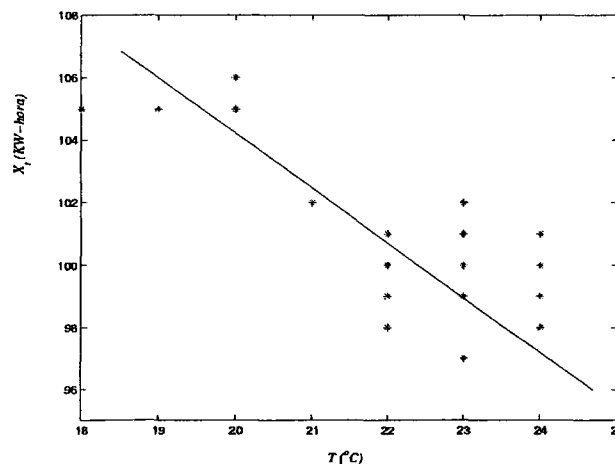


Figura 11 Relação entre Energia  $\times$  Temperatura para ( $T < 25 \text{ }^\circ\text{C}$ )

A mesma análise pode ser repetida para  $T \geq 25 \text{ }^\circ\text{C}$ , os resultados são apresentados a seguir e a Figura 12 mostra que nesta faixa de temperatura a relação linear tem inclinação positiva com  $\rho = 0.68$ .

	$X$ (Energia)	$T$ (Temperatura)
média	$\bar{X} = 101.37$	$\bar{T} = 28.39$
dp	$S_X = 2.2947$	$S_T = 2.5631$

$$\hat{\rho} = \frac{\frac{1}{n-1} \sum_{i=1}^n X_{i,1} X_{i,2} - \frac{n}{(n-1)} \bar{X}_1 \bar{X}_2}{S_1 S_2}$$

$$\hat{\rho} = \frac{2882.21 - (38/37)101.37 \times 28.39}{2.2947 \times 2.5631} = 0.68$$

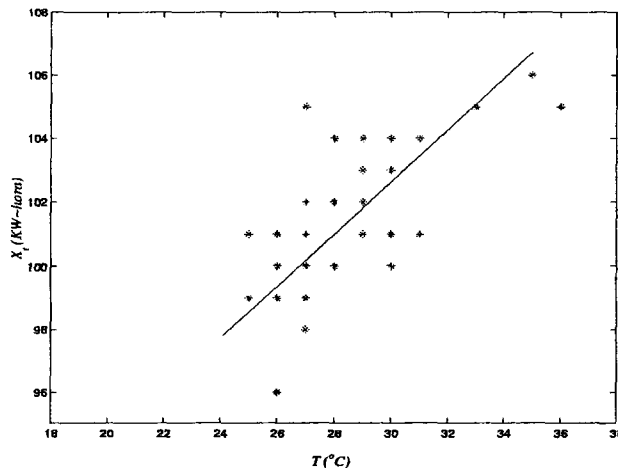


Figura 12 Relação entre Energia  $\times$  Temperatura para ( $T < 25^\circ\text{C}$ )

Considerando o consumo de energia elétrica  $X$  e a temperatura  $T$  como variáveis aleatórias normais, cujas médias e variâncias estão estimadas no exemplo anterior, podemos estabelecer a relação linear para cada uma das situações ( $T < 25^\circ\text{C}$ ) e ( $T \geq 25^\circ\text{C}$ ), usando a equação (1.43). Assim temos:

$$E(X_1|X_2 = x_2) = \mu_1 + (\rho\sigma_1/\sigma_2)(x_2 - \mu_2)$$

$$E(X_1|X_2 = x_2) = \begin{cases} 100.72 - 0.78 \times 2.5295/1.6800(T - 22.18), & T < 25^\circ\text{C} \\ 101.37 + 0.68 \times 2.2974/2.5631(T - 28.39), & T \geq 25^\circ\text{C} \end{cases}$$

Essa relação permite calcular o valor esperado do consumo de energia elétrica para uma previsão da temperatura. Relações mais complicadas entre variáveis aleatórias é o campo de estudo dos modelos de regressão. O exercício a seguir pode dar uma rápida idéia de como os modelos de regressão entre variáveis aleatórias podem ser ampliados.

**Exercício 32** Repetir o Exemplo anterior considerando como segunda variável agora o  $\log(T)$  em lugar de  $T$ . Após encontrar a relação linear entre  $X$  e  $T$ , calcule o resíduo (erro de ajuste) dado por  $\epsilon = X - T$  e faça uma análise descritiva deste erro.

## 1.11 Técnica de Regressão

Supondo que temos as medidas de duas variáveis ( $X, Y$ ) tomadas em pares  $\{(x_1, y_1), \dots, (x_n, y_n)\}$ . Sabemos que há uma relação entre estas variáveis,

mas devido às imprecisões e aleatoriedades não podemos determinar com exatidão essa relação, por exemplo: A relação mais simples é a reta, neste caso temos:

$$Y = \alpha X + \beta + \epsilon \quad (1.45)$$

Onde  $\epsilon$  é um erro de medida que introduz aleatoriedade no modelo. Podemos calcular as constantes  $\alpha$  e  $\beta$  tal que o erro quadrático médio entre os pontos  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  observados e a equação (1.45) seja mínimo. Assim precisamos encontrar  $\alpha, \beta$ , solução do problem:

$$\min_{\alpha, \beta} S(\alpha, \beta) = \frac{1}{n-1} \sum_{i=1}^n (y_i - \alpha x_i + \beta)^2 \quad (1.46)$$

Para encontrar  $\alpha, \beta$  que minimizam  $S(\alpha, \beta)$ , diferenciamos (1.46) em relação a cada um dos parâmetros e igualamos os resultados a zero,

$$\frac{\partial S(\alpha, \beta)}{\partial \alpha} = \frac{-2}{n-1} \sum_{i=1}^n (y_i - \alpha x_i + \beta) x_i = 0 \quad (1.47)$$

$$\frac{\partial S(\alpha, \beta)}{\partial \beta} = \frac{-2}{n-1} \sum_{i=1}^n (y_i - \alpha x_i + \beta) = 0 \quad (1.48)$$

Assim temos o seguinte sistema de equações lineares para resolver:

$$\left( \sum_{i=1}^n x_i^2 \right) \alpha + \left( \sum_{i=1}^n x_i \right) \beta = \left( \sum_{i=1}^n x_i y_i \right) \quad (1.49)$$

$$\left( \sum_{i=1}^n x_i \right) \alpha + n\beta = \left( \sum_{i=1}^n y_i \right) \quad (1.50)$$

A solução deste sistema leva aos valores de  $\hat{\alpha}$  e  $\hat{\beta}$  que minimizam o erro médio quadrático, ou seja:

$$\hat{\alpha} = \frac{\frac{1}{n-1} \sum_{i=1}^n x_i y_i - \frac{n}{(n-1)} \bar{X} \bar{Y}}{S_X^2} = \frac{C_{X_1 X_2}}{S_X^2} \quad (1.51)$$

$$\hat{\beta} = \bar{Y} - \hat{\alpha} \bar{X} \quad (1.52)$$

Onde  $\bar{X}$  e  $\bar{Y}$  são as médias amostrais de  $X$  e  $Y$  respectivamente,  $S_X^2$  é a variância amostral de  $X$  e  $C_{X_1 X_2}$  é a covariância amostral entre  $X$  e  $Y$ .

Comparando este resultado com a expressão do valor esperado condicional  $E(Y|X = x) = \mu_Y + (\rho\sigma_Y/\sigma_{Xy})(x - \mu_X)$  e sendo  $\hat{\rho} = C_{XY}/S_X S_Y$ , temos que  $\hat{\alpha} = \hat{\rho}S_Y/S_X$  e a equação da reta que minimiza o erro médio quadrático pode ser escrita como:

$$y_i = \bar{Y} + (\hat{\rho}S_Y/S_X)(x_i - \bar{X}) \quad (1.53)$$

Uma medida da qualidade do ajuste é dada pelo coeficiente de determinação, estimado como:

$$r^2 = \frac{\text{variação explicada pela reta}}{\text{variação total em } Y}$$

$$r^2 = \frac{\sum(Y_{\text{ajustado}} - \bar{Y})^2}{\sum(Y - \bar{Y})^2} \quad (1.54)$$

o coeficiente  $r^2$  pode ser interpretado como a fração da variação total que é explicada pela reta de regressão de mínimos quadrados, se toda variação é explicada pela reta, teremos  $r^2 = 1$ .

**Exerccio 33** *Verifique a relação linear entre os meses do ano e o consumo de energia elétrica (em MW-hora) no setor residencial de uma região urbana do Estado de São Paulo, e encontre o ajuste de mínimos quadrados para essa relação:*

Mês\Ano	75	76	77	78	79
jan	63	77	81	95	106
fev	62	66	76	92	100
mar	75	72	82	102	110
abr	59	74	81	97	109
mai	69	78	88	98	114
jun	65	75	82	99	104
jul	63	76	89	96	114
ago	68	78	89	99	106
set	69	79	90	101	113
out	74	80	92	105	114
nov	73	84	91	101	117
dez	68	83	103	107	125

Vamos denotar por  $Y$  o consumo de energia e por  $X$  o tempo que varia de 1, 2, ... 60 meses. A relação linear é verificada pelo coeficiente de correlação

linear estimado por:

$$\hat{\rho} = \frac{\frac{1}{n-1} \sum_{i=1}^n y_i x_i - \frac{n}{(n-1)} \overline{XY}}{S_X S_Y} = 0.97$$

A reta de mínimos quadrados ajustada é mostrada na Figura 13, juntamente com as observações utilizadas no ajuste:

$$Y = 60.1 + 0.925X$$

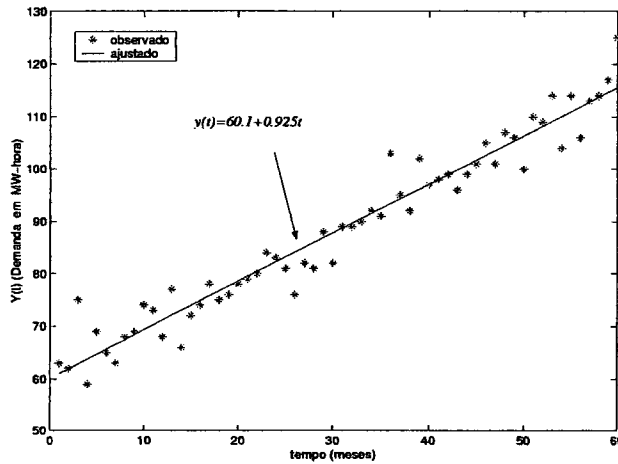


Figura 13 Ajuste da tendência da demanda de energia elétrica

O coeficiente de determinação desta reta é  $r^2 = 0.557$ , ou seja a reta de mínimos quadrados explica menos de 60% da variação total do consumo de energia. A análise do erro (o resíduo da regressão) permite avaliar a incerteza a qual estaremos sujeitos ao estimar o consumo para um mês futuro. O cálculo do resíduo dado por  $\epsilon_t = y_t - (\alpha x_t + \beta)$  para  $t = 1, \dots, 60$ , fornece um conjunto de medidas do erro cuja distribuição é dada na Figura 14 e a estatística descritiva é dada na Tabela 7.

Tabela 7 Estatística Descritiva dos Resíduos  $\epsilon_t$

Média	Mediana	Sd	I.C. 95%
0	0.0238	3.923	[-7.0397 ; 9.6132]

O histograma dos resíduos dados na Figura 14 mostra uma distribuição simétrica em torno da média indicando uma boa aproximação normal para  $\epsilon_t$ , exceto por uma assimetria para a direita, mas com um coeficiente de

assimetria  $CK = -0.0067$  considerado pequeno.

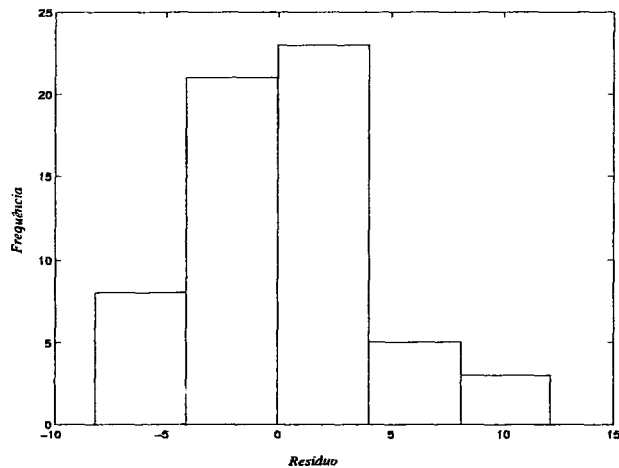


Figura 14 Histograma dos Resíduos  $\epsilon_t$

Uma forma de verificação gráfica da normalidade dos resíduos muito usada em modelos de regressão é o gráfico probabilístico normal, onde a escala do eixo vertical é calculada de forma que a distribuição de probabilidade da variável normal seja transformada em uma reta, como mostra a Figura 15.

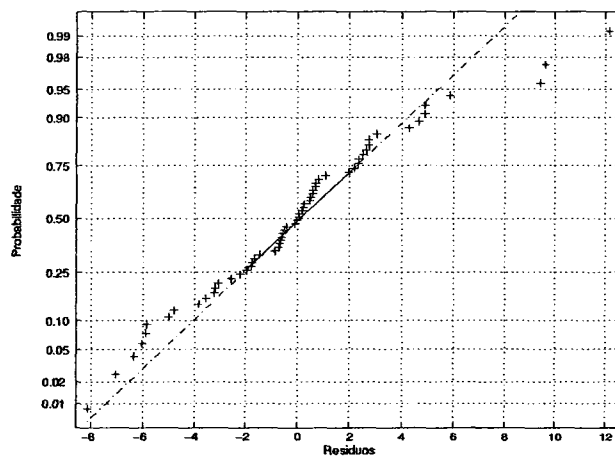


Figura 15 Gráfico Probabilístico dos Resíduos  $\epsilon_t$

O uso da equação ajustada para fazer previsão dos próximos 12 meses leva ao seguinte resultado:

$$Y_{t+k} = 60.1 + 0.925X_{t+k} \text{ para } t = 60 \text{ e } k = 1, 2, \dots, 12$$

Mês	Previsão	Observado	Erro (%)
jan	116	124	-6.04
fev	117	116	1.23
mar	118	119	-0.54
abr	119	114	4.63
mai	120	127	-5.35
jun	121	119	1.79
jul	122	121	0.87
ago	123	143	-13.99
set	124	112	10.63
out	125	123	1.49
nov	126	132	-4.73
dez	127	133	-4.75

O gráfico da Figura 16 mostra os valores previstos e observados juntamente com os dados históricos usados no ajuste do modelo de regressão. Note que o uso do modelo de regressão só permite prever a tendência (ou média) da demanda para os meses futuros.

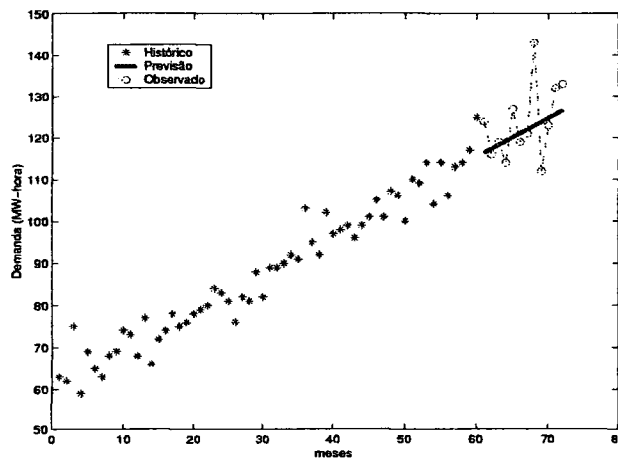


Figura 16 Valores previstos e observados para 12 meses

Dois medidas de avaliação da qualidade da previsão muito usadas são: O *Erro Quadrático Médio (EQM)* e *Erro Absoluto Percentual Médio (EAPM)*,



que podem ser calculados como:

$$REM Q = \frac{1}{N} \sum_{k=1}^N [Y_{obs}(t+k) - Y_{prev}(t+k)]^2 \quad (1.55)$$

$$EMAP(\%) = \frac{100}{N} \sum_{k=1}^N \left| \frac{Y_{obs}(t+k) - Y_{prev}(t+k)}{Y_{obs}(t+k)} \right| \quad (1.56)$$

Para esse exemplo, temos  $REM Q = 7.97$  e  $EMAP = 4.67\%$ . O  $REM Q$  é avaliado por comparação com o desvio-padrão do resíduo  $\epsilon_t$ , note neste caso que  $REM Q > 2Sd$  (dado na Tabela 7), portanto temos um erro grande indicando a imprecisão do modelo. O  $EMAP$  é avaliado de forma subjetiva e de acordo com a experiência do analista, neste caso temos  $EMAP < 5\%$  que em muitos problemas reais é considerado um erro pequeno. Quando o modelo linear não é satisfatório, podemos optar por modelos não lineares que são mais complexos, ou modelos de regressão múltipla.

#### • Parábola de Mínimos Quadrados

As idéias acima podem ser estendidas para ajuste de uma parábola dada por:

$$Y = a + bX + cX^2 \quad (1.57)$$

O ajuste da parábola de mínimos quadrados é obtido calculando-se os coeficientes  $a$ ,  $b$  e  $c$  que são solução do sistema de equações lineares:

$$na + b \sum x + c \sum x^2 = \sum y \quad (1.58)$$

$$a \sum x + b \sum x^2 + c \sum x^3 = \sum xy \quad (1.59)$$

$$a \sum x^2 + b \sum x^3 + c \sum x^4 = \sum x^2y \quad (1.60)$$

Esses modelos são geralmente usados para descrever tendência de crescimento não linear em séries de consumo de energia elétrica, onde  $Y$  representa o consumo de energia e  $X$  representa a escala de tempo ( mês, por exemplo).

#### • Regressão Múltipla

Quando é necessário considerar mais de uma variável para explicar o comportamento de uma variável de interesse, é necessário uma regressão

múltipla. No caso de duas variáveis explicativas estamos considerando o modelo do tipo:

$$Z = a + bX + cY \quad (1.61)$$

Para obter o ajuste de mínimos quadrados precisamos calcular os parâmetros  $a$ ,  $b$  e  $c$  resolvendo o seguinte sistema linear:

$$na + b \sum x + c \sum y = \sum z \quad (1.62)$$

$$a \sum x + b \sum x^2 + c \sum xy = \sum xz \quad (1.63)$$

$$a \sum y + b \sum xy + c \sum y^2 = \sum yz \quad (1.64)$$

Um exemplo onde esses modelos são aplicados são os modelos onde o consumo de energia de uma região  $Z$  está relacionado além da renda per capita  $X$ , também depende do preço de uma fonte de energia alternativa como o preço do petróleo,  $Y$ .

As generalizações para várias variáveis, assim como para modelos não lineares mais complexos são muito usadas em modelos de regressão como mostram os exemplos a seguir:

**Exemplo 34** Neste exemplo apresentamos uma classe de modelos que utilizam técnicas regressão múltipla visando avaliar o efeito de variáveis exógenas no consumo de energia elétrica. Nesta classe de modelos se assume que os bens de consumo que usam energia elétrica de uma sociedade  $S_t$  são medidos em quantidades de Watts que esses bens potencialmente podem consumir. Denotando por  $E_t$  a quantidade de energia elétrica, medida em KW-hora, consumida por essa sociedade, o modelo supõe que a relação entre essas duas variáveis é dada por:

$$E_t = u(X_t, Y_t, Z_t)S_t$$

onde  $u(\cdot)$  é a taxa de utilização de  $S_t$  e assumidamente dependente de variáveis como, Renda per capita  $X_t$ , Preço da energia elétrica  $Y_t$ , e outros fatores (econômicos, sociais e demográficos)  $Z_t$ . Uma suposição para a função utilidade  $u(\cdot)$ , é assumir que:

$$u(X_t, Y_t, Z_t) = a_0 + a_1X_t + a_2Y_t + a_3Z_t$$

um modelo possível e mais realista para a função utilidade  $u(\cdot)$  é:

$$u(X_t, Y_t, Z_t) = a_0 + a_1 \ln(X_t) + a_2 \ln(Y_t) + a_3 \ln(Z_t)$$

Essa classe de modelos são amplamente utilizadas para análise e previsão do consumo de energia a longo prazo.

# Capítulo 2

## Tópicos de Inferência Estatística

### 2.1 Inferência Estatística

Para colocar o problema básico de inferência estatística vamos considerar que  $\{x_1, x_2, \dots, x_n\}$  são observações de uma variável aleatória não necessariamente independentes ou identicamente distribuídas, com função distribuição de probabilidade conjunta  $F_{\mathbf{X}}(x_1, x_2, \dots, x_n | \Theta)$ , onde  $\Theta = (\theta_1, \dots, \theta_p) \in R^p$  é o vetor de parâmetros da distribuição  $F_{\mathbf{X}}(\cdot)$ . Denotamos um estimador de  $\Theta$ , como uma função  $t_n(x_1, x_2, \dots, x_n)$  das observações e usamos também a notação  $\hat{\Theta} = t_n(x_1, x_2, \dots, x_n)$  para indicar um estimador do vetor  $\Theta$ .

O estimador  $t_n(x_1, x_2, \dots, x_n)$  não é único e a escolha do melhor estimador depende de propriedades tais como tendência ou vício, convergência, eficiência, etc.

**Definio 35** Um estimador  $t_n(x_1, x_2, \dots, x_n)$  é um estimador não tendenciosos de  $\Theta$  se:

$$E\{t_n(x_1, x_2, \dots, x_n) - \Theta\} = 0 \quad (2.1)$$

**Definio 36** O vício ou tendência de um estimador  $t_n((x_1, x_2, \dots, x_n)$  de  $\Theta$  é dado por:

$$E\{t_n(x_1, x_2, \dots, x_n)\} - \Theta = b_n(\Theta) \quad (2.2)$$

Um estimador  $t_n(x_1, x_2, \dots, x_n)$  é frequentemente tendencioso ou viciado, mas em geral a tendência  $b_n(\Theta)$  vai para zero quando  $n \rightarrow \infty$ , por exemplo:

Se  $t_n = s_n + (1/n)$ , onde  $s_n$  é um estimador não tendencioso, então  $\lim t_n \rightarrow s_n$ . Neste caso se diz que  $t_n$  é assintoticamente não tendencioso.

**Definio 37** Um estimador  $t_n(x_1, x_2, \dots, x_n)$  é um estimador assintoticamente não tendenciosos de  $\Theta$  se:

$$\lim_{n \rightarrow \infty} b_n(\Theta) = 0 \quad \text{para todo } \Theta \in R^p \quad (2.3)$$

Uma outra propriedade que queremos para o estimador  $t_n(x_1, x_2, \dots, x_n)$  é que ele convirja em probabilidade para o parâmetro verdadeiro  $\Theta$  quando  $n \rightarrow \infty$ .

**Definio 38** Um estimador  $t_n((x_1, x_2, \dots, x_n))$  é um estimador consistente de  $\Theta$  se somente se:

$$\lim_{n \rightarrow \infty} t_n((x_1, x_2, \dots, x_n)) \xrightarrow{P} \Theta \quad \text{para todo } \Theta \in R^p$$

**Exemplo 39** Seja  $\{x_1, x_2, \dots, x_n\}$  observações de uma variável aleatória independente e identicamente distribuida (i.i.d.) com média  $\mu$  e variância  $\sigma^2$  e considere as estatísticas (estimadores):

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.4)$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) \quad (2.5)$$

$$U^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \left( \sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) \quad (2.6)$$

Queremos verificar se esses estimadores para  $\mu$  e  $\sigma^2$  são tendenciosos.

1) Considerando  $\bar{X}$  temos:

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(x_i) = \mu$$

2) Considerando  $U^2$ , temos:

$$E(U^2) = \frac{1}{n} \left( \sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2) \right)$$

$$E(U^2) = (\sigma^2 + \mu) - E(\bar{X}^2)$$

$$E(U^2) = (\sigma^2 + \mu) - \left( \frac{\sigma^2}{n} + \mu \right)$$

$$E(U^2) = \frac{n-1}{n} \sigma^2 \neq \sigma^2$$

3) Agora considerando  $S^2$ , temos:

$$S^2 = \frac{n}{n-1} U^2$$

$$E(S^2) = \frac{n}{n-1} E(U^2)$$

$$E(S^2) = \sigma^2$$

Portanto  $\bar{X}$  e  $S^2$  são estimadores não tendenciosos de  $\mu$  e  $\sigma^2$  respectivamente, mas  $U^2$  é um estimador tendenciosos de  $\sigma^2$ . A tendência do estimador  $U^2$  é dada por:

$$b_n(\sigma^2) = E(U^2) - \sigma^2$$

$$b_n(\sigma^2) = -\frac{\sigma^2}{n}$$

Note que  $\lim b_n(\sigma^2) \rightarrow 0$ , quando  $n \rightarrow \infty$ . Portanto  $U^2$  é um estimador assintoticamente não tendencioso de  $\sigma^2$ . Para analisar a consistência dos estimadores vamos considerar a desigualdade de Chebyshev para  $\bar{X}$ , dada por:

$$P(|\bar{X} - \mu| \geq \epsilon) \leq \frac{V(\bar{X})}{\epsilon^2}$$

Sendo:

$$V(\bar{X}) = V \left( \frac{1}{n} \sum_{i=1}^n x_i \right)$$

$$V(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n V(x_i) = \frac{\sigma^2}{n}$$

Temos que:

$$P(|\bar{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

$$\lim_{n \rightarrow \infty} P(|\bar{X} - \mu| \geq \epsilon) = 0$$

Portanto  $\bar{X} \xrightarrow{P} \mu$ . Com um pouco mais de cálculos algébricos podemos mostrar da mesma forma que,  $S^2 \xrightarrow{P} \sigma^2$  e  $U^2 \xrightarrow{P} \sigma^2$ . Assim pode ser mostrado que os estimadores  $\bar{X}$ ,  $S^2$  e  $U^2$  são estimadores consistentes de  $\mu$  e  $\sigma^2$  respectivamente.

**Exercício 40** Gerar amostras de tamanho  $n = 10, 30, \dots, 1000$  de uma variável aleatória  $N(5, 1)$  e calcular os valores dos estimadores  $\bar{X}$ ,  $S^2$  e  $U^2$  para cada uma das amostras. Faça um gráfico de  $\bar{X}$ ,  $S^2$  e  $U^2$  por  $n$  e verifique a convergência dos estimadores graficamente.

## 2.2 Estimativa de Máxima Verossimilhança (MLE)

**Definio 41** Se  $\{x_1, x_2, \dots, x_n\}$  são observações de uma variável aleatória não necessariamente independentes ou identicamente distribuídas, com função densidade de probabilidade conjunta  $f_{\mathbf{X}}(x_1, x_2, \dots, x_n | \Theta)$ , onde  $\Theta = (\theta_1, \dots, \theta_p)$  é o vetor de parâmetros de  $f_{\mathbf{X}}(\cdot)$ , a função de verossimilhança de  $\mathbf{X}$  é definida por:

$$L(\Theta) = f_{\mathbf{X}}(x_1, x_2, \dots, x_n | \Theta) \quad (2.7)$$

Se  $\{x_1, x_2, \dots, x_n\}$  é uma amostra aleatória independente então:

$$L(\Theta) = f_{\mathbf{X}}(x_1, x_2, \dots, x_n | \Theta) = \prod_{i=1}^n f_{\mathbf{X}}(x_i | \Theta) \quad (2.8)$$

Qualquer estimador  $\hat{\Theta} = t_n(x_1, x_2, \dots, x_n)$ , tal que:

$$L(\hat{\Theta}) = \sup\{L(\Theta) : \Theta \in R^p\}$$

É chamado estimador de máxima verossimilhança (**MLE**).

O estimador  $\hat{\Theta}$  que maximiza  $L(\Theta)$  é o mesmo que maximiza  $\log L(\Theta)$ , é mais prático então encontrar  $\hat{\Theta}$  maximizando  $l(\Theta) = \log L(\Theta)$ . No caso de uma amostra independente, temos:

$$\hat{\Theta} = \max_{\Theta} l(\Theta) = \max_{\Theta} \left\{ \sum_{i=1}^n \log f_{\mathbf{X}}(x_i | \Theta) \right\} \quad (2.9)$$

**Exemplo 42** Seja  $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$  observações independentes de uma variável aleatória  $X$  normalmente distribuída como média  $\mu$  e variância  $\sigma^2$ . Neste caso temos  $\Theta = (\mu, \sigma^2)$  e a função de verossimilhança dessa amostra é dada por:

$$L(\Theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x_i - \mu)^2}{2\sigma^2} \right\}$$

$$L(\Theta) = \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\}$$

Sendo:

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n x_i^2 - 2n\mu\bar{X} + n\mu^2 + n\bar{X}^2 - n\bar{X}^2$$

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n x_i^2 - n\bar{X}^2 + n(\mu^2 - 2\mu\bar{X} + \bar{X}^2)$$

$$\sum_{i=1}^n (x_i - \mu)^2 = (n-1)S^2(\mathbf{X}) + n(\bar{X} - \mu)^2$$

onde  $S^2(\mathbf{X}) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$

Assim a função de verossimilhança para  $\mathbf{X}$  pode ser escrita como:

$$L(\Theta) = \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} [n(\bar{X} - \mu)^2 + (n-1)S^2(\mathbf{X})] \right\}$$

Tomando-se o  $l(\Theta) = \log L(\Theta)$ , temos:

$$l(\Theta) = \frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} [n(\bar{X} - \mu)^2 + (n-1)S^2(\mathbf{X})]$$

Derivando  $l(\Theta)$  com relação aos parâmetros e igualando a zero, para encontrar os estimadores  $\hat{\mu}$  e  $\hat{\sigma}^2$  que maximizam  $l(\Theta)$ , temos:

$$\frac{\partial l(\Theta)}{\partial \mu} = \bar{X} - \mu = 0$$

$$\hat{\mu} = \bar{X}$$

$$\frac{\partial l(\Theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} [n(\bar{X} - \mu)^2 + (n-1)S^2(\mathbf{X})] = 0$$

fazendo  $\mu = \bar{X}$ ,

$$\hat{\sigma}^2 = \frac{n-1}{n} S^2(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{X}^2$$

Portanto o estimador dado por (1.1) é o estimador de máxima verossimilhança de  $\mu$ , mas o estimador de máxima verossimilhança de  $\sigma^2$  não é (1.2) e sim o estimador viciado dado em (2.6)

### 2.2.1 Qualidade do Estimador

Usa-se o conceito de concentração de uma variável aleatória em torno de um ponto para comparar dois estimadores. Um estimador  $\hat{\Theta}_1$  é mais concentrado em  $\Theta$  que outro  $\hat{\Theta}_2$  se para todo  $r > 0$ , temos:

$$P\left(\left|\hat{\Theta}_1 - \Theta\right| \leq r\right) \geq P\left(\left|\hat{\Theta}_2 - \Theta\right| \leq r\right) \quad (2.10)$$

Se a desigualdade (2.10) vale para pelo menos um  $r > 0$  então o estimador  $\hat{\Theta}_1$  é preferível ao estimador  $\hat{\Theta}_2$ .

**Definio 43** Um estimador  $t_n(x_1, x_2, \dots, x_n)$  é tal que  $\sqrt{n}(t_n - \Theta)$  tem distribuição assintótica  $N(0, J^2(\Theta))$  para alguma função  $J(\Theta)$  é chamado estimador consistente normal assintótico (CAN).

**Nota 44** Quando  $n$  é grande o estimador de máxima verossimilhança (MLE) é o estimador consistente normal assintótico (CAN) mais concentrado.



## 2.3 Inferência Bayesiana

Na abordagem bayesiana o vetor de parâmetros  $\Theta = (\theta_1, \dots, \theta_p) \in R^p$  a ser estimado é considerado um vetor aleatório, de forma que é razoável imaginar que um especialista possa conhecer alguma informação a *priori* sobre esses parâmetros. Esta informação pode ser resumida em uma densidade de probabilidade conjunta, **densidade a priori**  $\pi_0(\Theta)$ . Agora vamos considerar as observações  $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$  de uma variável aleatória não necessariamente independentes ou identicamente distribuídas, com função densidade de probabilidade condicionada nas variáveis  $\Theta$  que denotamos por  $f_{\mathbf{X}}(x_1, x_2, \dots, x_n | \Theta)$ . Usando o *Teorema de Bayes* (1.10) podemos calcular a densidade conjunta do vetor de parâmetros  $\Theta$ , condicionada nas observações  $\mathbf{X}$  por:

$$\pi_1(\Theta | \mathbf{X}) = \frac{f_{\mathbf{X}}(x_1, x_2, \dots, x_n | \Theta) \pi_0(\Theta)}{\int_{\Theta} f_{\mathbf{X}}(x_1, x_2, \dots, x_n | \Theta) \pi_0(\Theta) d\Theta} \quad (2.11)$$

A **densidade a posteriori**  $\pi_1(\Theta | \mathbf{X})$  é interpretada como a informação sobre os parâmetros, que se obtém após (a posteriori) a observação dos dados  $\mathbf{X}$ . Essa densidade incorpora ambas informações vindas da experiência prévia e dos dados observados. Portanto a técnica de inferência Bayesiana permite considerar a experiência de especialistas no modelo estatístico. Notando que a integral no denominador de (2.11) é sobre o espaço dos parâmetros, então esse termo só depende das observações conhecidas. Assim podemos sem perda, considerar somente o numerador de (2.11), ou seja:

$$\pi_1(\Theta | \mathbf{X}) \propto f_{\mathbf{X}}(x_1, x_2, \dots, x_n | \Theta) \pi_0(\Theta) \quad (2.12)$$

O símbolo “ $\propto$ ” indica proporcionalidade.

Queremos encontrar o estimador Bayesiano  $\hat{\Theta}$  que minimize uma função perda  $J(\Theta, \hat{\Theta}) = E(\Theta - \hat{\Theta})^2$ , ou seja uma perda quadrática. Pode ser mostrado que esse estimador é dado pelo valor esperado condicionado  $E(\Theta | \mathbf{X})$  com relação a densidade a posteriori  $\pi_1(\Theta | \mathbf{X})$ , ou seja:

$$\hat{\Theta} = \int_{\Theta} \Theta \pi_1(\Theta | \mathbf{X}) d\Theta \quad (2.13)$$

**Exemplo 45** *Supor que o conjunto de ( $n = 30$ ) observações da Tabela 1 é uma amostra aleatória independente de uma v.a.  $X$  com distribuição Normal. Para se fazer a inferência Bayesiana da média  $\mu$  e da variância  $\sigma^2$  desses*

dados, vamos considerar a função de verossimilhança dada por:

$$f_{\mathbf{X}}(\Theta|\mathbf{X}) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2} [n(\bar{X} - \mu)^2 + (n-1)S^2(\mathbf{X})]\right\}$$

Assumindo agora que a opinião de um especialista sobre esse equipamento assegura que a média  $\mu$  varia normalmente em torno de uma média  $\mu_0$  com um desvio-padrão de  $\sigma_0^2 = \sigma^2/k_0$ , onde  $k_0$  é uma constante conhecida, temos  $\mu|\sigma^2 \sim N(\mu_0, \sigma^2/k_0)$  e a variância  $\sigma^2$  tem distribuição  $\text{Inv-}\chi^2(\nu_0, \sigma_0^2)$  então a densidade a priori é dada por:

$$\begin{aligned} \pi_0(\Theta) &\propto \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}\right\} \left(\frac{1}{\sigma^2}\right)^{(\nu_0/2+1)} \exp\left\{-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right\} \\ \pi_0(\Theta) &\propto \sigma^{-1} (\sigma^2)^{-(\nu_0/2+1)} \exp\left\{-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2 + k_0(\mu - \mu_0)^2]\right\} \end{aligned}$$

Usando Teorema de Bayes para combinar as informações a priori de  $\pi_0(\Theta)$  e as informações nos dados em  $f_{\mathbf{X}}(\Theta|\mathbf{X})$ , temos a densidade a posteriori, dada por:

$$\pi_1(\Theta|\mathbf{X}) \propto f_{\mathbf{X}}(\Theta|\mathbf{X})\pi_0(\Theta)$$

$$\begin{aligned} \pi_1(\Theta|\mathbf{X}) &\propto \sigma^{-1} (\sigma^2)^{-(\nu_0/2+1)} \exp\left\{-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2 + k_0(\mu - \mu_0)^2]\right\} \times \\ &\times (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} [n(\bar{X} - \mu)^2 + (n-1)S^2(\mathbf{X})]\right\} \end{aligned}$$

Após alguns cálculos algébricos mostra-se que:

$$\pi_1(\Theta|\mathbf{X}) \propto \sigma^{-1} (\sigma^2)^{-(\nu_n/2+1)} \exp\left\{-\frac{1}{2\sigma^2}[\nu_n\sigma_n^2 + k_n(\mu - \mu_n)^2]\right\}$$

com:

$$\begin{aligned} \mu_n &= \frac{k_0}{k_0 + n}\mu_0 + \frac{n}{k_0 + n}\bar{X} \\ k_n &= k_0 + n \\ \nu_n &= \nu_0 + n \\ \nu_n\sigma_n^2 &= \nu_0\sigma_0^2 + (n-1)S^2(\mathbf{X}) + \frac{k_0n}{k_0 + n}(\bar{X} - \mu_0)^2 \end{aligned}$$

Note que os parâmetros da densidade a priori combinam as informações a priori e as informações contidas nos dados, por exemplo,  $\mu_n$  é a média ponderada da média a priori  $\mu_0$  e da média amostral  $\bar{X}$  com peso determinado pela precisão relativa das duas fontes de informação. Assim, se  $n$  é muito grande e  $k_0$  é muito pequeno a média a posteriori tende a se aproximar da média amostral  $\bar{X}$ , caso contrário a média a priori passa a assumir um papel importante na estimativa de  $\mu_n$ . A soma de quadrados a posteriori  $\nu_n \sigma_n^2$  combina a soma de quadrados a priori  $\nu_0 \sigma_0^2$ , a soma de quadrados amostral  $(n-1)S^2(\mathbf{X})$  e a incerteza adicional trazida pela diferença entre a média a priori e a média amostral.

**Exemplo 46** Usando os dados da Tabela 1, temos  $n = 30$ ,  $\bar{X} = 5.0910$ ,  $S^2 = 0.6137$ . Assumindo os parâmetros da priori dados por  $\mu_0 = 5$ ,  $\nu_0 = 10$  e  $k_0 = 1$  que implica em  $\sigma_0^2 = \sigma^2$ . Podemos estimar os parâmetros da posteriori como:

$$\begin{aligned}\mu_n &= \frac{k_0}{k_0 + n} \mu_0 + \frac{n}{k_0 + n} \bar{X} \\ \mu_n &= \frac{5}{31} + \frac{30 \times 5.0910}{31} = 5.0880\end{aligned}$$

Aumentando a precisão a priori, fazendo  $k_0 = 100$ , temos  $\mu_n = 5.021$ , ou seja uma estimativa muito mais próxima da média a priori  $\mu_0 = 5$ .

### 2.3.1 Algoritmo Amostrador de Gibbs

Considerando a densidade a posteriori  $\pi(\Theta|\mathbf{X}) = \pi(\theta_1, \dots, \theta_p|\mathbf{X})$ , podemos encontrar as densidades condicionais  $\pi_k(\theta_k|\Theta_{-k}, \mathbf{X})$ , onde estamos denotando por  $\Theta_{-k} = (\theta_1, \dots, \theta_{k-1}, \theta_{k+1}, \dots, \theta_p)$ , ou seja todos os parâmetros exceto o parâmetro  $\theta_k$ . Essas densidades condicionais podem ser usadas para estimar o parâmetro  $\theta_k$  usando algoritmo de simulação de Monte Carlo em Cadeia de Markov (MCMC). Vamos ilustrar o uso desta técnica com o exemplo anterior. Para isso temos as densidades condicionais:

$$p(\mu|\sigma^2, \mathbf{X}) \propto \exp \left\{ -\frac{1}{2\sigma^2/k_n} (\mu - \mu_n)^2 \right\} \quad (2.14)$$

$$(\mu|\sigma^2, \mathbf{X}) \sim \text{Normal} \left( \frac{k_0 \mu_0 + n \bar{X}}{k_0 + n}, \sigma^2/k_n \right) \quad (2.15)$$

e integrando a posteriori  $\pi_1(\mu, \sigma^2 | \mathbf{X})$  em relação a  $\mu$ , temos que:

$$q(\sigma^2 | \mu, \mathbf{X}) \propto (\sigma^2)^{-(\nu_n/2+1)} \exp \left\{ -\frac{\nu_n \sigma_n^2}{2\sigma^2} \right\} \quad (2.16)$$

$$(\sigma^2 | \mu, \mathbf{X}) \sim \text{Inv} - \chi^2(\nu_n, \sigma_n^2) \quad (2.17)$$

O algoritmo Amostrador de Gibbs, consiste nos seguintes passos:

**Algoritmo 47** *Amostrador de Gibbs*

- 1) Adote condições iniciais  $\mu^{(0)}, (\sigma^2)^{(0)}$  e faça o contador  $i = 1$ .
- 2) Gerar  $\mu^{(i)} \sim p(\mu | (\sigma^2)^{(i-1)}, \mathbf{X})$ .
- 3) Gerar  $(\sigma^2)^{(i)} \sim q(\sigma^2 | \mu^{(i)}, \mathbf{X})$ .
- 4) Faça  $i \leftarrow i + 1$  e repita os passos 2 e 3 até a convergência.

A amostra gerada  $\{\mu^{(i)}, (\sigma^2)^{(i)}, i = 1, \dots, M\}$  é usada para calcular estimativas de Monte Carlo dos parâmetros dadas de forma geral por:

$$\widehat{g}(\theta) = \frac{1}{M} \sum_{i=1}^M g(\theta^{(i)}) \quad (2.18)$$

Existem vários critérios de diagnóstico de convergência, assim como vários algoritmos de MCMC que podem ser usados para fazer inferência Bayesiana de parâmetros de modelos complicados. Mas esse assunto não é parte dos nossos objetivos aqui.

# Capítulo 3

## Tópicos de Processos Estocástico

Um problema muito frequente em várias áreas de aplicação surge quando é necessário a análise de uma variável aleatória que é observada ao longo do tempo. Como por exemplo o número de falhas que um equipamento apresenta durante 1 ano de uso, o número de pessoas que chegam a um balcão solicitando informação ou o consumo de energia elétrica de uma residência ou de uma área residencial atendida por uma empresa. Nestes casos a variável aleatória de interesse, aparece indexada  $\{X_t, t \geq 0\}$  e é denominada um *processo estocástico*. O principal objetivo ao se analisar um processo estocástico é determinar qual a lei probabilística que rege esse processo e se essa lei é invariante no tempo, ou seja se o processo é estacionário.

### 3.1 Estacionariedade

**Definio 48** Um processo estocástico  $\{X_t, t \geq 0\}$  é estacionário no sentido forte se sua distribuição de probabilidade conjunta é invariante no tempo, ou seja:

$$F(x_{t_1}, \dots, x_{t_n}) = F(x_{t_1+\tau}, \dots, x_{t_n+\tau}) \quad \forall n, \tau \in R \quad (3.1)$$

**Definio 49** Um processo estocástico  $\{X_t, t \geq 0\}$  é estacionário no sentido fraco se:

- 1  $E\{X_t\} = \mu$  (média constante independente de  $t$ )
- 2  $V(X_t) = \sigma^2$  (variância constante independente de  $t$ )
- 3  $Cov(X_t, X_{t+k}) = \gamma(k)$  (covariância entre  $X_t$  e  $X_{t+k}$  só depende de  $k$ )

• **Análise gráfica:**

Um primeiro passo na análise um processo estocástico é fazer o gráfico da trajetória do processo. Mas essa análise em geral fornece poucas informações sobre o comportamento probabilístico do processo. Mas permite avaliar se o processo é estacionário na média e variância. Para ilustrar vamos considerar alguns exemplos:

**Exemplo 50** Processo  $\{X_t, t \geq 0\}$  gerada com um modelo auto-regressivo -  $AR(1)$ :  $X_{t+1} = 50 + 1.01X_t + a_t$  onde  $\{a_t, t \geq 0\}$  é um ruído gaussiano com variância  $\sigma^2$ .

Note que para  $\sigma^2 = 10$  temos mais incerteza no comportamento do processo do que com  $\sigma^2 = 1$ , de forma que a observação do gráfico destes dados não nos assegura sobre qual modelo matemático deve ser adotado para descrever o comportamento do processo  $\{X_t, t \geq 0\}$ . A análise gráfica neste caso só nos assegura que não há estacionariedade na média (existe uma tendência).

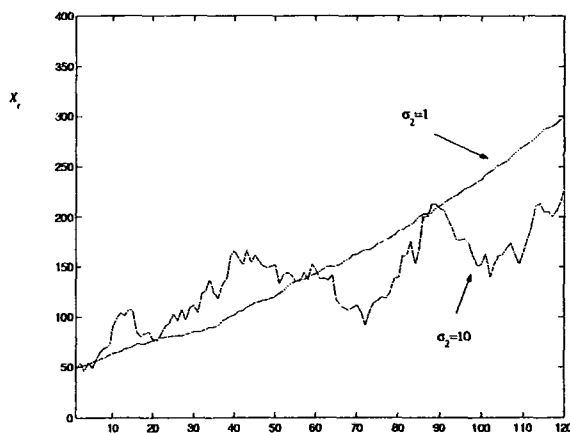


Figura 17 - Processo gerado com  $\sigma^2 = 1$  e  $\sigma^2 = 10$

**Exemplo 51** Neste exemplo temos uma série de carga horária para o período de 5 dias úteis (120 horas). A série periódica não tem estacionariedade na média, mas observando a Figura 18, aparentemente a variância é a mesma para as diferentes horas do dia.

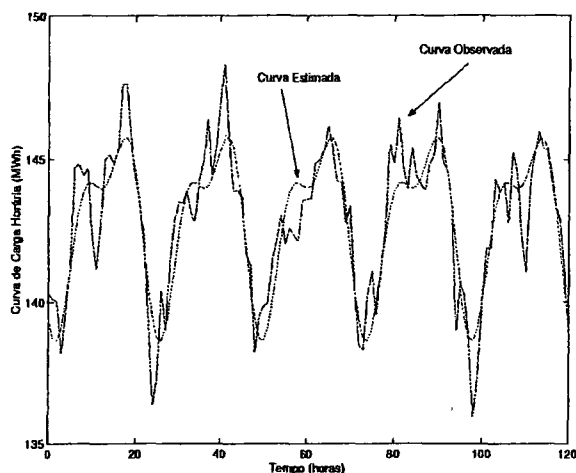


Figura 18 Série de carga horária (série periódica)

**Exemplo 52** *Análise do consumo de energia elétrica em MW-hora, no setor residencial na área de concessão de uma empresa no período de janeiro de 1975 a dezembro de 1985.*

*Tabela 8 Consumo de energia elétrica no setor residencial*

Mês\Ano	75	76	77	78	79	80	81	82	83	84	85
jan	63	77	81	95	106	124	156	154	165	168	184
fev	62	66	76	92	100	116	139	135	155	174	180
mar	75	72	82	102	110	119	151	146	151	177	173
abr	59	74	81	97	109	114	127	146	173	167	187
mai	69	78	88	98	114	127	139	143	152	172	184
jun	65	75	82	99	104	119	129	149	170	172	185
jul	63	76	89	96	114	121	138	150	161	168	166
ago	68	78	89	99	106	143	134	147	158	172	187
set	69	79	90	101	113	112	143	157	156	184	194
out	74	80	92	105	114	123	137	149	170	166	188
nov	73	84	91	101	117	132	143	160	166	193	208
dez	68	83	103	107	125	133	138	153	160	171	188

O gráfico deste processo mostra um crescimento linear deste consumo ao longo do tempo e uma provável mudança na variância nos anos 80 e 81.

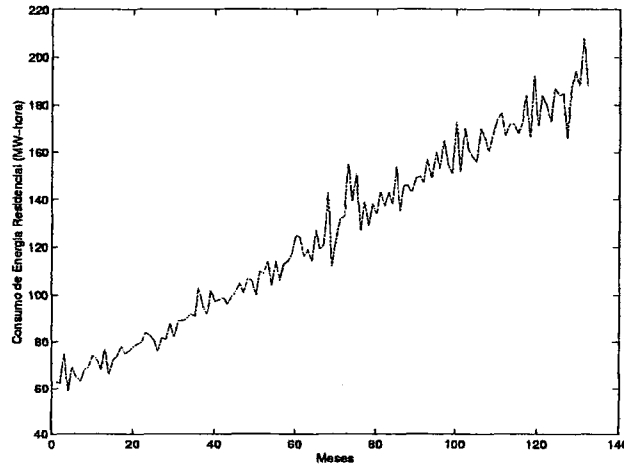


Figura 19 Consumo de energia elétrica residencial de 1975 a 1985

### 3.2 Função de Autocorrelação (FAC)

Uma estatística importante na análise de um processo estocástico é a autocorrelação, ou seja a correlação serial do processo. Essa estatística indica se há relação linear entre as observações  $X_t$  e  $X_{t+k}$ . A função de autocorrelação ou correlograma de um processo estocástico é definido por:

$$\rho_k = \frac{Cov(X_t, X_{t+k})}{V(X_t)} = \frac{\gamma(k)}{\gamma(0)} \quad \text{para } k = 1, 2, \dots \quad (3.2)$$

A função de autocorrelação pode ser estimada para um processo estacionário a partir das observações de uma realização do processo como:

$$\rho_k = \frac{C_k}{C_0} \quad (3.3)$$



onde

$$C_k = \frac{1}{N-k} \sum_{t=1}^{T-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad (3.4)$$

$$C_k = \frac{1}{N-k} \sum_{t=1}^{T-k} X_t X_{t+k} - \bar{X}^2 \quad (3.5)$$

$$C_0 = \frac{1}{N-1} \sum_{t=1}^T X_t^2 - \bar{X}^2 \quad \text{e} \quad \bar{X} = \frac{1}{N} \sum_{t=1}^T X_t \quad (3.6)$$

Para exemplificar o uso da função autocorrelação na análise de um processo estocástico, vamos considerar o processo que representa o consumo de energia elétrica no setor comercial, cuja trajetória observada nos últimos 11 anos é dada na Tabela 8. Aplicando a análise de regressão linear ajustando-se uma reta para a tendência de crescimento linear do processo, usando como variável explicativa  $X = t$  (meses), temos:

$$Y_t = \alpha_0 + \alpha_1 t + \epsilon_t$$

$$Y_t = 58.0449 + 1.0056t + \epsilon_t$$

Os gráficos da Figura 20, mostram a reta ajustada e os resíduos  $\epsilon_t$ . Vamos considerar  $\epsilon_t$  como um processo estocástico.

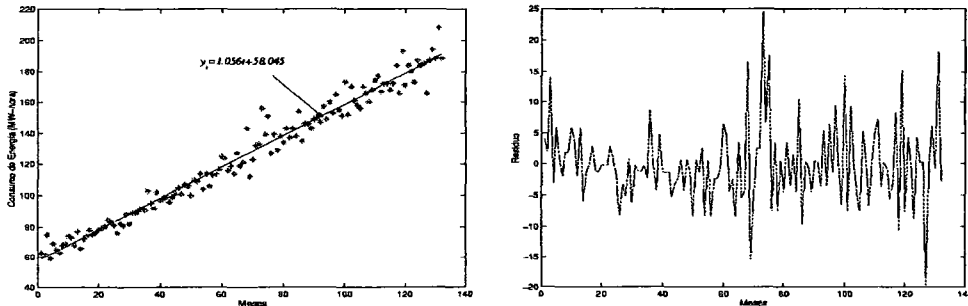


Figura 20 A tendência e os resíduos para o consumo de energia residencial

Sob a hipótese de estacionariedade, podemos calcular sua função de autocorrelação dos resíduos. A Figura 21 mostra o gráfico desta função. Para  $k = 1, 2, \dots, 20$ , e os gráficos que mostram a relação linear entre  $\epsilon_t$  e  $\epsilon_{t-1}$  e

$\epsilon_{t-2}$  respectivamente.

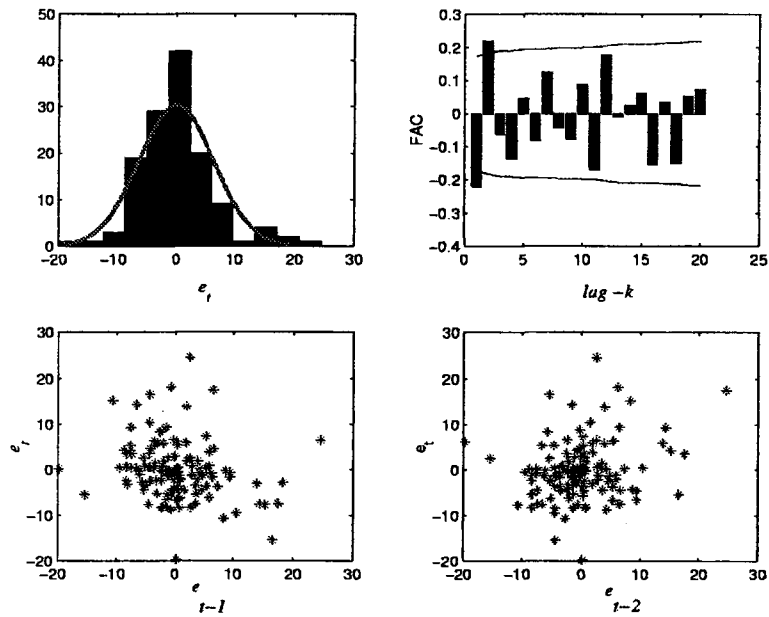


Figura 21 Análise dos resíduos como um processo estocástico

Os gráficos da Figura 21 assim como a FAC indicam que podemos considerar um modelo linear de auto-regressão (geralmente denominado modelo auto-regressivo) para escrever o resíduo  $\epsilon_t$  como uma função linear dos valores  $\epsilon_{t-1}$  e  $\epsilon_{t-2}$ . Assim podemos pensar em um modelo estocástico para representar o consumo de energia elétrica neste setor residencial como:

$$y_t = \alpha_0 + \alpha_1 t + \epsilon_t \quad (3.7)$$

onde  $\epsilon_t$  é modelado como:

$$\epsilon_t = \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + a_t \quad (3.8)$$

Onde  $\{a_t, t \geq 0\}$  é um processo não correlacionado (ruído branco). Assim o modelo estocástico completo usado para representar o consumo de energia elétrica é dado por:

$$y_t = \alpha_0 + \alpha_1 t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + a_t \quad (3.9)$$

O estudo detalhado que visa identificar e ajustar essa classe de modelos auto-regressivos é denominado análise de séries temporais. Esse é o tema abordado nos próximos Capítulos.

**Parte II**

**Modelos para Séries Temporais**

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that this is crucial for ensuring transparency and accountability in the organization's operations.

2. The second part of the document outlines the various methods and tools used to collect and analyze data. It highlights the need for robust data management systems that can handle large volumes of information efficiently and securely.

3. The third part of the document focuses on the role of technology in enhancing data collection and analysis. It discusses the use of advanced software and hardware solutions to streamline the data processing workflow and improve the accuracy of the results.

4. The fourth part of the document addresses the challenges associated with data collection and analysis. It identifies common issues such as data quality, data integration, and data security, and provides strategies to overcome these challenges.

5. The fifth part of the document concludes by summarizing the key findings and recommendations. It stresses the importance of continuous monitoring and evaluation of the data collection and analysis process to ensure its effectiveness and relevance over time.

# Capítulo 4

## Modelos com Tendência e Sazonalidade

### 4.1 Introdução

Um modelo clássico para séries temporais supõe que a série temporal  $\{X_t, t \geq 0\}$  pode ser escrita como a soma de três componentes: Uma tendência  $T_t$ , uma componente sazonal  $S_t$ , e um termo estocástico  $Z_t$ . Assim, temos a série  $X_t$  representada pelo modelo composto:

$$X_t = T_t + S_t + Z_t \quad (4.1)$$

A componente sazonal só aparece e pode ser estimada quando a série é observada por vários períodos, por exemplo em séries de período anual, vários anos devem ser observados e usados para se identificar a componente  $S_t$ , já em séries de período diário (24 horas), vários dias típicos precisam ser observados. Outros modelos que podem ser usados para séries temporais é o modelo multiplicativo, ou seja:

$$X_t = T_t S_t Z_t \quad (4.2)$$

Neste caso o modelo pode ser transformado no modelo aditivo tomando-se o logaritmo da série, ou seja:

$$\log(X_t) = \log(T_t) + \log(S_t) + \log(Z_t) \quad (4.3)$$

Um modelo misto  $X_t = T_t S_t + Z_t$  também pode ser usado quando necessário.

## 4.2 Modelo com Tendência

Vamos supor que a componente sazonal não esteja presente e vamos considerar o modelo aditivo somente com tendência:

$$X_t = T_t + Z_t \quad (4.4)$$

Após ajustar a tendência podemos calcular a série livre de tendência  $Z_t = X_t - T_t$  e analisar a série  $Z_t$  como um processo estocástico estacionário. Há vários métodos de estimar tendência, aqui vamos considerar somente os ajustes polinômiais e os métodos não paramétricos que usam a diferenciação discreta

### 4.2.1 Métodos de regressão para tendência

Neste caso a tendência da série é modelada como um polinômio de grau  $m$  e temos:

$$T_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_m t^m \quad (4.5)$$

O grau  $m$  do polinômio deve ser bem menor que a quantidade de observações  $N$  disponível. Para estimar os parâmetros  $\alpha_j, j = 1, \dots, m$  utilizamos o método dos mínimos quadrados que minimiza a função:

$$f(\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_m) = \sum_{t=1}^N [X_t - (\alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_m t^m)]^2 \quad (4.6)$$

Essa técnica já foi apresentada antes e vamos nos restringir à apresentação de um exemplo, como se segue.

**Exemplo 53** *Considerando os dados do consumo de energia elétrica dado no Capítulo 2, temos:*

$$X_t = \alpha_0 + \alpha_1 t + Z_t$$

*Assim as equações normais de mínimos quadrados são:*

$$\begin{aligned} N\alpha_0 + \alpha_1 \sum_{t=1}^N t &= \sum_{t=1}^N X_t \\ \alpha_0 \sum_{t=1}^N t + \alpha_1 \sum_{t=1}^N t^2 &= \sum_{t=1}^N tX_t \end{aligned}$$

A solução leva às estimativas:

$$\hat{\alpha}_1 = \frac{\sum_{t=1}^N tX_t - \frac{1}{N} \left( \sum_{t=1}^N t \right) \left( \sum_{t=1}^N X_t \right)}{\sum_{t=1}^N t^2 - \frac{1}{N} \left( \sum_{t=1}^N t \right)^2}$$

$$\hat{\alpha}_0 = \bar{X}_t - \hat{\alpha}_1 \bar{t}$$

onde:

$$\bar{X}_t = \frac{1}{N} \sum_{t=1}^N X_t \quad e \quad \bar{t} = \frac{1}{N} \sum_{t=1}^N t = \frac{(N+1)}{2}$$

temos:

$$\hat{\alpha}_0 = 58.0449 \quad e \quad \hat{\alpha}_1 = 1.0056$$

O cálculo da componente estocástica pode ser feito pela diferença  $Z_t = X_t - (\hat{\alpha}_0 + \hat{\alpha}_1 t)$ . A única questão que resta até o momento é saber o quanto a componente estocástica está influenciando a estimativa da tendência e vice-versa.

### 4.2.2 Diferenciação discreta

Quando a tendência de uma série é determinada por um polinômio, então é fácil mostrar que tomando-se um número apropriado de diferenças da série chega-se à uma série livre de tendência., isto é:

$$X_t = T_t + Z_t$$

$$T_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_m t^m$$

Denotando o operador diferença por:

$$\Delta X_t = X_t - X_{t-1}$$

$$\Delta^2 X_t = \Delta(\Delta X_t) = X_t - 2X_{t-1} + X_{t-2}$$

$$\vdots = \vdots$$

$$\Delta^d X_t = \Delta(\Delta^{d-1} X_t)$$

Aplicando o operador diferença para a tendência  $T_t$ , temos que:

$$\Delta^d T_t = \begin{cases} m! & \text{se } d = m \\ 0 & \text{se } d > m \end{cases}$$

Considerando o exemplo anterior, onde  $X_t = T_t + Z_t = \alpha_0 + \alpha_1 t + Z_t$ , e fazendo a primeira diferença, temos:

$$\begin{aligned}\Delta X_t &= X_t - X_{t-1} = \Delta T_t + \Delta Z_t \\ \Delta X_t &= [\alpha_0 + \alpha_1 t] - [\alpha_0 + \alpha_1(t-1)] + Z_t - Z_{t-1} \\ \Delta X_t &= \alpha_1 + (Z_t - Z_{t-1})\end{aligned}$$

Quando a tendência é um polinômio de grau  $m$ , então são necessárias  $m$  diferenciações. Se  $Z_t$  é um ruído branco, a série resultante  $\Delta^d Z_t$  é uma série estacionária na média (com média zero e variância  $m!\sigma^2$ ) e livre de tendência.

**Exercício 54** *Considere o processo representado pelo modelo  $X_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + Z_t$ . Mostre que fazendo duas diferenciações chega-se ao processo livre de tendência:*

$$\Delta^2 X_t = 2\alpha_2 + Z_t - 2Z_{t-1} + Z_{t-2}$$

Desta maneira, utilizando o procedimento de diferenças, temos que para um processo com tendência polinomial de grau  $m$ , a série livre de tendência será dada por  $Y_t = \Delta^m X_t$ . Se a série apresentar um crescimento exponencial (ou se o modelo for multiplicativo  $X_t = T_t Z_t$ ), o procedimento consiste em tomar a diferença do logaritmo da série original, isto é:

$$\begin{aligned}\Delta X_t^* &= \Delta \log(X_t) = \log(X_t) - \log(X_{t-1}) \\ \Delta X_t^* &= \log(X_t/X_{t-1})\end{aligned}$$

Existem duas desvantagem na estimação de tendência através desse método:  
1) Não obtemos estimativa da tendência para os primeiros  $m$  valores da série e não existe maneira de fazer previsão da tendência.

**Exemplo 55** *Faça uma análise da componente livre de tendência para a série de consumo de energia elétrica do estado do Espírito Santo observada entre janeiro de 1977 e dezembro de 1978*



Tabela 9 Série original e diferença de ordem 1

$t$	$X_t$	$\Delta X_t$	$t$	$X_t$	$\Delta X_t$
1	84.6	...	13	110.3	0.7
2	89.9	5.3	14	118.1	7.8
3	81.9	-8.0	15	116.5	-1.6
4	95.4	13.5	16	134.2	17.7
5	91.2	-4.2	17	134.7	0.5
6	89.8	-1.4	18	144.8	10.1
7	89.7	-0.1	19	144.4	-0.4
8	97.9	8.2	20	159.2	14.8
9	103.4	5.5	21	168.2	9.0
10	107.6	4.2	22	175.2	7.0
11	120.4	12.8	23	174.5	-0.7
12	109.6	-10.8	24	173.7	-0.8

A análise da componente livre de tendência é feita pelos gráficos mostrados na Figura 22.

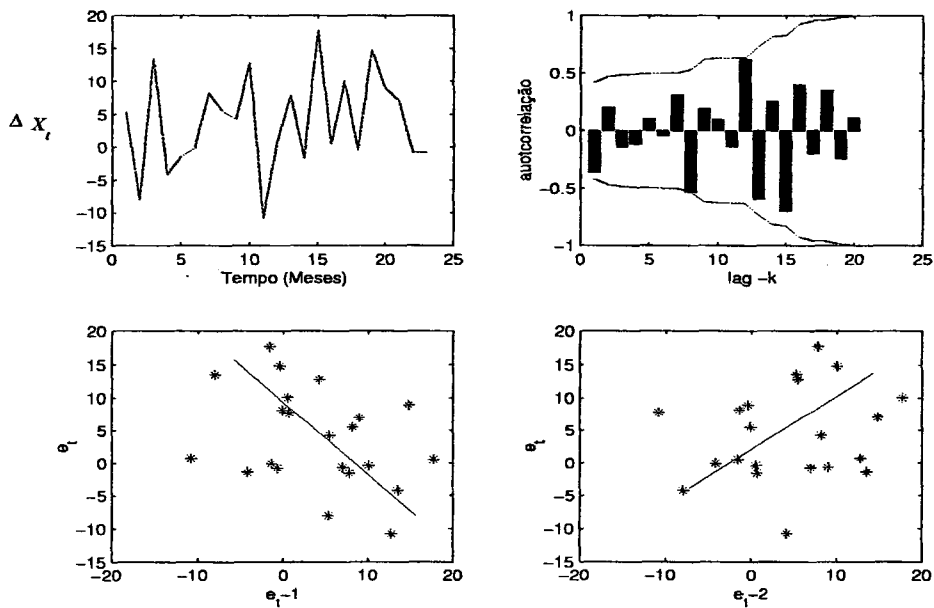


Figura 22 Séries diferenciada e função de autocorrelação dos resíduos

por:

$$\hat{\sigma}_i = \frac{1}{12} \sum_{j=1}^{12} (X_{i,j} - \bar{X}_i)^2, i = 1, \dots, n \quad (4.14)$$

$$\hat{\sigma}_j = \frac{1}{n} \sum_{i=1}^n (X_{i,j} - \bar{X}_j)^2, j = 1, \dots, 12 \quad (4.15)$$

$$\hat{\sigma} = \frac{1}{12n} \sum_{i=1}^n \sum_{j=1}^{12} (X_{i,j} - \bar{X})^2 \quad (4.16)$$

Quando além da média sazonal, temos a variância sazonal, a componente da série livre de sazonalidade é calculada como:

$$Z_t = \frac{X_t - \hat{S}_t}{\hat{\sigma}_j} \quad (4.17)$$

**Exemplo 56** Considere os dados de consumo de energia no setor rural de uma certa região do estado de São Paulo no período de jan/1982 a dez/1986.

Tabela 11 Consumo de energia no setor rural (MWh/1000)

Mês\Ano	1982	1983	1984	1985	1986	$\hat{S}_t$
jan	32.82	34.22	36.52	40.16	45.27	37.79
fev	31.20	33.85	38.75	38.88	44.09	37.35
mar	32.72	32.86	39.05	38.42	40.93	36.79
abr	33.06	35.82	40.09	39.71	42.73	38.28
mai	35.20	36.23	41.21	43.07	46.55	40.45
jun	37.56	38.79	46.28	52.50	50.34	45.09
jul	38.69	42.76	52.58	53.63	57.77	49.08
ago	41.47	44.24	54.02	61.82	59.78	52.26
set	42.23	48.30	51.62	59.32	53.79	51.05
out	40.65	42.05	45.26	55.79	57.93	48.33
nov	37.33	39.50	49.14	56.17	57.43	47.91
dez	34.49	37.13	40.57	49.18	51.96	42.66
Média	36.45	38.81	44.59	49.05	50.71	43.92

A Figura 23 apresenta o gráfico do ciclo  $\hat{S}_t$  estimada pelas médias mensais  $\bar{X}_j$  dadas na última coluna da Tabela 11

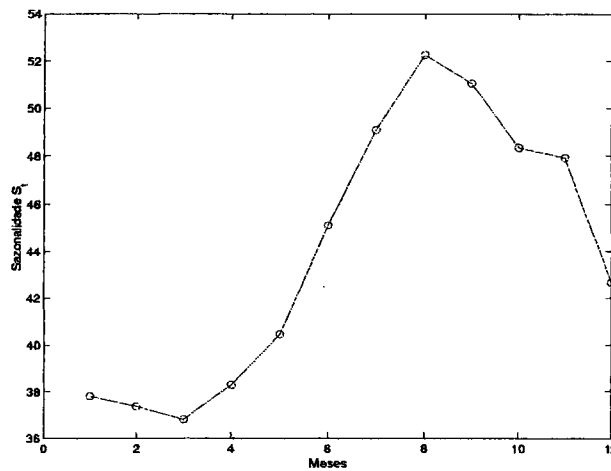


Figura 23 Sazonalidade estimada pela média mensal

As Figuras 24(a) e 24(b) apresentam a série original  $X_t$  e a série livre de sazonalidade  $Z_t = X_t - \hat{S}_t$  respectivamente. Note que a série livre de sazonalidade ainda apresenta uma tendência que pode ser estimada por um polinômio, como já mostramos.

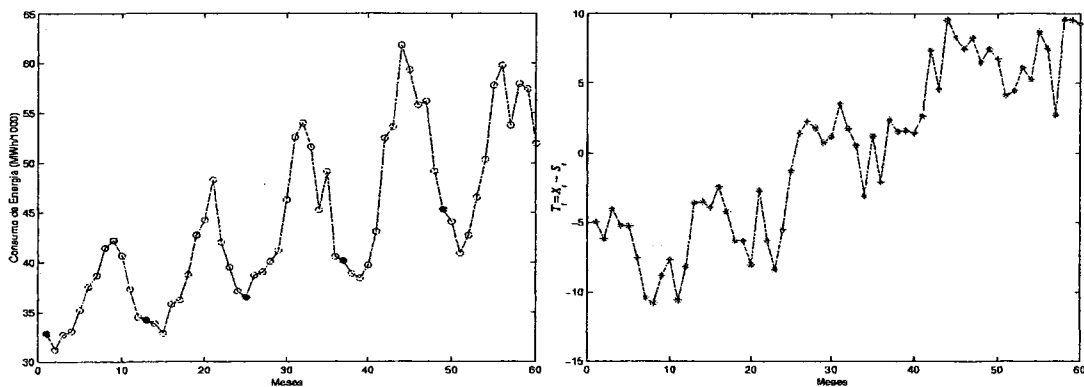


Figura 24 (a) Consumo de energia no setor rural - (b) Série sem sazonalidade

**Exercício 57** Utilizando os dados da Tabela 11 ajuste a tendência  $T_t$  e faça a análise dos resíduos obtidos após a extração de ambas componentes tendência e sazonalidade. Verifique se os resíduos são correlacionados. Considere o modelo completo  $X_t = T_t + S_t + \epsilon_t$ , com  $\{\epsilon_t, t \geq 0\}$  um ruído branco e faça a previsão para 6 passos à frente.

**Exercício 58** Continuando a análise dos dados da Tabela 11, suponha o modelo  $W_t = \log(X_t) = \log(T_t) + \log(S_t) + \epsilon_t$ . Ajuste a tendência e sazonalidade para a série transformada pelo logarítmo. Faça a previsão para 6 passos à frente com este novo modelo.

### 4.3.2 Método de regressão (série de Fourier)

Outra abordagem para analisar séries temporais periódicas é supor que a série é gerada pela superposição de senos e cossenos com diferentes frequências. Uma ferramenta utilizada nesta metodologia é o periodograma que foi originalmente desenvolvido para identificar componentes de senos e cossenos de frequências conhecidas contaminados por ruídos. Para ilustrar o cálculo do periodograma vamos supor que o número de observações  $N = 2q + 1$  é ímpar. O modelo por série de Fourier é:

$$X_t = \alpha_0 + \sum_{i=1}^q \alpha_i \sin(2\pi f_i t) + \beta_i \cos(2\pi f_i t) + Z_t, t = 1, 2, \dots, N \quad (4.18)$$

Aqui estamos considerando que a série  $X_t$  não tem tendência, só tem a sazonalidade (modelada pela série de Fourier) e o ruído  $Z_t$ . Denotando por  $f_i = i/N$  a frequência do  $i$ -ésimo harmônico. A estimativa de mínimos quadrados dos coeficientes  $\alpha_0, (\alpha_i, \beta_i), i = 1, \dots, q$ , é:

$$\hat{\alpha}_0 = \bar{X}_t = \frac{1}{N} \sum_{t=1}^n X_t \quad (4.19)$$

$$\hat{\alpha}_i = \frac{2}{N} \sum_{t=1}^n X_t \sin(2\pi f_i t) \quad (4.20)$$

$$\hat{\beta}_i = \frac{2}{N} \sum_{t=1}^n X_t \cos(2\pi f_i t) \quad (4.21)$$

$$i = 1, 2, \dots, q$$

O periodograma consiste de  $q = (N - 1)/2$  valores  $I(f_i)$  chamados de intensidade da frequência  $f_i$ , calculados por:

$$I(f_i) = \frac{N}{2} (\hat{\alpha}_i^2 + \hat{\beta}_i^2) \text{ para } i = 1, 2, \dots, q \quad (4.22)$$

Quando  $N$  é par, fazemos  $N = 2q$  e aplicamos as equações (4.19)-(4.21) e (4.22) para  $i = 1, 2, \dots, q - 1$  e calculamos  $\hat{\alpha}_q = \frac{1}{N} \sum_{t=1}^n (-1)^t X_t$ ,  $\hat{\beta}_q = 0$

e  $I(f_q) = I(0.5) = N\hat{\alpha}_q^2$ . Note que a frequência mais alta é 0.5 ciclo por intervalo de tempo, porque o menor período é 2 intervalos de tempo. Para ilustrar a aplicação desta técnica vamos utilizar, no exemplo que se segue, dados de uma curva de carga horária para 5 dias úteis, onde só temos a presença de sazonalidade e do ruído, ou seja, a série de carga horária é livre de tendência.

**Exemplo 59** *Considere os dados de carga horária apresentados na Tabela 12 para 5 dias úteis (120 horas) de uma semana típica, sem feriados. Queremos modelar a sazonalidade desta série como uma série de Fourier.*

Tabela 12 Carga horária (MW-hora) para 5 dias úteis de uma semana

<i>hora\dia</i>	<i>seg</i>	<i>ter</i>	<i>qua</i>	<i>qui</i>	<i>sex</i>	<i>Média</i>
1	140.00	140.65	139.72	138.99	139.85	139.84
2	137.42	139.89	139.23	137.50	137.24	138.26
3	139.57	138.83	138.67	139.31	137.75	138.83
4	138.46	138.61	139.25	139.23	139.15	138.94
5	138.77	139.29	139.54	140.62	140.62	139.77
6	138.50	141.59	140.59	142.46	141.84	141.00
7	140.90	143.22	142.12	141.03	142.49	141.95
8	143.25	143.65	142.80	143.50	145.16	143.67
9	143.16	144.43	143.96	142.64	145.33	143.90
10	142.31	143.39	144.71	143.87	143.37	143.53
11	145.19	143.32	144.20	144.28	145.52	144.50
12	145.70	141.80	144.95	142.67	144.62	143.95
13	143.82	145.64	143.73	142.67	145.91	144.36
14	144.43	143.06	145.33	144.29	144.67	144.36
15	145.70	144.69	143.31	144.43	144.39	144.50
16	144.29	146.03	144.92	145.71	145.24	145.24
17	144.60	145.38	143.96	143.44	147.40	144.96
18	146.31	144.12	144.88	146.25	145.72	145.46
19	145.66	143.96	144.80	146.24	143.44	144.82
20	145.00	146.69	145.87	146.86	143.58	145.60
21	144.20	143.65	144.26	144.55	143.58	144.05
22	143.50	143.66	142.59	143.55	142.01	143.06
23	143.03	142.87	142.03	141.95	141.18	142.22
24	141.00	141.85	140.48	139.60	140.15	140.62

A Figura 25 mostrar as médias horárias dadas na última coluna da Tabela 12 e representa uma estimativa da curva de carga de um dia típico.

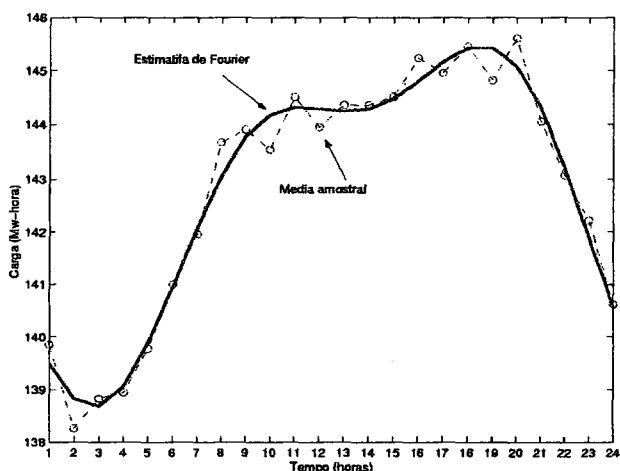


Figura 25 Estimativa da curva de carga de um dia típico

A Figura 26 mostra o periodograma da série da Tabela 12 e a intensidade dos 12 primeiros harmônicos são apresentados na Tabela 13. Estes cálculos identificam como principais harmônicos os de frequência  $f_5 = 1/24$  e  $f_{10} = 1/12$ .

Tabela 13 Cálculo do periodograma da séries de carga horária

$i$	$f_i$	$I(f_i)$	$T_i$	$a_i$	$b_i$
1	0.0083	0.04	120.00	0.0250	-0.0030
2	0.0167	3.55	60.00	0.1810	-0.1627
3	0.0250	6.45	40.00	-0.1160	-0.3067
4	0.0333	5.92	30.00	-0.2183	-0.2257
5	0.0417	507.77	24.00	-2.2399	-1.8563
6	0.0500	1.78	20.00	0.0380	-0.1681
7	0.0583	1.59	17.14	0.1000	-0.1284
8	0.0667	5.80	15.00	0.1091	0.2912
9	0.0750	1.57	13.33	-0.1040	0.1242
10	0.0833	99.32	12.00	-1.2303	-0.3764
11	0.0917	4.82	10.90	0.2765	-0.0619
12	0.1000	3.37	10.00	0.1599	0.1749

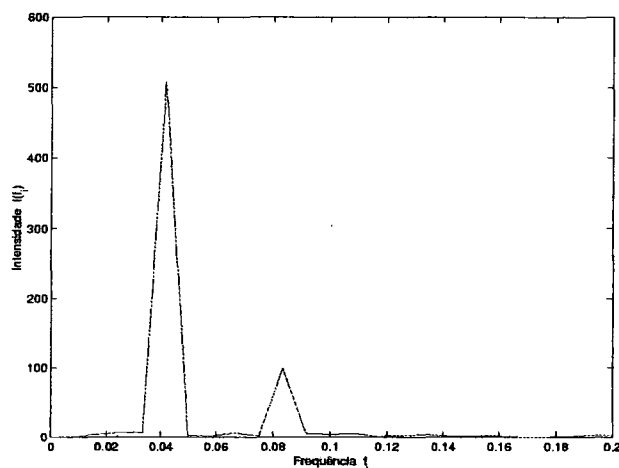


Figura 26 Periodograma da série de carga horária

Usando a Tabela 13, a sazonalidade pode ser modelada como:

$$\begin{aligned} \hat{S}_t &= \hat{\alpha}_0 + \hat{\alpha}_5 \sin(2\pi f_5 t) + \hat{\alpha}_{10} \sin(2\pi f_{10} t) + \\ &+ \hat{\beta}_5 \cos(2\pi f_5 t) + \hat{\beta}_{10} \cos(2\pi f_{10} t) \\ \hat{S}_t &= 142.8116 - 2.2399 \sin(2\pi f_5 t) - 1.2303 \sin(2\pi f_{10} t) + \\ &- 1.8563 \cos(2\pi f_5 t) - 0.3764 \cos(2\pi f_{10} t) \end{aligned}$$

A série de carga horária e a sazonalidade ajustadas por uma série de Fourier são mostradas na Figura 27.

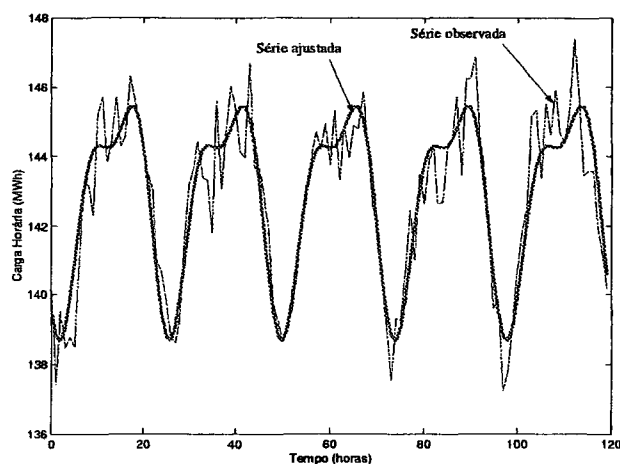


Figura 27 Série de carga horária e sazonalidade ajustada

A análise da série de carga horária fica completa, retirando-se a sazonalidade da série original, e analisando-se os resíduos para encontrar correlação serial. Esta análise é deixada como exercício.

**Exercício 60** *Cosidere os dados mensais de consumo de energia elétrica do setor rural dados na Tabela 11. Faça o ajuste da tendência linear e ajuste a sazonalidade por uma série de Fourier, em seguida analise os resíduos verificando se há correlação serial.*

### 4.3.3 Diferenciação discreta sazonal

Da mesma forma que foi feito com a tendência, a sazonalidade de uma série pode ser extraída por diferenciação discreta sazonal que denotaremos por  $\Delta_S$  onde  $S$  é o período da série. Desta forma definimos o operador diferença sazonal por:

$$\begin{aligned}\Delta_S X_t &= X_t - X_{t-S} \\ \Delta_S^2 X_t &= \Delta_S(\Delta_S X_t) = X_t - 2X_{t-S} + X_{t-2S} \\ &\vdots \\ \Delta_S^d X_t &= \Delta_S(\Delta_S^{d-1} X_t)\end{aligned}$$

A principal desvantagem dessa abordagem é que a série deve ser muito longa quando o período  $S$  é grande, por exemplo para a série de carga horária apresentada no exemplo anterior,  $S = 24$  horas e  $N = 120$  portanto só temos 5 períodos na série histórica a diferenciação resulta, neste caso em uma série com somente 5 observações, que é muito pouco para qualquer análise estatística dos resíduos. Além disso a diferenciação sazonal, como a diferenciação simples, não permite fazer previsão da sazonalidade. Por estas razões não vamos nos deter mais nesta técnica de extração da sazonalidade para séries de cargas.



# Capítulo 5

## Modelos Lineares Estacionários

Os modelos aqui apresentados são casos particulares de um modelo de filtro linear. Este modelo supõe que uma série temporal é gerada através de um sistema linear, cuja entrada é um ruído branco como é ilustrado na Figura 2.1. Denotamos por  $\Psi(B)$  a função de transferência do sistema, onde  $B$  é um operador retardo, definido por  $B^l a_t = a_{t-l}$ .

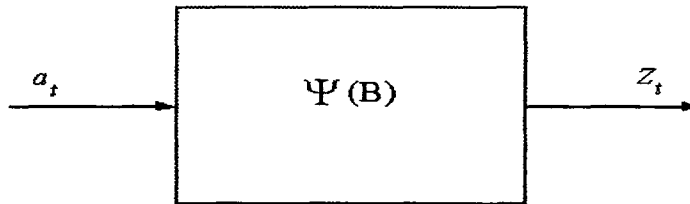


Figura 28 Filtro linear com entrada  $a_t$ , saída  $Z_t$

Desta forma podemos escrever que:

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots$$

denotando por:

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$$

temos:

$$Z_t = \mu + \psi(B)a_t \quad (5.1)$$

A condição de estacionariedade do modelo linear (5.1) está associado ao polinômio  $\psi(B)$ . Pode ser mostrado que a série  $Z_t = \psi(B)a_t$  é estacionária se o polinômio  $\psi(B)$  converge para  $|B| \leq 1$

Uma forma equivalente de representação de uma série temporal é escrevê-la como uma combinação linear de seus valores passados, ou seja:

$$Z_t = \mu + a_t - \pi_1 Z_{t-1} - \pi_2 Z_{t-2} - \dots$$

definindo o polinômio:

$$\pi(B) = 1 + \pi_1 B + \pi_2 B^2 + \dots$$

podemos escrever que:

$$\pi(B)Z_t = \mu + a_t \quad (5.2)$$

Com essa representação, temos associado a propriedade de inversibilidade, ou seja o modelo (5.2) é inversível quando o polinômio  $\pi(B)$  converge para todo  $|B| \leq 1$ . Os modelos (5.1) e (5.2) são equivalentes, se escolhermos  $\pi(B)$  e  $\psi(B)$  tal que:

$$\psi(B)\pi(B) = 1$$

Podemos eliminar o termo constante  $\mu$  sem perda de generalidade, para isso basta considerar a série  $\tilde{Z}_t = Z_t - E(Z_t)$ . Os três modelos de processos estocástico lineares mais usados para modelar séries temporais estacionárias, são os processos Auto-Regressivos (AR), Médias Móveis (MA) e os mixtos Auto-Regressivos-Médias Móveis (ARMA). Apresentamos a seguir as características de cada um desses modelos,

## 5.1 Processo AR( $p$ )

Uma série  $Z_t$ ,  $t = 1, 2, \dots$ , é um processo auto-regressivo AR( $p$ ) quando:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t \quad (5.3)$$

Definindo o polinômio de grau  $p$ ,  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ , podemos escrever modelo (5.3) como:

$$\phi(B)Z_t = a_t$$

Sendo  $\phi(B)$  um polinômio de grau finito, não há restrições sobre os parâmetros para assegurar a inversibilidade de  $Z_t$ . A condição de estacionariedade deste modelo pode ser analisada escrevendo o polinômio  $\phi(B)$  como:

$$\phi(B) = (1 - G_1B)(1 - G_2B) \dots (1 - G_pB) \quad (5.4)$$

onde,  $G_i^{-1}$ ,  $i = 1, \dots, p$ , são as raízes da equação característica. Expandindo (5.4) em frações parciais, temos:

$$\phi^{-1}(B) = \sum_{i=1}^p \frac{A_i}{1 - G_iB} \quad (5.5)$$

onde  $A_i$ ,  $i = 1, \dots, p$ , são constantes. e denotando por  $\psi(B) = \phi^{-1}(B)$  desde que a estacionariedade é assegurada se  $\psi(B)$  convergir para  $|B| \leq 1$ , esta condição é satisfeita para (5.5) se  $|G_i| \leq 1$ ,  $i = 1, \dots, p$ . Portanto asseguramos que um processo AR( $p$ ) é estacionário se as raízes do polinômio  $\phi(B) = 0$  estão fora do círculo de raio unitário,  $|G_i|^{-1} > 1$ ,  $i = 1, \dots, p$ . Para ilustrar, considere como exemplo o modelo AR( $p$ ) com  $p = 2$ .

**Exemplo 61** *Processo AR(2):*

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t \quad (5.6)$$

neste exemplo a condição de estacionariedade pode ser escrita como:

$$\begin{aligned} \phi_1 + \phi_2 &< 1 \\ \phi_2 - \phi_1 &< 1 \\ |\phi_2| &< 1 \end{aligned}$$

### 5.1.1 Função de Autocorrelação do AR( $p$ )

A função de autocorrelação (FAC) é uma ferramenta muito útil na identificação do modelo que gerou a série temporal. Para um processo AR( $p$ ) dado por (5.3) esta função pode ser calculada como:

$$\rho_k = \gamma_k / \gamma_0$$

onde  $\gamma_k = E(Z_t Z_{t-k})$  e  $\gamma_0 = E(Z_t^2)$ , é a variância de  $Z_t$ . Assim temos:

$$\gamma_k = E\{(a_t + \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p}) Z_{t-k}\} \quad (5.7)$$

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p} \quad (5.8)$$

Dividindo (5.8) por  $\gamma_0$  temos:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad (5.9)$$

A equação (5.9) é conhecida como equação de Yule-Walker e sendo  $G_1^{-1}, \dots, G_p^{-1}$  as raízes do polinômio  $\phi(B) = 0$ , podemos escrever a solução da equação (5.9), como:

$$\rho_k = C_1 G_1^k + C_2 G_2^k + \dots + C_p G_p^k \quad (5.10)$$

onde  $C_i$ ,  $i = 1, \dots, p$  são constantes. Se o modelo é estacionário então  $|G_i| < 1$ , conseqüentemente  $\rho_k$  decai exponencialmente com  $k$ . Se  $G_i^{-1}$  são raízes complexas conjugadas então  $G_i^{-1}$  podem ser escrito como  $|G_i| e^{i(2k\pi + f)}$ . Neste caso  $\rho_k$  decai exponencialmente oscilando.

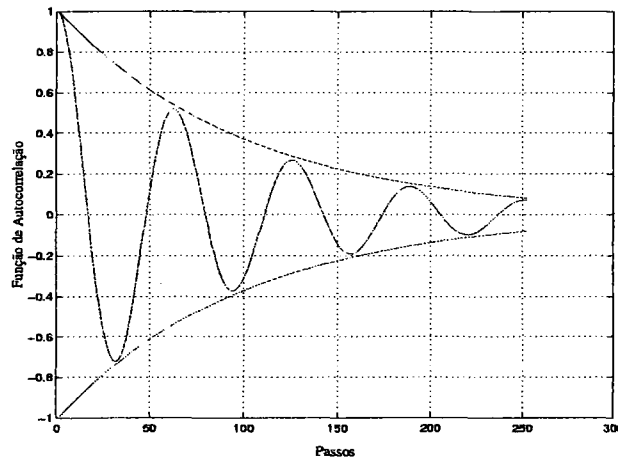


Figura 29 Função de Autocorrelação de um Processo  $AR(p)$

### 5.1.2 Função de Autocorrelação Parcial do $AR(p)$

Outra ferramenta utilizada na identificação dos modelos de séries temporais é a função de autocorrelação parcial (FACP). Esta função que denotamos por  $\phi_{kk}$  para  $k = 1, 2, \dots$ , é definida como:

**Definio 62** O coeficiente de autocorrelação parcial de ordem  $k$ ,  $\phi_{kk}$ , é o último coeficiente de um modelo  $AR(k)$ , ajustado a série temporal  $Z_t$ ,  $t = 1, 2, \dots$

Supondo-se que a série  $Z_t$ ,  $t = 1, 2, \dots$ , foi gerada por um modelo AR(1), então:

$$Z_t = \phi_1 Z_{t-1} + a_t \Rightarrow \phi_{11} = \phi_1$$

Supondo agora que  $Z_t$  foi gerada por um modelo AR(2), então:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t \Rightarrow \phi_{22} = \phi_2$$

Procedendo desta maneira, temos no caso geral, supor que  $Z_t$  foi gerada por um modelo AR( $k$ ), então:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_k Z_{t-k} + a_t \Rightarrow \phi_{kk} = \phi_k$$

Uma propriedade importante da FACP é que se a série  $Z_t$  é gerada por um modelo AR( $p$ ) temos  $\phi_{kk} = 0$  para  $k > p$ .

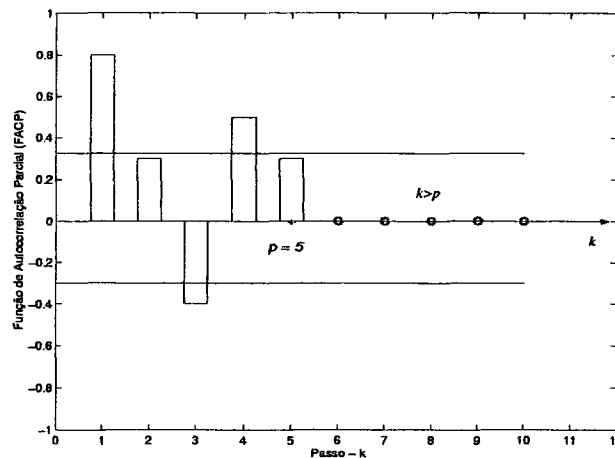


Figura 30 Função de Autocorrelação Parcial do Modelo AR( $p$ )

**Exemplo 63** Considere o processo AR(2) gerado pelo modelo  $Z_t = 0.8Z_{t-1} - 0.5Z_{t-2} + a_t$ , com  $a_t$  normal  $N(0, 1)$ , cujas observações são mostradas na Figura 31.

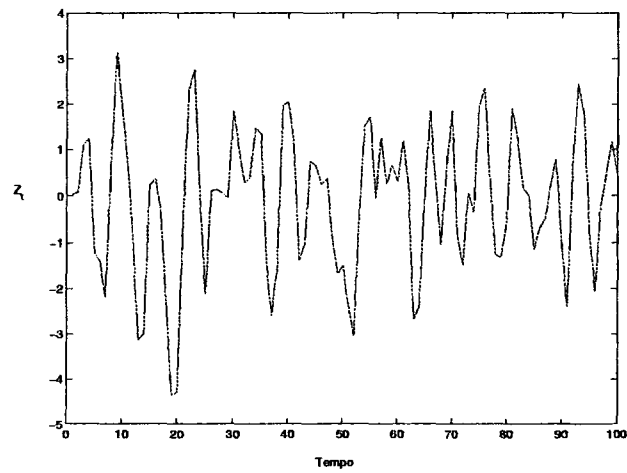


Figura 31 Processo gerado por um modelo  $AR(2)$

Os gráficos das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) mostrados na Figura 32 ilustram como essas funções podem ser usadas para identificação da ordem do modelo.

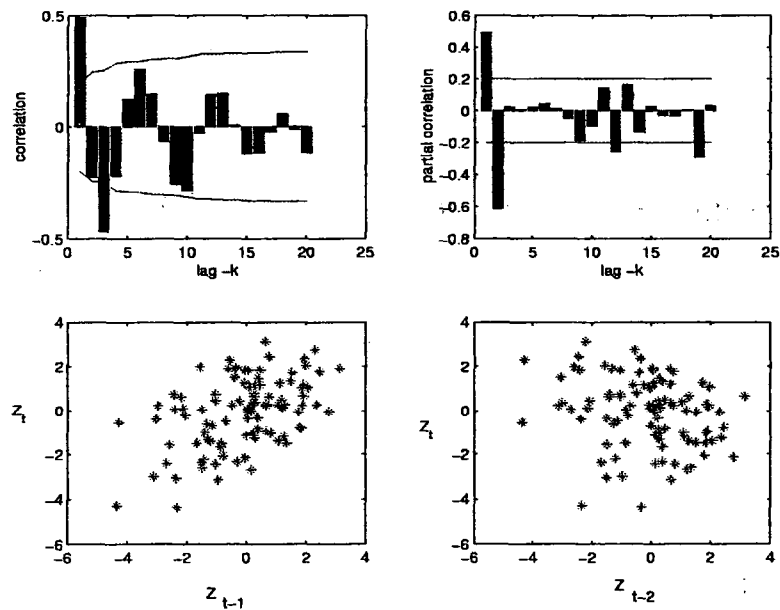


Figura 32 FAC e FACP do processo  $AR(2)$ .

## 5.2 Processo MA( $q$ )

Uma série  $Z_t$ ,  $t = 1, 2, \dots$ , é um processo de Médias Móveis de ordem  $q$ , MA( $q$ ), se satisfaz o seguinte modelo:

$$Z_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (5.11)$$

denotando por  $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ , podemos escrever esse modelo como:

$$Z_t = \theta(B)a_t \quad (5.12)$$

Comparando-se o modelo (5.2) com (5.3) temos,  $\psi(B) = \theta(B)$  e sendo  $\theta(B)$  um polinômio finito, então não há restrições sobre os parâmetros  $\theta_j$  para que  $\psi(B)$  convirja. Portanto todo processo MA( $q$ ) é estacionário. Podemos verificar que um processo MA( $q$ ) é inversível quando as raízes da equação característica  $\theta(B) = 0$  estão fora do círculo de raio unitário. Para ilustrar esse modelo, vamos considerar como exemplo  $q = 2$ . Temos:

**Exemplo 64** *Processo MA(2):*

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \quad (5.13)$$

neste caso as condições de inversibilidade são:

$$\begin{aligned} \theta_1 + \theta_2 &< 1 \\ \theta_2 - \theta_1 &< 1 \\ |\theta_2| &< 1 \end{aligned}$$

### 5.2.1 Função de Autocorrelação do MA( $q$ )

A função de autocorrelação (FAC) de um processo MA( $q$ ) pode ser calculada por:

$$\rho_k = \gamma_k / \gamma_0 \quad (5.14)$$

sendo  $\gamma_k = E(Z_t Z_{t-k})$ , temos:

$$\gamma_k = E\{(a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q})(a_{t-k} - \theta_1 a_{t-k-1} - \dots - \theta_q a_{t-k-q})\}$$

sendo  $E\{a_t a_{t-i}\} = 0$  se  $i \neq 0$  e  $E\{a_t^2\} = \sigma_a^2$ , temos:

$$\gamma_k = \begin{cases} (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_a^2 & k = 0 \\ (-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q) \sigma_a^2 & k = 1, 2, \dots, q \\ 0 & k > q \end{cases} \quad (5.15)$$

sendo  $\gamma_0$  a variância do processo temos:

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2} & k = 1, 2, \dots, q \\ 0 & k > q \end{cases} \quad (5.16)$$

Uma característica da FAC de um processo MA( $q$ ) é que ela se anula após a ordem  $q$  do modelo.

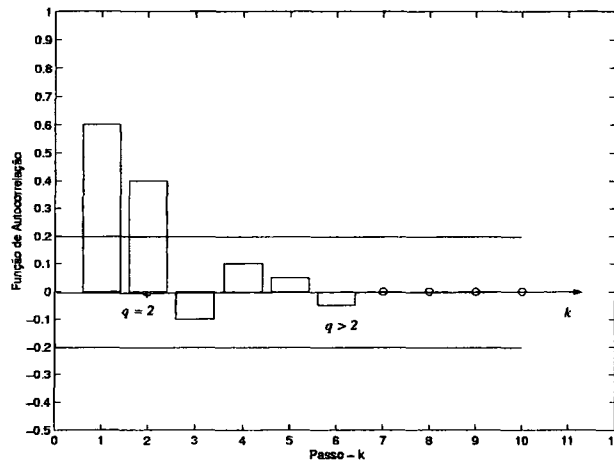


Figura 33 Função de Autocorrelação do Modelo MA( $q$ )

### 5.2.2 Função de Autocorrelação Parcial do MA( $q$ )

Sendo um processo MA( $q$ ) equivalente a um processo AR( $\infty$ ), a função de autocorrelação parcial (FACP) para processos MA( $q$ ) tem o comportamento que assintoticamente decai exponencialmente ou oscilando quando  $k \rightarrow \infty$ , semelhante a forma da função de autocorrelação de um processo AR( $p$ ), como mostra a Figura 28.

**Exemplo 65** Um processo MA(2) gerado pelo modelo  $Z_t = a_t + 0.8a_{t-1} -$



$0.5a_{t-2}$ , com  $a_t$  um ruído branco  $N(0, 1)$  é mostrado na Figura 34.

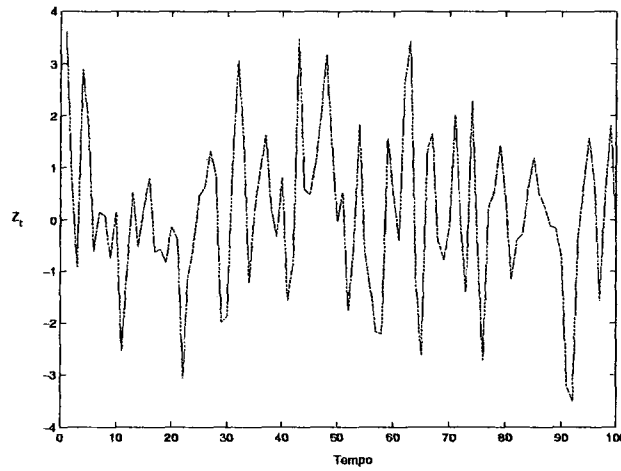


Figura 34 Processo gerado por um modelo MA(2)

A FAC e FACP para esse processo são apresentados na Figura 35 e ilustram como essas funções podem ser usadas para estimar a ordem do modelo.

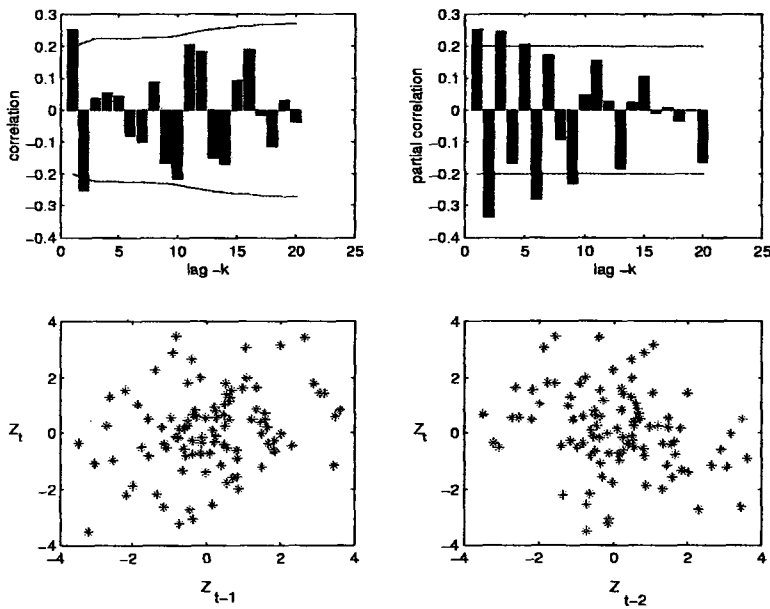


Figura 35 FAC e FACP do processo MA(2)

### 5.3 Processo ARMA( $p, q$ )

Uma série  $Z_t$ ,  $t = 1, 2, \dots$  gerada por um modelo misto Auto-Regressivo-Médias -Móveis, ARMA( $p, q$ ) é descrita como:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (5.17)$$

Se  $\phi(B)$  e  $\theta(B)$  são os polinômios dos modelos auto-regressivos e de médias-móveis, respectivamente, introduzidos anteriormente, podemos escrever (5.17) como:

$$\phi(B)Z_t = \theta(B)a_t \quad (5.18)$$

Os processos ARMA( $p, q$ ) são estacionários se as raízes de  $\phi(B) = 0$  estão todas fora do círculo de raio unitário e serão inversíveis se todas as raízes de  $\theta(B) = 0$  estiverem fora do círculo de raio unitário. como exemplo considere o processo ARMA(1,1).

**Exemplo 66** *Processo ARMA(1, 1).*

$$Z_t = \phi_1 Z_{t-1} + a_t - \theta_1 a_{t-1} \quad (5.19)$$

*neste caso temos a região de permissividade (estacionariedade e inversibilidade) definidas pelas regiões  $|\phi_1| < 1$  e  $|\theta_1| < 1$ .*

#### 5.3.1 Função de Autocorrelação do ARMA( $p, q$ )

A função de autocorrelação dos modelos ARMA( $p, q$ ), pode ser calculada como:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (5.20)$$

onde,  $\gamma_k = E(Z_t Z_{t-k})$  é dado por:

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} + \gamma_{za}(k) - \theta_1 \gamma_{za}(k-1) - \dots - \theta_q \gamma_{za}(k-q) \quad (5.21)$$

onde,  $\gamma_{za}(k-j) = E(Z_{t-k} a_{t-j})$  sendo  $\gamma_{za}(k) = 0$  para  $k > 0$  e  $\gamma_{za}(k) \neq 0$  para  $k \leq 0$ . Portanto quando  $k > q$  todos os termos  $\gamma_{za}$  se anulam e a FAC do processo ARMA se reduz a FAC da parte AR do processo, ou seja:

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p}, \quad k > q \quad (5.22)$$

da mesma forma como já foi feito anteriormente temos:

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p}, \quad k > q \quad (5.23)$$

No caso dos processos ARMA( $p, q$ ) devemos observar alguns pontos característicos da função de autocorrelação:

- i) A FAC de um modelo ARMA( $p, q$ ) é uma mistura de exponenciais e/ou senóides amortecidas a partir do passo  $q - p + 1$  e as raízes de  $\phi(B) = 0$  determinam esse comportamento.
- ii) Os  $q$ -primeiros valores  $\rho_1, \dots, \rho_q$  dependem tanto dos parâmetros da parte AR como dos parâmetros da parte MA. Além disso as  $p$ -autocorrelações  $\rho_{q-p+1}, \dots, \rho_q$  servem como condições iniciais para a solução da equação  $\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p}$  quando  $k > q$ .

### 5.3.2 Função de Autocorrelação Parcial do ARMA( $p, q$ )

A função de autocorrelação parcial (FACP) do modelo ARMA( $p, q$ ) também tem um comportamento que decai assintoticamente exponencial e/ou oscilando quando  $k \rightarrow \infty$ .

**Nota 67** *A equivalência entre os modelos AR, MA e ARMA:*

*Um modelo AR(1) pode ser escrito como um modelo MA( $\infty$ ). Essa propriedade é mostrada facilmente, considerando o modelo AR(1)  $\phi(B)Z_t = a_t$ , com  $\phi(B) = (1 - \phi B)$ . Escrevendo esse modelo como um MA, temos:*

$$Z_t = (1 - \phi B)^{-1} a_t$$

$$Z_t = a_t + \sum_{j=1}^{\infty} \phi^j B^j a_{t-j}$$

*Da mesma forma que um modelo MA(1) é equivalente a um AR( $\infty$ ), assim como um modelo ARMA(1,1) é pode ser escrito como um AR( $\infty$ ) ou como um MA( $\infty$ ). No entanto a escolha do modelo para representar a série temporal depende do uso que se quer fazer do modelo. Por exemplo, ajustar e fazer previsão com modelos AR em geral é muito mais simples que modelos MA ou ARMA mas, pode levar a imprecisão e um esforço computacional maior se for necessário operar com muitos parâmetros (ordem  $p$  muito grande). Por outro lado ajustar modelos MA e ARMA leva a solução de sistemas de*

equações não-lineares que exigem muito esforço computacional e só podem ser resolvidos com aproximação numérica e portanto também estão sujeitos a erros de aproximação. Portanto aqui vamos apresentar somente o ajuste de modelos AR os quais facilitam as técnicas de ajuste e o cálculo de previsões.

## 5.4 Ajuste de Modelo Linear Estacionário

Considere o modelos ARMA( $p, q$ ) e o residuo  $a_t$ , então podemos escrever:

$$a_t = Z_t - \sum_{j=1}^p \phi_j Z_{t-j} + \sum_{j=1}^q \theta_j a_{t-j} \quad (5.24)$$

Supondo  $a_t$  *i.i.d.*  $N(0, \tau)$  onde:  $\tau = 1/\sigma^2 > 0$ , e considere o vetor de observações de tamanho  $N$ , dado por  $\mathbf{D} = (Z_1, \dots, Z_N)'$  e denotamos o vetor de observações de tamanho  $N - p$  por  $\mathbf{Z} = (Z_{p+1}, \dots, Z_N)'$  os dados usados para ajustar o modelo e denotamos por:  $\Phi = (\phi_1, \dots, \phi_p)' \in R^p$  e  $\Theta = (\theta_1, \dots, \theta_q)' \in R^q$  o vetor de parâmetros. Então a função de verossimilhança, condicionada as  $p$  primeiras observações  $Z_1, \dots, Z_p$ , pode ser escrita como:

$$\mathcal{L}(\Phi, \Theta, \tau | \mathbf{Z}) \propto \tau^{\frac{N}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{t=p+1}^N a^2(t) \right\} \quad (5.25)$$

$$\mathcal{L}(\Phi, \Theta, \tau | \mathbf{Z}) \propto \tau^{\frac{N}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{t=p+1}^N \left( Z_t - \sum_{j=1}^p \phi_j Z_{t-j} + \sum_{j=1}^q \theta_j a_{t-j} \right)^2 \right\}$$

Aqui estamos considerando  $a_0 = \dots = a_{p+q} = 0$ . Usando a notação matricial:

$$\mathbf{a} = (a_{q+1}, \dots, a_N)'$$

$$\mathbf{X} = \begin{pmatrix} Z_{p+1} & Z_p & Z_{p-2} & \dots & Z_1 \\ Z_{p+2} & Z_{p+1} & Z_p & \dots & Z_2 \\ Z_{p+3} & Z_{p+2} & Z_{p+1} & \dots & Z_3 \\ \dots & \dots & \dots & \dots & \dots \\ Z_N & Z_{N-1} & Z_{N-2} & \dots & Z_N \end{pmatrix}_{N \times p}$$

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -\theta_1 & 1 & 0 & \dots & 0 \\ -\theta_2 & -\theta_1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}_{N \times N}$$

Podemos escrever o modelo ARMA( $p, q$ ) como:

$$\mathbf{Z} = \mathbf{X}\Phi + \mathbf{Ga} \quad (5.26)$$

Assim a função de verossimilhança pode ser escrita como:

$$\mathcal{L}(\Phi, \tau | \mathbf{Z}) \propto \tau^{N/2} |V|^{1/2} \exp\left\{-\frac{\tau}{2} (\mathbf{Z} - \mathbf{X}\Phi)' V (\mathbf{Z} - \mathbf{X}\Phi)\right\} \quad (5.27)$$

onde  $V = (GG')^{-1}$ . A relação (5.27) envolve os vetores de parâmetros  $(\Phi, \Theta)$ , em uma relação não linear e de difícil manuseio para encontrar os estimadores de máxima verossimilhança.

### 5.4.1 Ajustando um modelo AR( $p$ )

Para o caso particular dos modelos auto-regressivos AR( $p$ ), temos  $\theta_j = 0, j = 1, \dots, q$  e a matriz  $G = 1$ . Assim a função de máxima verossimilhança torna-se:

$$\mathcal{L}(\Phi, \tau | \mathbf{Z}) \propto \tau^{N/2} \exp\left\{-\frac{\tau}{2} (\mathbf{Z} - \mathbf{X}\Phi)' (\mathbf{Z} - \mathbf{X}\Phi)\right\} \quad (5.28)$$

Tomando-se o logaritmo de (5.28), temos:

$$l(\Phi | \mathbf{Z}) = \frac{N}{2} \log(\tau) - \frac{\tau}{2} (\mathbf{Z} - \mathbf{X}\Phi)' (\mathbf{Z} - \mathbf{X}\Phi) \quad (5.29)$$

podemos calcular os estimadores de máxima verossimilhança (EMV) para  $\phi_i, i = 1, \dots, p$  e  $\tau$ , derivando (5.29) com relação a  $\phi_i$  e  $\tau$ , e igualando a zero, onde temos:

$$\begin{aligned} \frac{\partial l(\Phi, \tau | \mathbf{Z})}{\partial \phi_i} &= -\tau \mathbf{X}' (\mathbf{Z} - \mathbf{X}\Phi) = 0, \quad i = 1, \dots, p \\ \frac{\partial l(\Phi, \tau | \mathbf{Z})}{\partial \tau} &= \frac{N}{2} \tau^{-1} - \frac{1}{2} (\mathbf{Z} - \mathbf{X}\Phi)' (\mathbf{Z} - \mathbf{X}\Phi) \end{aligned}$$

Os EMV são dados por:

$$\begin{aligned} \hat{\Phi} &= (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{Z}) \\ \hat{\tau}^{-1} &= \frac{1}{N} (\mathbf{Z} - \mathbf{X}\hat{\Phi})' (\mathbf{Z} - \mathbf{X}\hat{\Phi}) \end{aligned}$$

**Exemplo 68** Considerando os dados do processo  $AR(2)$  dado na Figura 31 temos as estimativas de máxima verossimilhança dos parâmetros dadas por:

$$\hat{\Phi} = \begin{pmatrix} 0.778 \\ -0.597 \end{pmatrix}$$

$$\hat{\tau}^{-1} = 1.193$$

O modelo ajustado para esta série é dado por:

$$Z_t = 0.778Z_{t-1} - 0.597Z_{t-2} + a_t$$

A figura 36 mostra a análise dos resíduos  $\{a_t, t \geq 0\}$ , calculado após o ajuste do modelo:

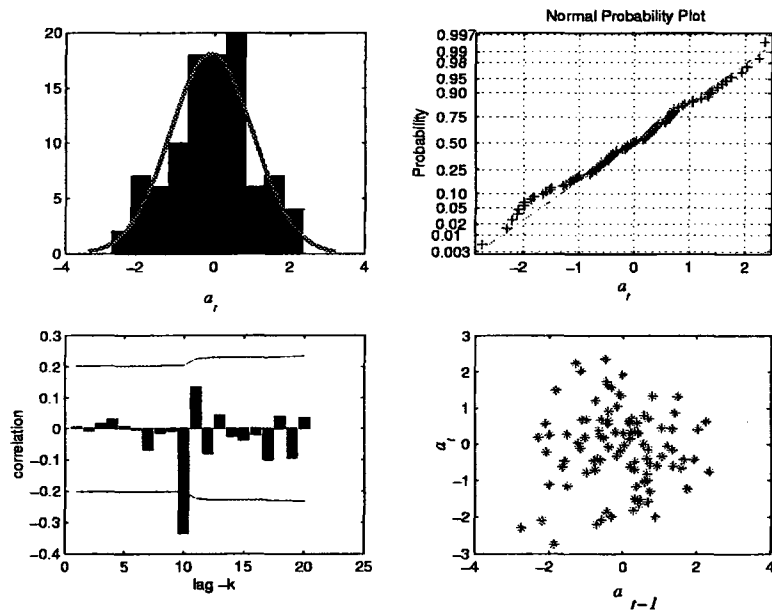


Figura 36 Análise dos resíduos do modelo ajustado

A análise dos resíduos mostra que o ajuste do modelo  $AR(2)$  eliminou a correlação temporal de ordem  $k < 10$  entre as observações da série e a correlação de ordem  $k = 10$  é pequena, não valendo a pena aumentar a complexidade do modelo para eliminá-la.

## 5.5 Critério de Seleção de Modelo

Identificar a ordem de um modelo auto-regressivo consiste em determinar o valor de  $p$ , ou seja determinar o número adequado de termos que devemos

considerar no modelo auto-regressivo. A escolha de um valor pequeno de  $p$ , por exemplo,  $p = 1$  pode levar a uma representação inadequada da série, por outro lado a escolha de um valor alto de  $p$ , por exemplo,  $p = 6$  pode levar a um modelo com um grau de complexidade desnecessário. O critério de escolha do modelo adequado com o menor número de parâmetro possível é denominado como critério de parcimoniosidade. Dois critérios de seleção de modelo, propostos na literatura que levam em conta o critério de parcimoniosidade são: o critério AIC, proposto por Akaike em 1973 e o critério de informação Bayesiano BIC, proposto por Schwarz em 1978, (ver [2]). De acordo com esses critérios escolhe-se o modelo  $AR(p)$  que minimize os valores dos critérios de AIC e BIC, dados por:

$$AIC(p) \simeq -\log \hat{\tau} + \frac{2(p+1)}{N}$$

$$BIC(p) \simeq -\log \hat{\tau} + (p+1) \frac{\log(N)}{N}$$

Onde  $\hat{\tau}$  denota o EMV de  $\tau$ .

**Exemplo 69** *Para exemplificar o uso do critério AIC e BIC, vamos considerar os dados do processo  $MA(2)$  dados na Figura 33, podemos ajustar modelos  $AR(1)$ ,  $AR(2)$  e  $AR(3)$  e escolher pelos critérios acima qual o modelo que melhor representa os dados. O resultado do ajuste é apresentado na Tabela 14*

Tabela 14 Critério de seleção de modelos  $AR(p)$

$p$	AIC	BIC
1	0.7188	0.7712
2	0.6271	0.7062
3	0.5896	0.6957
4	0.5478	0.6813
5	0.5427	0.7040
6	0.5017	0.6911
7	0.5064	0.7243
8	0.5360	0.7827

*Pelo critério de Akaike (AIC) seleciona-se o modelo de ordem  $p = 7$  mas pelo critério BIC seleciona-se o modelo de ordem  $p = 4$ . A análise dos resíduos desses dois modelos apresentada nas Figuras 37 e 38 mostram que*

ambos modelos apresentam resíduos com correlação serial não significativa. O critério BIC assegura mais a parcimônia do modelo que o AIC.

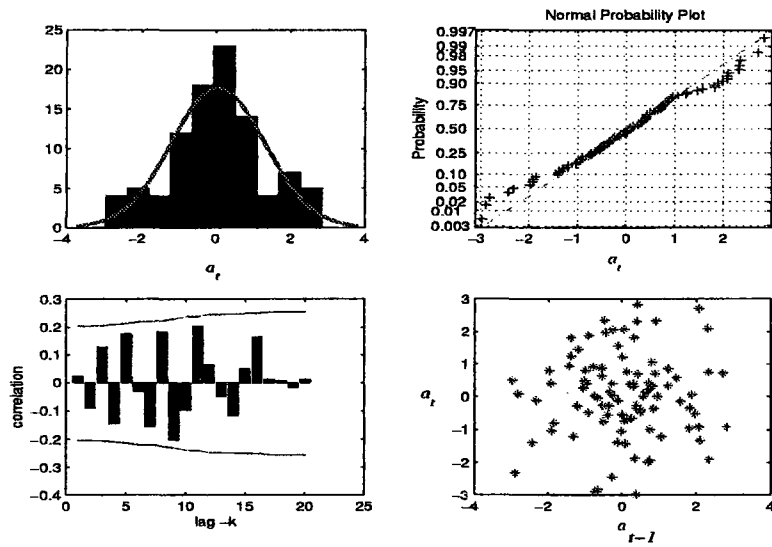


Figura 37 Análise dos resíduos para o modelo AR(4)

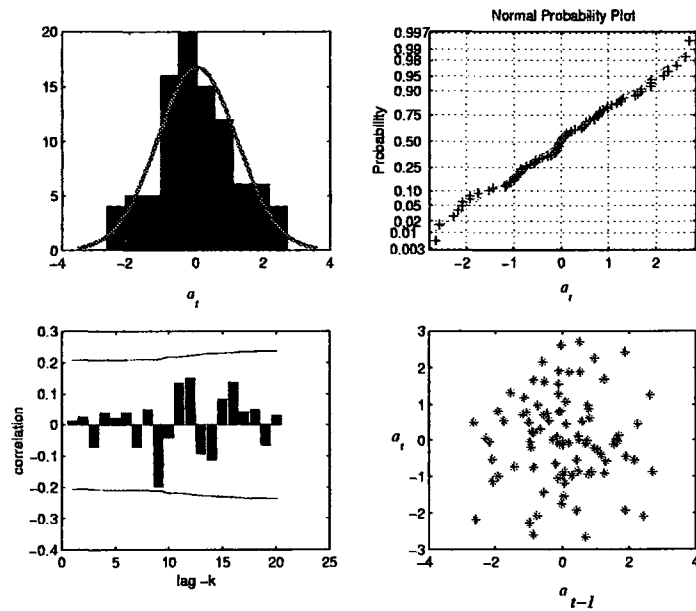


Figura 38 Análise dos resíduos para o modelo AR(7)



A Figura 39 mostra a análise dos resíduos para o modelo AR(2) esta análise mostra que esta ordem é visível mente insuficiente para expressar toda correlação da série temporal.

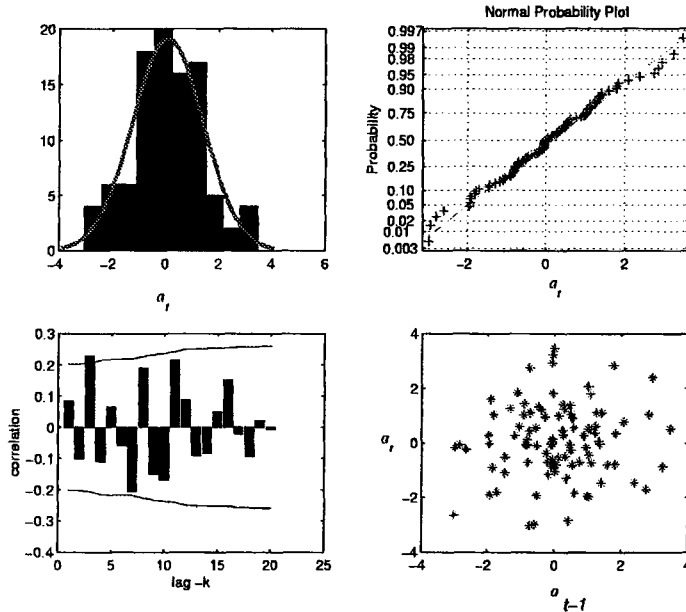


Figura 39 Análise dos resíduos para o modelo AR(2)

Portanto o modelo selecionado é AR(4) e os parâmetros ajustados são:

$$Z_t = 0.4701Z_{t-1} - 0.4852Z_{t-2} + 0.2701Z_{t-3} - 0.1463Z_{t-4} + a_t$$

com  $a_t$  um ruído branco normal com média zero e variância estimada em  $\sigma^2 = 1.558$ .

## 5.6 Previsão

Para fazer previsão de uma variável aleatória  $Y$  baseado na observação de outra variável  $X$  precisamos encontrar uma função  $g(X)$  que relacione a variável  $X$  com  $Y$ , de tal forma a previsão da variável  $Y$  dada a observação  $X = x$  pode ser calculada por  $Y = g(x)$ . Claramente gostaríamos de escolher a função  $g(x)$  tal que  $g(X)$  esteja o mais perto possível de  $Y$ . Um critério possível para escolha da função  $g$  é minimizar  $E\{[Y - g(X)]^2\}$ . Mostraremos que sob esse critério o melhor preditor possível é  $g(X) = E\{Y|X\}$ .

**Proposition 70**  $E\{[Y - g(X)]^2\} \geq E\{[Y - E(Y|X)]^2\}$

*Prova.* Para demonstrar essa proposição vamos considerar a igualdade:

$$\begin{aligned} E\{[Y - g(X)]^2|X\} &= E\{[Y - E(Y|X) + E(Y|X) - g(X)]^2|X\} \\ &= E\{[Y - E(Y|X)]^2|X\} + \\ &\quad + E\{[E(Y|X) - g(X)]^2|X\} + \\ &\quad + 2E\{[Y - E(Y|X)][E(Y|X) - g(X)]|X\} \end{aligned}$$

Dado  $X$ ,  $E(Y|X) - g(X)$  é uma função de  $X$  e pode ser tratado como uma constante, então:

$$\begin{aligned} E\{[Y - E(Y|X)][E(Y|X) - g(X)]|X\} &= [E(Y|X) - g(X)] \times \\ &\quad \times E\{[Y - E(Y|X)]|X\} \\ &= [E(Y|X) - g(X)] \times \\ &\quad \times [E(Y|X) - E(Y|X)] \\ &= 0 \end{aligned}$$

Sendo  $E\{[E(Y|X) - g(X)]^2|X\} > 0$ , temos que:

$$E\{[Y - g(X)]^2|X\} \geq E\{[Y - E(Y|X)]^2|X\} \quad (5.30)$$

Tomando o valor esperado com relação a  $X$  de ambos os lados de (5.30) e usando a propriedade do valor esperado condicional afirmando que  $E_A\{E_B(B|A)\} = E(B)$ , temos:

$$E\{[Y - g(X)]^2\} \geq E\{[Y - E(Y|X)]^2\}$$

■

### 5.6.1 Previsão com modelo ARMA( $p, q$ )

Agora considerando o modelo ARMA( $p, q$ ), onde fazemos  $Y = Z_t$  e  $X$  é o vetor de observações  $Z = (Z_1, \dots, Z_N)$ , e a função  $g$  que relaciona  $Z_t$  com o vetor de observações é o modelo ARMA( $p, q$ ),  $Z_t = a_t + \sum_{i=1}^p \phi_i Z_{t-i} + \sum_{j=1}^q \theta_j a_{t-j}$ . Então a previsão de um valor futuro  $Z_{N+h}$  para qualquer  $h \geq 1$  é estimada escrevendo:

$$Z_{t+h} = a_{t+h} + \sum_{i=1}^p \phi_i Z_{t+h-i} + \sum_{j=1}^q \theta_j a_{t+h-j}$$

calculado o valor esperado condicionado, temos:

$$E\{Z_{t+h}|\mathbf{Z}\} = E\{a_{t+h} + \sum_{i=1}^p \phi_i Z_{t+h-i} + \sum_{j=1}^q \theta_j a_{t+h-j} | \mathbf{Z}\}$$

$$E\{Z_{t+h}|\mathbf{Z}\} = E\{a_{t+h}|\mathbf{Z}\} + \sum_{i=1}^p \phi_i E\{Z_{t+h-i}|\mathbf{Z}\} + \sum_{j=1}^q \theta_j E\{a_{t+h-j}|\mathbf{Z}\}$$

Vamos denotar por  $\hat{Z}_{t+h} = E\{Z_{t+h}|\mathbf{Z}\}$  e  $\hat{a}_{t+h} = E\{a_{t+h}|\mathbf{Z}\}$  e utilizar o seguinte fato:

$$E\{Z_{t+h}|\mathbf{Z}\} = \begin{cases} Z_{t+h} & h \leq 0 \\ \hat{Z}_{t+h} & h > 0 \end{cases}$$

$$E\{a_{t+h}|\mathbf{Z}\} = \begin{cases} a_{t+h} & h \leq 0 \\ 0 & h > 0 \end{cases}$$

Os valores de  $Z_{t+h}$  e  $a_{t+h}$  para  $h < 0$  são extraídos do vetor de observações  $\mathbf{Z}$  e os valores de  $\hat{Z}_{t+h}$  são os valores previstos previamente. Para exemplificar como a previsão é calculada vamos considerar o caso particulares dos modelo AR( $p$ ).

### 5.6.2 Previsão com modelo AR( $p$ )

Neste caso o modelo é dado por:  $Z_{t+h} = a_{t+h} + \sum_{i=1}^p \phi_i Z_{t+h-i}$ , para  $h \geq 1$ . A previsão é dada por:

$$\hat{Z}_{t+h} = \sum_{i=1}^p \phi_i E\{Z_{t+h-i}|\mathbf{Z}\}$$

Pois sabemos que  $E\{a_{t+h}|\mathbf{Z}\} = 0$  para  $h \geq 1$  e considerando a relação:

$$E\{Z_{t+h-i}|\mathbf{Z}\} = \begin{cases} Z_{t+h-i} & h \leq i, i = 1, \dots, p \\ \hat{Z}_{t+h} & h > 1 \end{cases}$$

Podemos escrever a previsão para os seguintes passos  $h = 1, 2, \dots, p + 1$ :

$$\begin{aligned}
 h = 1 & \implies \widehat{Z}_{t+1} = \sum_{i=1}^p \phi_i Z_{t+1-i} \\
 & \widehat{Z}_{t+1} = \phi_1 Z_t + \dots + \phi_p Z_{t+1-p} \\
 h = 2 & \implies \widehat{Z}_{t+2} = \sum_{i=1}^p \phi_i Z_{t+2-i} \\
 & \widehat{Z}_{t+2} = \phi_1 \widehat{Z}_{t+1} + \dots + \phi_p Z_{t+2-p} \\
 & \vdots \quad \quad \quad \vdots \\
 h = p & \implies \widehat{Z}_{t+p} = \sum_{i=1}^p \phi_i Z_{t+p-i} \\
 & \widehat{Z}_{t+p} = \phi_1 \widehat{Z}_{t+p-1} + \dots + \phi_{p-1} \widehat{Z}_{t+1} + \phi_p Z_t \\
 h = p + 1 & \implies \widehat{Z}_{t+p+1} = \sum_{i=1}^p \phi_i Z_{t+p+1-i} \\
 & \widehat{Z}_{t+p+1} = \phi_1 \widehat{Z}_{t+p} + \dots + \phi_{p-1} \widehat{Z}_{t+2} + \phi_p \widehat{Z}_{t+1}
 \end{aligned}$$

Neste caso particular do modelo  $AR(p)$  notamos que as previsões para um passo  $h \geq p + 1$  são calculadas usando somente valores anteriormente previstos, ou seja já não são considerados dados observados. Este fato se reflete em um crescente do erro de previsão, pois introduz uma superposição de erros proveniente das previsões anteriores. Para enfatizar essa limitação dos modelos, vamos supor que  $p = 2$  (modelo  $AR(2)$ ). Neste caso somente as duas últimas observações são consideradas no cálculo da previsão. As previsões para um horizonte  $h > 2$  só levarão em conta valores anteriormente previstos. ( $\widehat{Z}_{t+h-1}, \widehat{Z}_{t+h-2}$ ). No entanto em muitos problemas o horizonte  $h = 2$  é extremamente pequeno.

A ordem do modelo reflete a memória do processo, portanto um modelo  $AR(1)$  conhecido como modelo Markoviano tem como propriedade a falta de memória, ou seja o futuro ( $Z_{t+1} = a_t + \phi_1 Z_t$ ) só depende do presente ( $Z_t$ ) não dependendo de qualquer informação passada. Processos com “*pouca memória*” são de difícil previsão para horizontes grandes.

Considerando o modelo  $AR(2)$  ajustados no Exemplo anterior, temos:

$$\widehat{\Phi} = \begin{pmatrix} 0.778 \\ -0.597 \end{pmatrix}$$

A Figura 40 mostra as últimas 10 observações da série juntamente com a previsão para  $t = 91, \dots, 100$ , considerando que as novas observações são incorporadas aos dados a cada passo (previsão 1-passo a frente) e a cada

três passos (previsão **3**-passos a frente).

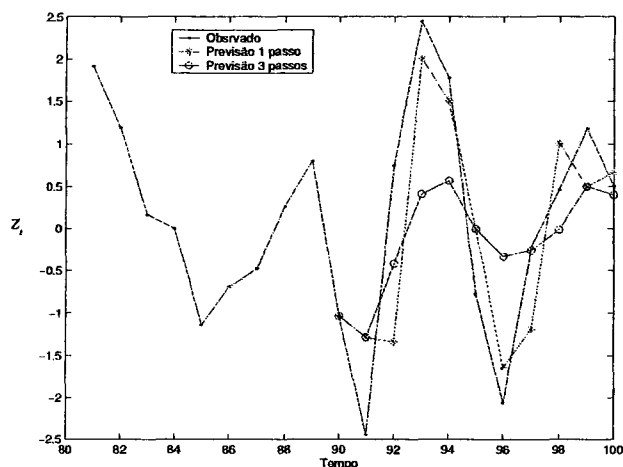


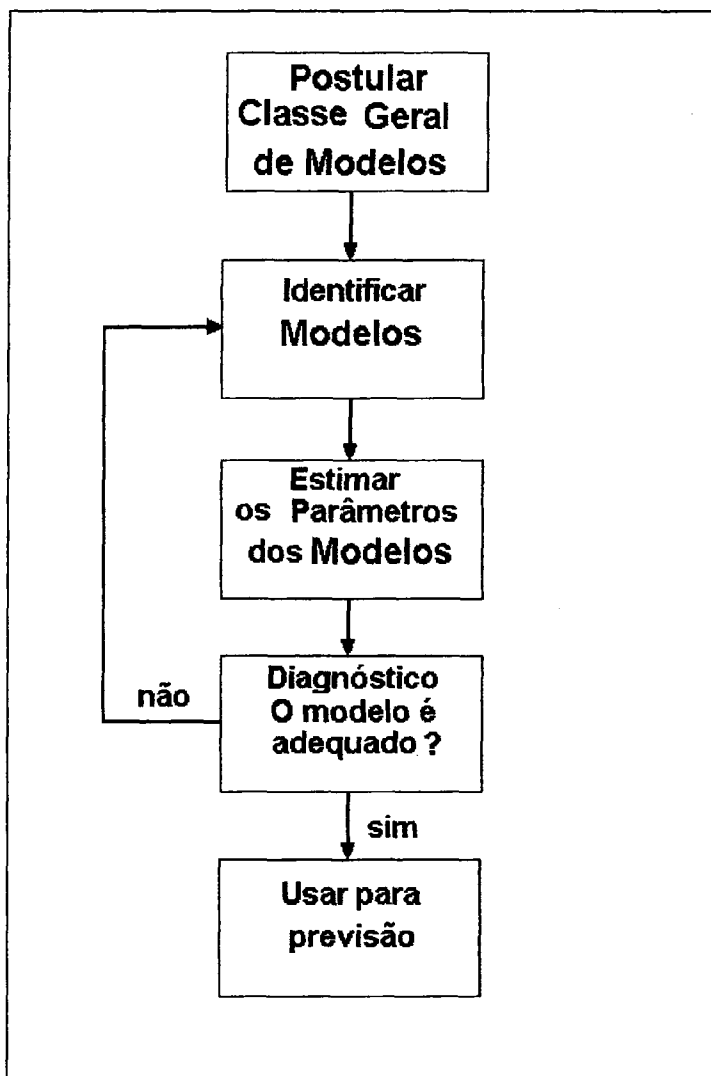
Figura 40 Dados observados e Previsão com modelo AR(2)

Notamos neste exemplo como a previsão se aproxima da média da série quando o horizonte de previsão aumenta refletindo a perda de memória do modelo.

## 5.7 As Etapas do Ajuste de um modelo linear

O uso teórico da FAC e FACP para identificar o modelo que descreve uma série temporal nem sempre pode ser seguido na prática, pois a estimativa destas funções está sempre sujeita as variabilidades amostrais que podem levar a escolha do modelo errado. Por isso um procedimento iterativo de identificação, estimativa e diagnóstico de validação para um modelo escolhido é sempre recomendado. O diagrama apresentado a seguir sumariza o procedimento iterativo que deve ser seguido para ajustar um modelo para

análise e previsão de uma séries temporal.



O bloco correspondente ao diagnóstico inclui testes de hipótese sobre os parâmetros, o critério de seleção de modelo e a análise dos resíduos para verificar se não há mais correlação serial.

## Capítulo 6

# Um Exemplo de Previsão de Carga.

Neste capítulo vamos apresentar um exemplo de modelagem, análise e previsão de carga, aqui vamos considerar uma série de consumo mensal de energia no setor industrial de uma região do estado de São Paulo no período de 1975 a 1986. O objetivo aqui identificar a tendência e sazonalidade da série e analisar a componente estocástica desta série através de sua função de autocorrelação (FAC) e auto correlação parcial (FACP). Ajustando um modelo autorregressivo usado para prever esta componente visando diminuir os erros de previsão. Também é feita uma avaliação da previsão através dos erros quadrático médio (EQM) e erros absolutos percentuais médios (EAPM) para diferentes procedimento de atualização das observações e horizontes de previsão necessários para atender os objetivos da operação e planejamento dos sistemas geradores e distribuidores de energia elétrica.

### 6.1 Consumo Industrial de Energia Elétrica

A Tabela 15 apresenta uma série de consumo mensal de energia elétrica no período de 1975 a 1986 no setor industrial de uma região do estado de

São Paulo. Essa série é apresentada de forma gráfica na Figura 41.

Tabela 15 Consumo de Energia elétrica no Setor Industrial em MW-hora

Mês	75	76	77	78	79	80	81	82	83	84	85	86
jan	138	155	170	209	227	240	305	262	276	306	360	393
fev	131	152	177	201	205	243	270	264	290	333	400	409
mar	150	164	182	214	240	259	318	268	282	332	373	344
abr	134	177	187	219	242	257	258	303	339	344	399	370
mai	149	173	195	214	230	274	284	286	300	344	415	411
jun	160	190	217	250	257	295	310	340	359	407	457	449
jul	167	211	243	276	299	308	318	337	350	395	392	449
ago	185	225	252	247	275	324	325	361	360	441	479	506
set	190	219	254	273	297	345	336	368	390	464	466	451
out	186	224	244	280	294	312	306	333	366	433	438	469
nov	175	245	243	256	288	333	302	347	366	459	488	486
dez	165	202	258	263	299	309	289	314	332	398	427	406

Uma simples análise visual do gráfico da A Figura 21 mostra que essa série tem ma tendência linear e uma sazonalidade.

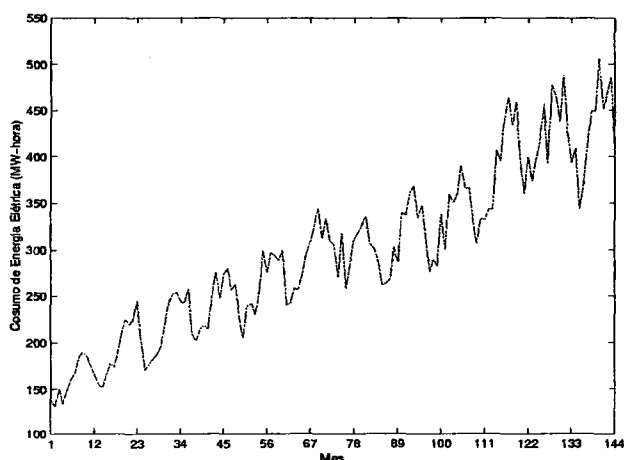


Figura 41 Consumo de Energia elétrica do setor industrial de 1975 a 1986

Vamos considerar o modelo aditivo para esta série de consumo, assim temos:

$$X_t = T_t + S_t + Z_t$$



### 6.1.1 A Tendência

A justando uma reta para modelar a tendência linear, temos:

$$T_t = \alpha_0 + \alpha_1 t$$

As estimativas de mínimos quadrados para  $\alpha_0, \alpha_1$  são:

$$\hat{\alpha}_1 = \frac{\sum_{t=1}^N tX_t - \frac{1}{N} \left( \sum_{t=1}^N t \right) \left( \sum_{t=1}^N X_t \right)}{\sum_{t=1}^N t^2 - \frac{1}{N} \left( \sum_{t=1}^N t \right)^2}$$

$$\hat{\alpha}_0 = \bar{X}_t - \hat{\alpha}_1 \bar{t}$$

onde:

$$\bar{X}_t = \frac{1}{N} \sum_{t=1}^N X_t \quad \text{e} \quad \bar{t} = \frac{1}{N} \sum_{t=1}^N t = \frac{(N+1)}{2}$$

temos:

$$\hat{\alpha}_0 = 149.73 \quad \text{e} \quad \hat{\alpha}_1 = 2.0296$$

O ajuste desta tendência juntamente com a série original são mostrados na Figura 42.

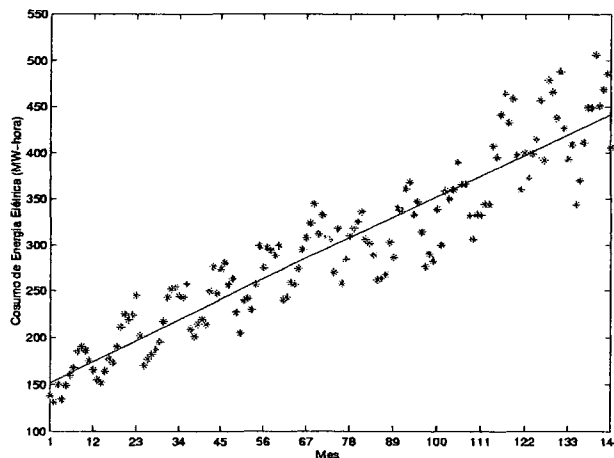


Figura 42 Série consumo observada e tendência estimada

A tendência ajustada pode ser utilizada para estimar a taxa de variação média anual no consumo de energia elétrica deste setor. Essa taxa é calculada como:

$$R_{k+1} = \frac{\bar{X}_{k+1} - \bar{X}_k}{\bar{X}_k} \times 100\%, \quad k = 1, 2, \dots, n-1$$

onde  $\bar{X}_k, k = 1, 2, \dots, n - 1$ , são as médias de consumo anuais. Considerando o período da carga como  $T = 12$  meses podemos calcular  $\bar{X}_k$ , considerando o ajuste linear, como:

$$\bar{X}_k = \frac{1}{T} \int_{(k-1)T}^{kT} (\alpha_0 + \alpha_1 t) dt, \quad k = 1, \dots, n - 1$$

$$\bar{X}_k = \alpha_0 + \alpha_1 \frac{(2k + 1)T}{2}, \quad k = 1, \dots, n - 1$$

Este resultado teórico (calculado com base no ajuste linear) pode ser comparado a posteriori com as taxa observadas, calculadas com as médias amostrais, dadas por:

$$\hat{R}_{k+1} = \frac{\hat{X}_{k+1} - \hat{X}_k}{\hat{X}_k} \times 100\%, \quad k = 1, 2, \dots, n - 1$$

onde as médias amostrais anuais,  $\hat{X}_k, k = 1, 2, \dots, n - 1$ , são calculadas com os dados históricos observados, por:

$$\hat{X}_k = \frac{1}{T} \sum_{t=(k-1)T+1}^{kT} X_t, \quad k = 1, 2, \dots, n$$

Fazendo o cálculo dessa taxa os dados da série de consumo no setor industrial, onde  $k = 1, 2, \dots, 12$  (ou seja para os 12 anos do histórico), podemos comparar a taxa estimada com a taxa observada como mostra a Figura 43

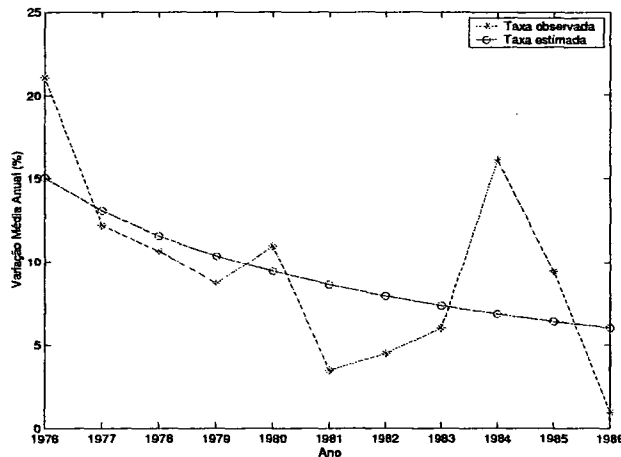


Figura 43 Taxa de variação no consumo médio anual

Podemos observar portanto uma taxa de variação média anual decrescente, caído de 15% em 1976 para 5% em 1986, indicando provavelmente uma retração no setor ou uma recessão econômica.

### 6.1.2 A Sazonalidade

Usando-se a tendência ajustada para eliminar o crescimento na série observada, temos uma série livre de tendência. A série livre de tendência é usada para ajustar a sazonalidade do consumo de energia. O cálculo do periodograma mostra que a frequência de maior intensidade é a de período  $1/f_k = 12$  meses, este resultado está mostrado na Figura 43.

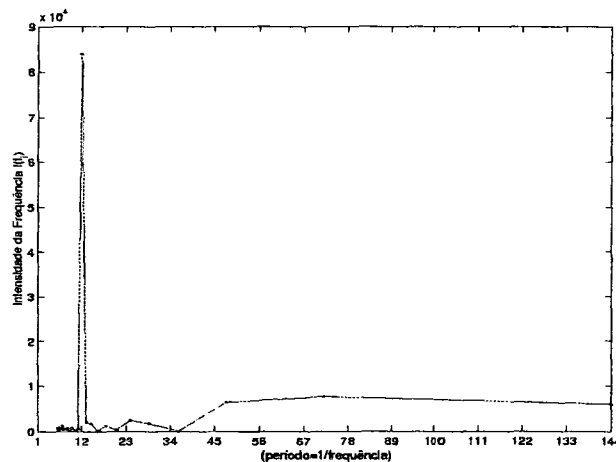


Figura 43 Periodograma da componente sazonal da série de consumo

O ajuste da série de Fourier, leva a seguinte modelo para a sazonalidade:

$$S_t = a_0 + a_{12} \sin(2\pi f_{12}t) + b_{12} \cos(2\pi f_{12}t)$$

com as seguintes estimativas:

$$\hat{a}_0 = 0, \hat{a}_{12} = -33.9167, \hat{b}_{12} = -4.0533$$

A série livre de tendência e a sazonalidade ajustada por série de Fourier, são mostradas na Figura 44.

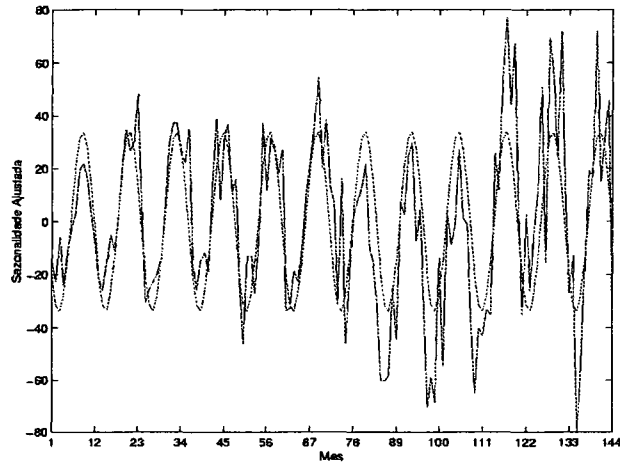


Figura 44 Ajuste da componente sazonal da série de consumo

### 6.1.3 A Componente Estocástica

A análise dos resíduos, ou componente estocástica da série pode ser feita eliminandose da série original a tendêncai e a sazonalidade, assim a série de compnente estocástica pode ser calculada por:

$$Z_t = X_t - T_t - S_t, t = 1, \dots, 144$$

O gráfico deste processo  $\{Z_t, t = 1, \dots, 144\}$ , e suas funções de autocorrelação (FAC) e autocorrelação parcial (FACP) são apresentados nas Figura 45.e 46

respectivamente.

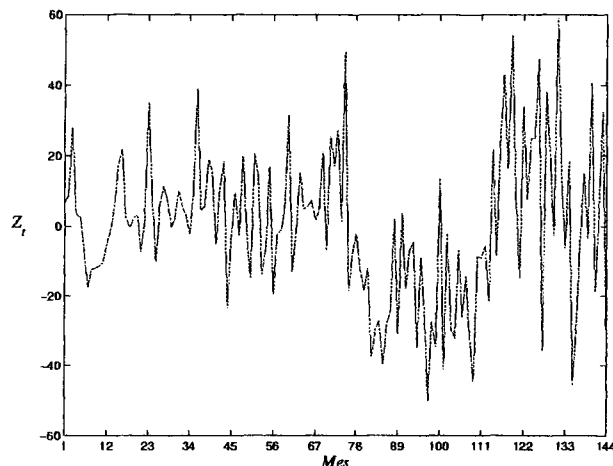


Figura 45 Componente estocástica  $\{Z_t, t = 1, \dots, 144\}$  da série de consumo  
A análise da FAC e FACP do processo  $Z_t$ , indicam que um modelo AR(3) pode representar bem o processo.

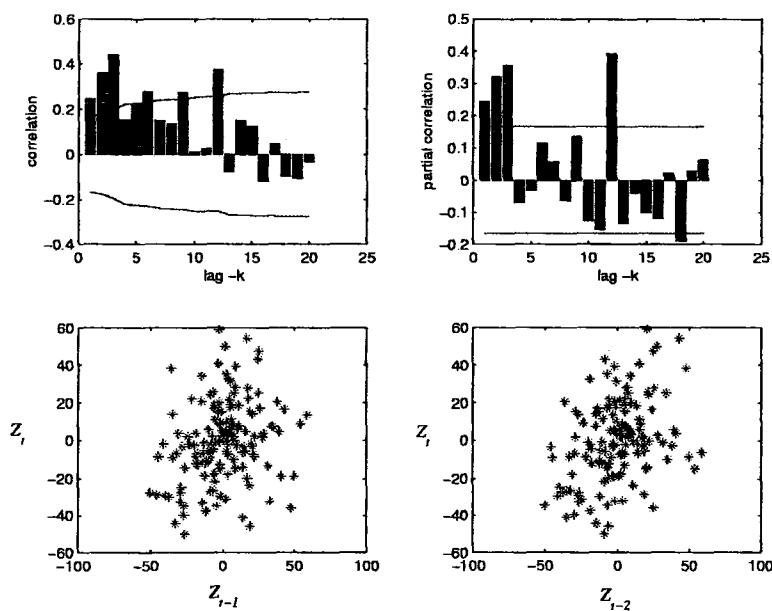


Figura 46 FAC e FACP de  $\{Z_t, t = 1, \dots, 144\}$

Ajustando-se modelos AR( $p$ ) com  $p = 1, 2, 3, 4$  e usando o cálculo dos critérios  $AIC$  e  $BIC$ , mostrados na Tabela 16, podemos confirmar a escolha do modelo

AR(3) para representar a estocasticidade residual da série.

Tabela 16 Valores de AIC e BIC

$p$	AIC	BIC
1	6.0915	6.1330
2	6.0053	6.0678
3	5.8812	5.9648
4	5.8964	6.0015

O ajuste do modelo AR(3) leva ao seguinte modelo:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + a_t$$

com

$$\phi_1 = 0.0436, \phi_2 = 0.2576, \phi_3 = 0.3608$$

A análise dos resíduos mostrados nos gráficos da Figura 47 assegura que este modelo é adequado para representar a componente  $Z_t$ .

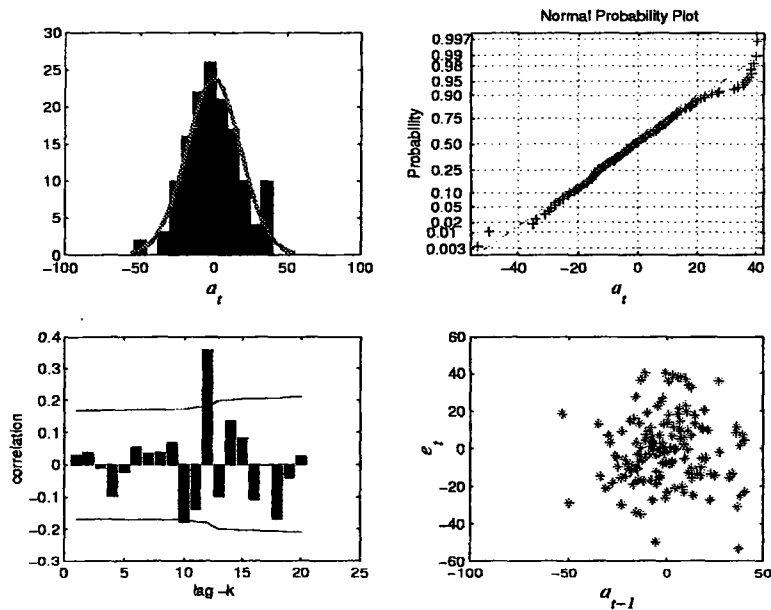


Figura 47 Análise dos resíduos do modelo AR(3)

O modelo completo para a série de consumo de energia elétrica no setor industrial, ajustado segundo os dados da Tabela 15 é:

$$X_t = T_t + S_t + Z_t$$

com:

$$T_t = 149.73 + 2.0296t$$

$$S_t = -33.917 \sin\left(\frac{2\pi t}{12}\right) - 4.053 \cos\left(\frac{2\pi t}{12}\right)$$

$$Z_t = 0.0436Z_{t-1} + 0.258Z_{t-2} + 0.361Z_{t-3} + a_t$$

#### 6.1.4 A Previsão

Para fazer um exercício de previsão com os dados do consumo de energia elétrica dados na Tabela 15, vamos deixar os últimos 12 meses fora do ajuste do modelo, em seguida faremos a previsão destes consumos, ou seja a previsão do consumo para 12 meses então podemos comparar esta previsão com os valores observados, deixados fora do ajuste. O modelo ajustado com as 132 observações é:

$$X_t = T_t + S_t + Z_t$$

$$T_t = 149.58 + 2.0356t$$

$$S_t = -32.483 \sin\left(\frac{2\pi t}{12}\right) - 3.855 \cos\left(\frac{2\pi t}{12}\right)$$

$$Z_t = 0.0544Z_{t-1} + 0.305Z_{t-2} + 0.378Z_{t-3} + a_t$$

A análise com as 132 observações é feita de forma semelhante a análise feita com os dados completos, ocorrendo somente uma pequena variação amostral nas estimativas dos parâmetros dos modelos devido a retirada dos últimos 12 meses. A Figura 48 apresenta os dados observados e os ajustes da tendência e sazonalidade  $T_t + S_t$ , para os dados observados e a extrapolação para os 12

meses de 1986.

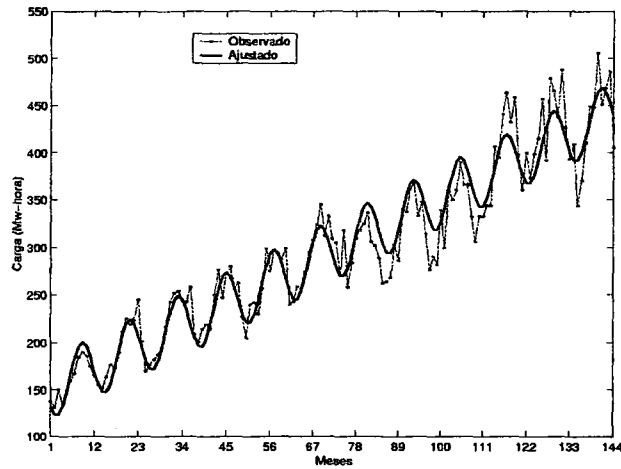


Figura 48 Ajuste e extrapolação da tendência e sazonalidade da série

A previsão para os valores de  $t = 133, \dots, 144$ . é calculada extrapolando a tendência e a sazonalidade para este horizonte e calculando-se o valor esperado condicionado da componente estocástica, como se segue:

$$\hat{Z}_{t+h} = 0.0544\hat{Z}_{t+h-1} + 0.305\hat{Z}_{t+h-2} + 0.378\hat{Z}_{t+h-3}$$

onde temos que  $\hat{Z}_{t+h-i} = Z_{t+h-i}$ ,  $h \leq i$ ,  $i = 1, 2, 3$ , assim temos que:

$$h = 1 \implies \hat{Z}_{t+1} = 0.0544Z_t + 0.305Z_{t-1} + 0.378Z_{t-2}$$

$$h = 2 \implies \hat{Z}_{t+2} = 0.0544\hat{Z}_{t+1} + 0.305Z_t + 0.378Z_{t-1}$$

$$h = 3 \implies \hat{Z}_{t+3} = 0.0544\hat{Z}_{t+2} + 0.305\hat{Z}_{t+1} + 0.378Z_t$$

Para  $h > 3$  usa-se os valores previstos recursivamente, ou seja, para  $h = 4, 5, \dots, 12$ , temos:

$$\hat{Z}_{t+4} = 0.0544\hat{Z}_{t+3} + 0.305\hat{Z}_{t+2} + 0.378\hat{Z}_{t+1}$$

$$\hat{Z}_{t+5} = 0.0544\hat{Z}_{t+4} + 0.305\hat{Z}_{t+3} + 0.378\hat{Z}_{t+2}$$

$\vdots = \vdots$

$$\hat{Z}_{t+12} = 0.0544\hat{Z}_{t+11} + 0.305\hat{Z}_{t+10} + 0.378\hat{Z}_{t+9}$$

A previsão do consumo portanto é calculada como:

$$\hat{X}_{t+h} = T_{t+h} + S_{t+h} + \hat{Z}_{t+h}$$



Os resultados dessa previsão juntamente com os valores observados para os 12 meses do ano de 1986, são apresentados graficamente na Figura 49.

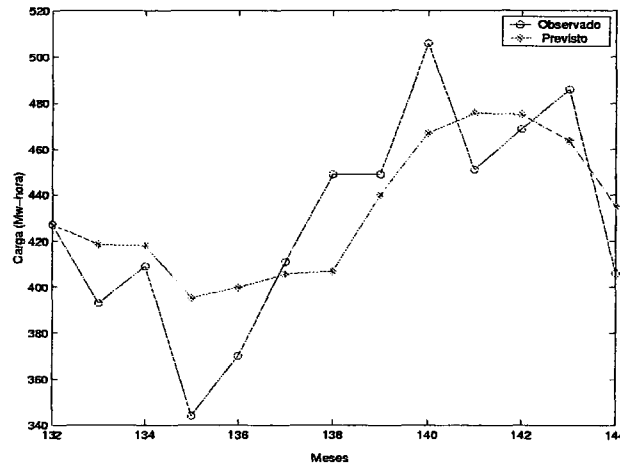


Figura 49 Gráfico dos valores previstos e observados para 1986

A avaliação desta previsão é feita com o cálculo do erro quadráticos médios e erro absoluto percentual médio, entre os valores previstos e observados. previsão, dados por:

$$REM Q = \frac{1}{L} \sum_{h=1}^L [X_{obs}(t+h) - X_{prev}(t+h)]^2 = 740.2884$$

$$EMAP(\%) = \frac{100}{L} \sum_{h=1}^L \left| \frac{X_{obs}(t+h) - X_{prev}(t+h)}{x_{obs}(t+h)} \right| = 5.4248\%$$

Por fim as Figura 50a e 50b apresentam os valores observados e previstos para 60 meses, quando essa previsão é feita com realimentação de novas observações mês-a-mês (Figura 50a) e quando não é feita realimentação (Figura 50b), neste caso notamos como a continuidade da aplicação desta técnica de previsão, sem a realimentação com novas observações, conduz a previsão

básicamente da tendência e sazonalidade da carga.

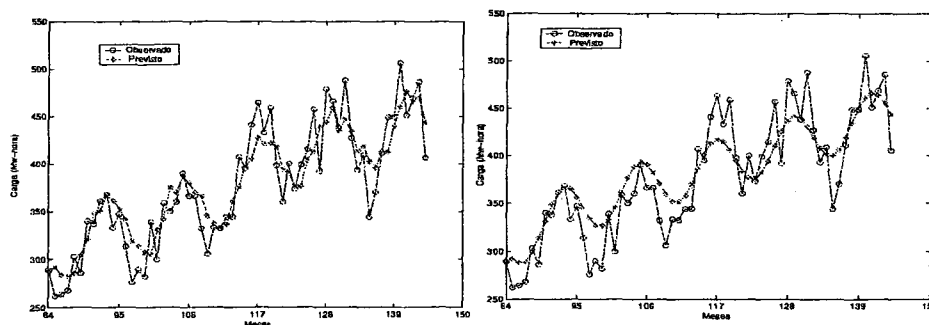


Figura 50-(a)Previsão com atualização mensal, (b) sem atualização

Este procedimento de previsão é refletido no erro de previsão como mostra a Tabela 17.

Tabela 17 Erro de Previsão

$L$	$l$	$EQM$	$EAPM(\%)$
12	12	740.2884	5.4248
60	1	635.7901	5.6563
60	60	802.8544	6.3746

*L-horizonte de previsão*

*l-período de realimentação*

### 6.1.5 Conclusão

A análise da série de 12 anos de consumo mensal de energia elétrica no setor industrial de uma região do estado de São Paulo nós permite concluir que este consumo é sazonal com um período de 12 meses e uma tendência linear.

A análise da tendência revela uma taxa de variação do consumo médio anual decrescente nestes anos, indicando uma retração desse setor ou uma recessão econômica nos anos de 1975 a 1986.

A análise da componente estocástica revela uma correlação de 3 meses nesta componente, mostrando com isso uma dependência previsível com até três meses de antecedência para este efeito aleatório.

A previsão deste consumo pode ser feito para um horizonte de longo prazo com erros médios da ordem de 6.5% e no curto prazo (menos de 6 meses) as previsões de consumo apresentam um erro médio da ordem de 5.5%.

# Referências Bibliográficas

- [1] Andrade, M.G., Carvalho, M.F.H., Barone, J.C., (1987), *Modelos Para Previsão do Consumo de Energia Elétrica, Rel. Técnico RT-OSSE 03-01/87, Cia. Paulista de Força e Luz (CPFL), Campinas-SP.*
- [2] Box, G.E.P., Jenkins, G.M., Reinsel, G.C., (1994), *Time Series Analysis Forecasting and Control, 3th Ed., Prentice Hall, New Jersey*
- [3] Galian, F.D., Handschin, E., Fiechter, A.R., (1974), *Identification of Stochastic Electric Load Models from Physical Data, IEEE Tras. Autom. Control, vol. AC-19, 6,887-893.*
- [4] Gupta, P.C., (1969), *Statistical and Stochastic Methods for Probabilistic Forecasting of Electrical Demands, in Utility Systems, PhD Thesis, Purdue University, Indiana.*
- [5] Gupta, P.C., (1971), *A Stochastic Approach to Peak Power Demand Forecasting in Electric Utility Systems, IEEE Tras, PAS Vol. 90, pp 824-832.*
- [6] Gupta, P.C., Yamada, K., (1972), *Adaptive Short Term Forecasting of Hourly Load Using Weather Information, IEEE Trans. PAS Vol. 91, 2085-2094*
- [7] Stanton, K.N., Gupta, P.C.(1969), *Forecasting Annual or Seasonal Peak Demand in Electric Utility Systems, IEEE Trans. PAS vol. 88, 22-27.*
- [8] Stanton, K.N, (1971), *Medium-Range, Weekly and Seasonal Peak Demand Forecasting by Probabulyty Methodos, IEEE Trans. PAS, vol90,pp.1183-1189.*
- [9] Taylor, L.D. (1975), *The Demand of Electricity: A Survey, Bell Journal of Economics, Vol.6, No.1, 74-110.*

- [10] Rao, S.C., Singh, S., (1981), *Statistical Correlation Method to Predict Peak Demand*, IE(I) Journal-EL, Vol. 61. pp189-1982
- [11] Vemuri, S., Hill, E.F., Balasubramaniam, R., (1973), *Load Forecasting Using Stochastic Models*, 8th P.I.C.A., pp 31-37, Mineapolis.