

UNIVERSIDADE FEDERAL DE SANTA CATARINA
COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO
EM ENGENHARIA

UMA ABORDAGEM À PREVISÃO ATRAVÉS
DE MODELAGEM ESTOCÁSTICA E IDENTIFICAÇÃO

HÉLIO ENIR MARODIN

FLORIANÓPOLIS
SANTA CATARINA - BRASIL
SETEMBRO - 1973

UNIVERSIDADE FEDERAL DE SANTA CATARINA
COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO
EM ENGENHARIA

UMA ABORDAGEM À PREVISÃO ATRAVÉS
DE MODELAGEM ESTOCÁSTICA E IDENTIFICAÇÃO

HELIO ENIR MARODIN

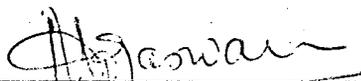
TESE SUBMETIDA A APRECIÇÃO COMO REQUISITO
PARCIAL PARA A OBTENÇÃO DO GRAU DE:
MESTRE EM CIÊNCIAS EM ENGENHARIA INDUSTRIAL

FLORIANÓPOLIS
SANTA CATARINA - BRASIL
SETEMBRO - 1973

ESTA TESE FOI JULGADA ADEQUADA PARA A
OBTENÇÃO DO TÍTULO DE

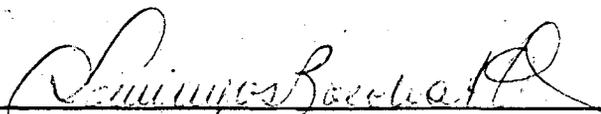
"MESTRE EM CIÊNCIAS"

E APROVADA EM SUA FORMA FINAL PELO O-
RIENTADOR E PELO CURSO DE PÓS-GRADUAÇÃO



Prof. Rajamani Doraiswami

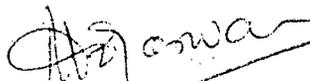
Orientador

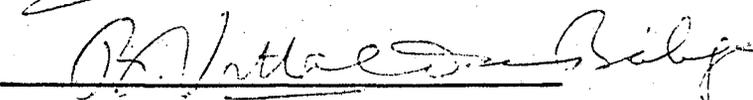


Prof. Domingos Boechat Alves

Integrador do Curso de Pós-Graduação

BANCA EXAMINADORA:





Paulo Nascimento

A minha esposa

A meu filho

R E S U M O

Nesse trabalho nós propomos um esquema de previsão para um processo estocástico mascarado por Ruído Branco. Os processos estocásticos desconhecidos são considerados Gaussianos e nenhuma suposição é feita no que se refere ao ruído nas observações.

O esquema de trabalho foi sugerido pela sequência lógica que uma pessoa segue quando se propõe a fazer previsões.

Primeiramente, nós tratamos o problema de determinação da forma funcional do Modelo Dinâmico. Assumimos que os processos estocásticos são a resposta de um sistema linear excitado por Ruído Branco. Usamos a representação por variáveis de estado para o sistema. Então, propomos um método para determinação dos parâmetros do modelo a partir de dados, supostos isentos de erro. Finalmente, conhecendo-se o modelo para o processo estocástico desconhecido, propomos um esquema de previsões. Este consiste na determinação de ser o processo de observações Gaussiano ou não. Um esquema linear é sugerido para processos Gaussianos e um esquema não linear para processos não Gaussianos.

Uma correção adaptativa é proposta de modo que os erros de previsão são atenuados, corrigindo-se a função não linear.

Mostramos claramente a superioridade de previsões não lineares para processos não Gaussianos e de previsões lineares somente para processos Gaussianos.

A B S T R A C T
=====

In this work we propose a forecasting scheme for a stochastic process corrupted by white noise. The Unknown Stochastic Process is assumed Gaussian and no assumption is made regarding the observation noise.

The scheme proposed follows a logical sequence from the acquisition of the data to the final forecasting algorithm.

First we treat the problem of determining the functional form of the dynamic model. We assume the stochastic process to be an output of a linear system excited by white noise. We use state-variable representation for the system. Then, we propose a method for determining the parameters of the model from the data assuming no error. Finally, knowing the model for the unknown stochastic process, we propose a forecasting scheme. It consists of determining whether the observation process is Gaussian or not. A linear model is suggested for Gaussian and nonlinear for non-Gaussian processes.

An adaptive correction is proposed so that the forecasting errors are corrected by correcting the nonlinear function.

We show clearly the superiority of nonlinear forecast for non-Gaussian and linear forecast only for Gaussian processes.

A G R A D E C I M E N T O S

- Quero aqui agradecer a todos meus Professores, colegas e amigos que muito me auxiliaram nesse trabalho com seus conhecimentos e seu incentivo sempre presentes. Em particular quero agradecer aos colegas Eng^{os}. João Nildo de Souza Vianna e Wilson Luiz Bannach, que com seu espírito crítico e auxílio muito contribuíram. Aos professores e auxiliares do Departamento de Ciências Estatísticas e da Computação da Universidade Federal de Santa Catarina, especialmente aos Professores Renato e Márcia Rabuske, que sempre souberam auxiliar na superação das dificuldades com o equipamento, na execução da simulação.

- À CAPES e ao BNDE pelo apoio financeiro.

- Muito especialmente ao Professor Rajamani Doraiswami, Ph.D., meu Orientador, por sua infinita paciência e dedicação, que contribuiu muito especialmente para a concretização deste trabalho.

Í N D I C E

PÁG.

CAPÍTULO I

1. INTRODUÇÃO	1
---------------------	---

CAPÍTULO II

2. DETERMINAÇÃO MATEMÁTICA DE UM MODELO ECONOMÉTRICO LI NEAR	3
2.1 - Introdução	3
2.2 - Processos Estocásticos	4
2.2.1 - Definição	4
2.2.2 - Caracterização dos Processos Estocásticos	5
2.2.3 - Classificação dos Processos Estocásticos	6
2.2.4 - Correlação, Ortogonalidade e Independência de Processos Estocásticos	10
2.2.5 - Espaço de Processos Estocásticos	11
2.2.6 - Representação por uma Base de um Sub-Espaço de Hilbert	12
2.3 - Representação de um Processo Estocástico por Va - riáveis de Estado	14
2.3.1 - Exemplo 1	14
2.3.2 - Exemplo 2	17
2.4 - Análise Comparativa e Aplicação do Modelo	19
2.5 - Modelos de Observação	22
2.6 - Conclusão do Segundo Capítulo	24

CAPÍTULO III

3. SIMULAÇÃO	26
3.1 - Introdução	26

3.2 - Testes de Comportamento Estatístico dos Geradores de Números Aleatórios	27
3.2.1 - Testes de Média e Variância	27
3.2.2 - Teste de Chi-Quadrado	27
3.3 - Geração de Números Aleatórios com Distribuição Normal (0, 1)	32
3.4 - Gerador de Números Aleatórios Independentes com Distribuição de Laplace	34
3.5 - Gerador de Números Aleatórios Independentes com Distribuição de Cauchy	36
3.6 - Simulação dos Modelos	37
3.7 - Simulação dos Modelos de Observação	43
3.8 - Considerações e Observações Sobre Esta Simulação.	44

CAPÍTULO IV

4. UM ESTUDO SOBRE IDENTIFICAÇÃO DE MODELOS	62
4.1 - Introdução	62
4.2 - O Problema Geral de Estimação	63
4.2.1 - Tipos de Estimadores	63
4.3 - Obtenção dos Parâmetros do Modelo a Partir dos Dados	64
4.3.1 - Estimadores e Critérios de Avaliação	64
4.3.2 - Propriedades do Estimador de Máxima Probabilidade	66
4.3.3 - Determinação dos Parâmetros de um Modelo cuja Forma Funcional é conhecida	67
4.3.4 - O Problema de Estimação dos Parâmetros de um Modelo cuja Forma Funcional é Desconhecida	71
4.4 - Escolha do Modelo Mais Adequado	72
4.4.1 - Exemplo de Aplicação	75
4.5 - Observações Sobre Este Capítulo	78

CAPÍTULO V

5. UMA ABORDAGEM ÀS PREVISÕES	79
-------------------------------------	----

5.1 - Introdução	79
5.2 - O Problema de Previsão Estocástica	80
5.3 - Estimação com uma Única Observação	80
5.3.1 - Análise do Comportamento do Estimador Proposto..	82
5.4 - Previsão Dispondo-se de um Conjunto de Observa - ções	87
5.5 - Predição	91
5.6 - Determinação da Função de Correção	91
 <u>CAPÍTULO VI</u>	
6. CONCLUSÃO	94
 NOTAÇÃO E SIMBOLOGIA USADA	 98
 REFERÊNCIAS BIBLIOGRÁFICAS	 101

C A P Í T U L O I

1 - INTRODUÇÃO

Com esse trabalho procuramos estabelecer um novo método de previsões. Visando alcançar este objetivo, discorremos ao longo deste trabalho sobre Modelagem Estocástica, Simulação, Identificação de Modelos para, finalmente, definirmos um esquema recursivo e auto-ajustável de previsões. Tal linha de ação nos foi sugerido pela sequência lógica que uma pessoa enfrenta quando se propõe a fazer previsões ou estimativas.

Assim sendo, procuramos, a partir do segundo capítulo, desenvolver os temas acima referidos, iniciando pelo que chamamos de Modelagem, ou seja, a obtenção da forma matemática de um modelo de sistemas estocásticos. Nesse capítulo analisamos inicialmente os processos estocásticos, pois sabe-se que o modelo geral de um sistema linear pode ter uma parcela estocástica. Utilizando-se de um sub-espaço do espaço de Hilbert, expressamos um processo estocástico, pertencente a esse sub-espaço, como uma combinação linear de processos de Ruido Branco. Baseados nisso definimos um modelo estocástico geral para representar o comportamento dinâmico dos sistemas. Depois tratamos dos modelos de observação, e fazemos três hipóteses representativas de confiabilidade de dados.

No terceiro capítulo simulamos os modelos dinâmicos e de observação, anteriormente encontrados. Obtemos gráficos do comportamento de um processo de Ruido Branco, dos modelos dinâmicos de primeira e segunda ordem e dos modelos de observação.

Analizando esses resultados, encontramos claras vinculações entre as variáveis do processo e os modelos representados.

No quarto capítulo, que denominamos Identificação do Modelo, procuramos analisar as técnicas usadas para determinar os parâmetros dos modelos, conhecendo-se um conjunto de dados. Três casos são frequentemente encontrados na solução prática de problemas de previsão e são tratados em separado. Inicialmente tratamos

da determinação do modelo conhecendo-se a sua forma funcional e um conjunto de dados; depois tratamos da determinação do modelo quando desconhecemos sua forma funcional, embora suponhamos que seja linear. Por último, abordamos o caso em que a forma funcional do modelo é desconhecida, mas encontra-se em um conjunto finito de modelos possíveis.

A abordagem a esses três casos propõe uma solução para cada um, sendo o primeiro caso solucionado pela técnica do estimador de máxima probabilidade (Maximum Likelihood Estimator). A solução do segundo tipo de problema recai no primeiro caso quando a teoria desenvolvida no primeiro capítulo é usada. O último caso é abordado com o auxílio da teoria da decisão. Para esse problema, desenvolvemos uma política de tomada de decisão determinística, pois assumimos como critério de avaliação uma dicotomização do espaço amostral.

Finalmente, no quinto e último capítulo, tratamos dos métodos e propomos uma técnica para determinação de previsões. Esta técnica explora as características estatísticas dos processos estocásticos envolvidos, e tem a vantagem de ser recursiva, permitindo a fácil programação em computadores sem sobrecarregar a memória ou utilizar artifícios de programação. Assim, determinamos estimadores cuja forma funcional depende exclusivamente das distribuições de probabilidade dos processos estocásticos envolvidos e são analisados vários casos, inclusive o caso mais comumente encontrado na prática. Por último, sugerimos um esquema adaptativo que ajusta os coeficientes dos modelos a medida que o tempo decorre e novas informações são obtidas.

C A P Í T U L O I I

2. DETERMINAÇÃO MATEMÁTICA DE UM MODELO ECONOMETRICO LINEAR

2.1 - INTRODUÇÃO

Os modelos econométricos de sistemas de consumo e renda têm sido largamente utilizados e têm demonstrado ser preciosos auxiliares em previsões a curto e médio prazo, e também no estabelecimento de políticas estratégicas.

Nosso objetivo será estabelecer, de uma forma unificada, um modelo matemático linear para um processo de consumo de bens normais. Tal modelo deverá ser uma síntese de todos ou pelo menos da maioria dos modelos lineares existentes.

Os processos de consumo e de demanda têm, como de um modo geral todos os processos econômicos, uma característica inerente que é a aleatoriedade das funções representativas com relação ao tempo. Por isso, a propriedade principal de tais processos é apresentarem um comportamento estocástico. Evidentemente o modelo proposto deverá conter uma parcela que represente esse comportamento aleatório.

Outra característica esperada do modelo é a capacidade de poder-se utilizá-lo eficientemente para a computação dos resultados, permitindo a rápida operação de massas de dados sem exceder a capacidade de memória ou exigir artifícios de programação, considerando-se, sempre, o custo das estimativas fornecidas pelo modelo. Em outras palavras, o modelo deverá ser tal que não seja preciso carregar toda a informação sobre o comportamento passado do sistema, pois de outra forma a memória requerida pelo computador seria impraticavelmente grande.

Nesse capítulo nós expressaremos um processo estocástico como uma combinação linear de vetores de uma base ortogonal de processos estocásticos de ruído branco (White noise processes). Posteriormente, observamos os processos estocásticos como a res-

posta de um sistema linear excitado por um processo de Ruido Branco, expressaremos o processo usando o modelo das variáveis de estado. Então a representação por variáveis de estado é mostrada ser o modo mais geral de representação de um processo estocástico segundo um modelo linear.

2.2 - PROCESSOS ESTOCÁSTICOS

2.2.1 - Definição

Um processo estocástico pode ser definido como uma coleção de variáveis aleatórias.

Seja Ω o espaço amostral, e $w \in \Omega$ um dado experimento; seja I o conjunto dos números reais; então o conjunto das funções reais ou complexas $\{X(w,t)\}$ com $w \in \Omega$, $t \in I$ é um processo aleatório ou processo estocástico.

Para interpretarmos o significado de $\{X(w,t)\}$ tomemos w_i um dado experimento e $t \in I$. Então, para um dado $w = w_i$, $X(w_i,t)$ representa uma particular realização do processo e $X(w_i,t) = \phi(t)$ é uma função de t . Considerando $t = t_i$, $X(w,t_i)$, é uma quantidade que somente depende de w , portanto é uma variável aleatória. Os dois casos acima podem ser graficamente representados como nas figuras 1 e 2.

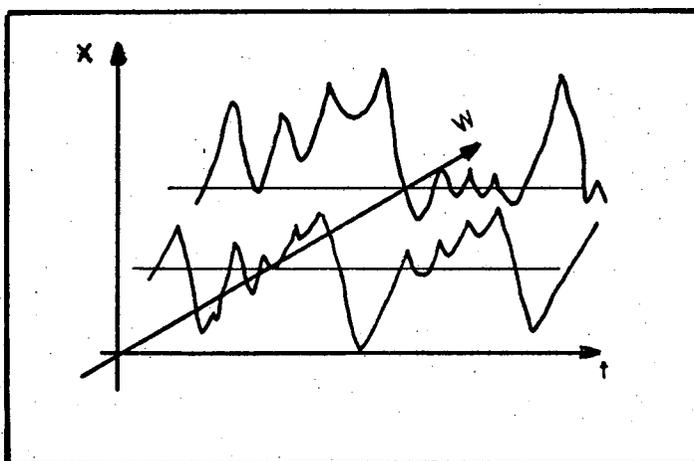


FIG. 1 - Realizações particulares do processo
(w fixo)

Portanto, um processo estocástico pode ser considerado como uma função de duas variáveis, no caso w e t . Em nossas aplicações e exemplos, consideraremos a variável t como sendo representativa -

do tempo e w denotará os diferentes experimentos, porém, de um modo geral, w e t são quaisquer.

Por exemplo, $X(w_i, t)$ fornece a evolução temporal do fenômeno aleatório para um experimento particular $w = w_i$.

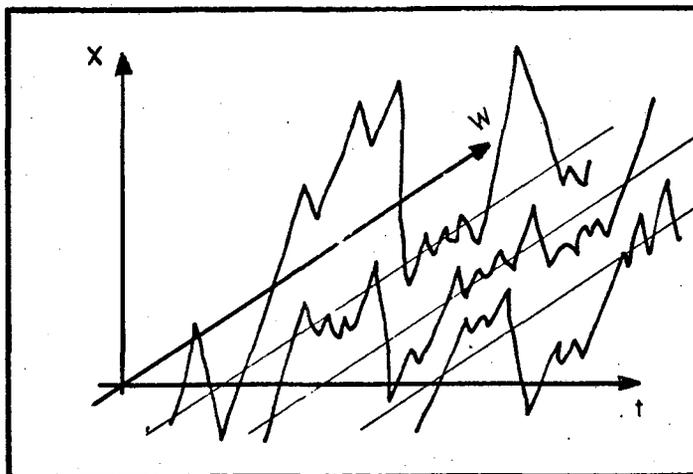


FIG. 2 - Variáveis aleatórias resultantes de diversos experimentos.
(t fixo)

2.2.2 - Caracterização dos Processos Estocásticos

A função densidade de probabilidade conjunta de $\{X(w, t)\}$ descreve completamente o processo, porém é muito difícil usá-la na prática, como modelo para um processo estocástico*. Por outro lado, a função densidade de probabilidade conjunta apresenta o comportamento macroscópico final do sistema, envolvendo o conhecimento da máxima informação possível sobre o processo.

No entanto, nos problemas usuais, a informação é muito limitada, e mesmo nos casos em que grande quantidade de informação é disponível, o modelo resultante não é simples, nem geral. Além disso, exige extenso conhecimento de métodos estatísticos, divergindo do objetivo que é solucionar um problema de modelagem e estimativa.

$$* - f(X(w,1), X(w,2), \dots, X(w,n)) = \frac{\partial^n F(X(w,1), X(w,2), \dots, X(w,n))}{\partial X(w,1) \partial X(w,2) \dots \partial X(w,n)},$$

onde $F(X(w,i))$ é a função distribuição de probabilidade conjunta acumulada de $X(w,i)$, ($i=1,2,\dots,n$); portanto a determinação da função densidade de probabilidade conjunta implica na existência de derivadas parciais de ordem 1 a n .

2.2.3 - Classificação dos processos estocásticos

a) Quanto à continuidade das variáveis: Podemos classificar os processos como discretos ou contínuos, dependendo de:

- i) ser t discreto ou contínuo
- ii) ser X discreto ou contínuo

Dentro dessa classificação poderemos ter, então:

1) Processo de variável aleatória e tempo discreto = são os processos em que t pertence ao conjunto dos números inteiros positivos, negativos ou nulos. O processo em si pode também ser considerado como discreto, isto é, $X(\omega, t)$ poderá somente assumir valores discretos, tais como $X(\omega, 0)$, $X(\omega, 1)$, $X(\omega, 2)$, O processo esboçado na figura 3 é um exemplo de processo de variável aleatória e tempo discretos.

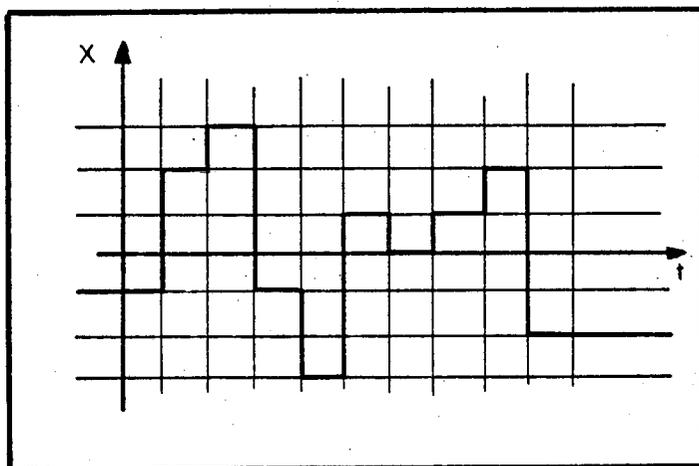


FIG. 3 - Processo estocástico com tempo aleatório.

2) Processo aleatório contínuo com tempo discreto = quando o processo admite valores contínuos entre sucessivos valores de tempo t , como representado na figura 4.

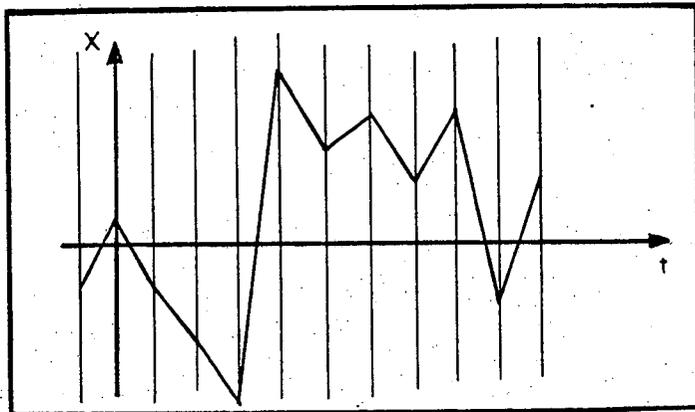


FIG. 4 - Processo estocástico contínuo e tempo discreto.

3) Processo aleatório com tempo contínuo = é o caso em que o processo só pode assumir valores discretos através do tempo. Um exemplo desse caso é mostrado na figura 5.

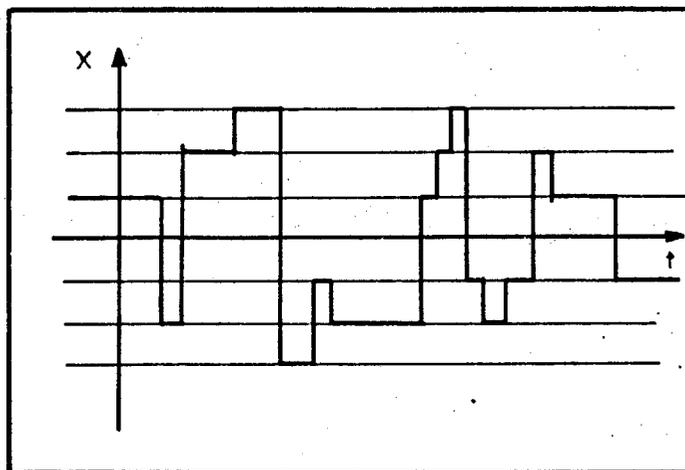


FIG. 5 - Processo aleatório com tempo contínuo.

4) Processo contínuo no tempo = é o caso mais geral, em que tanto o processo como o tempo de medida são contínuos, como o mostrado na figura 6.

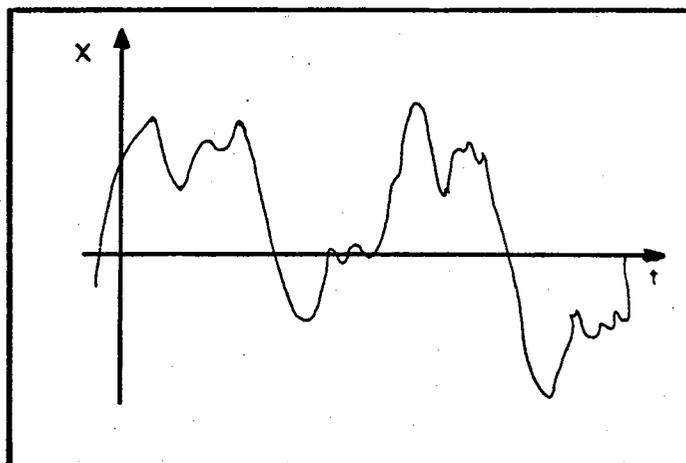


FIG. 6 - Processo perfeitamente contínuo

b) Quanto à existência física do processo: Podemos classificar os processos estocásticos como sendo ou não um processo físico. Se o valor do processo em qualquer instante de tempo é estatisticamente independente dos valores futuros, então é um processo causal, i.é, um processo estocástico é causal se $X(\omega, t)$ é estatisticamente independente de $X(\omega, \xi)$ para todo $\xi > t$, ou ainda:

$$P \{X(\omega, t) \mid X(\omega, \xi), \xi > t\} = P \{X(\omega, t)\}$$

c) Quanto ao campo de definição da variável X : os processos podem ser classificados como:

- Processos aleatórios de valores reais: quando $X(w,t)$ é um número real para todo t e para qualquer experimento w .
- Processos aleatórios de valores complexos: quando $X(w,t)$ pertence ao campo dos números complexos para qualquer t e w . Nesse caso, $X(w,t)$ pode ser expresso por:

$$X(w,t) = X'(w,t) + j X''(w,t),$$

onde j é a unidade imaginária $\sqrt{-1}$.

Em nosso trabalho, a menos que explicitamente dito, consideraremos somente os processos estocásticos físicos, causais, reais com tempo discreto e variável aleatória contínua. Não consideraremos os processos de tempo contínuo porque podemos aproximar um processo de tempo discreto a outro de tempo contínuo tanto quanto queiramos, tomando-se para isso um intervalo entre t e $t+1$ tão pequeno quanto se queira.

Além disso, o processo a tempo discreto elimina muitas dificuldades de ordem matemática, envolvidas nas definições de derivadas. Por isso, a partir desse ponto, notaremos os processos estocásticos a tempo discreto $X(w,t)$ por $X(w,n)$, $n = \dots, -k, \dots, 0, 1, 2, \dots, k, \dots$

d) Quanto à independência estatística no tempo: consideremos as relações estatísticas em relação ao tempo:

- 1 - Caso geral: seja um processo estocástico $X(w,n)$. No caso mais geral de dependência temporal $X(w,n)$ é dependente de todos $X(w,m)$, para $m = n-1, n-2, \dots, n-k+1, n-k$, isto é,

$$P\{X(w,n) | X(w,m), m < n\} = P\{X(w,n) | X(w,n-1), X(w,n-2), \dots, X(w,n-k)\}$$

- 2 - Processo Markoviano: quando $X(w,n)$ é estatisticamente dependente somente de $X(w,n-1)$, ou seja:

$$P\{X(w,n) | X(w,m), m < n\} = P\{X(w,n) | X(w,n-1)\}$$

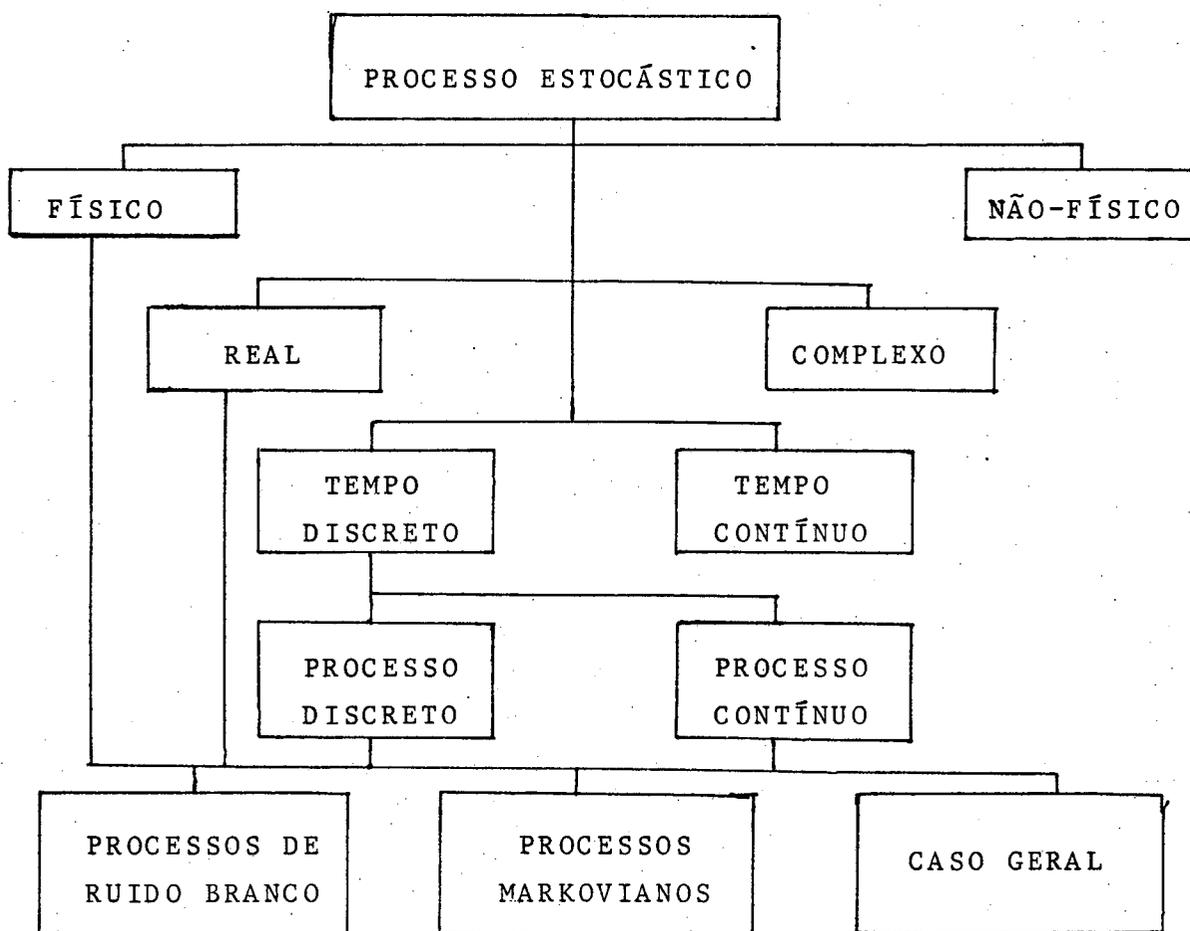
- 3 - Processo de ruído branco (white noise process): um processo $X(w,n)$ é chamado de processo de ruído branco se $X(w,n)$ é estatisticamente independente de $X(w,m)$ para

qualquer m . De outra forma podemos expressar um processo de ruído branco como:

$$P \{X(w,n) \mid X(w,m), m \in I\} = P \{X(w,n)\}$$

Portanto, um processo de ruído branco é uma sequência de processos independentes. Devido à grande importância que os processos de ruído branco assumirão no decorrer desse trabalho, os denotaremos por:

$$W(w,n), \quad n = 0, \pm 1, \pm 2, \dots, \pm \infty$$



QUADRO 1 - CLASSIFICAÇÃO DOS PROCESSOS ESTOCÁSTICOS

Os processos de ruído branco apresentam a importante propriedade de serem ortogonais em relação ao valor esperado ; se $W(w,n)$ é um conjunto de processos com média nula e variância unitária,

$$E [W(w,n) \cdot W(w,m)] = \delta_{nm}$$

$$\text{onde } \delta_{nm} = \begin{cases} 0, p / m \neq n \\ 1, p / m = n \end{cases}$$

Finalizando a classificação dos processos estocásticos, apresentamos no Quadro 1 um esquema geral dessa classificação.

2.2.4 - Correlação, ortogonalidade e independência de processos estocásticos.

Dois processos $X_1(w,n)$ e $X_2(w,j)$ são chamados de não-correlacionados se para qualquer n e j ,

$$E [X_1 \cdot X_2] = E [X_1] \cdot E [X_2]$$

onde $X_1 = X_1(w,n)$ e $X_2 = X_2(w,j)$.

Portanto, dois processos não correlacionados apresentam covariância $C_{X_1X_2}$ nula,

$$C_{X_1X_2} = E \{ (X_1 - \mu_{X_1})(X_2 - \mu_{X_2}) \} = E \{ X_1X_2 \} - E \{ X_1 \} \cdot E \{ X_2 \}$$

pois,

$$\mu_{X_1} = E [X_1] \quad \text{e} \quad \mu_{X_2} = E [X_2] .$$

No caso geral de n processos X_1, X_2, \dots, X_n define-se a matriz de covariância $COV(X_1, X_2, X_3, \dots, X_n)$, cujo elemento de ordem i, j é definido como $E \{ [X_i - E(X_i)] \cdot [X_j - E(X_j)] \}$; nesse caso o não correlacionamento de dois processos X_i e X_j implica na nulidade do elemento de ordem ij da matriz covariância

$$COV (X_1, \dots, X_j) .$$

Dois processos são ditos ortogonais se $X_i \perp X_j$ para todos os pares X_i e X_j , $i \neq j$.

Dois processos estocásticos $X(w,n)$ e $Y(w,j)$ são chamados de independentes se a função densidade de probabilidade conjunta de $X(w,n)$ e $Y(w,j)$ satisfaz:

$$f(X(w,n), Y(w,j)) = f(X(w,n)) \cdot f(Y(w,j)) .$$

Não há dificuldade de se estender essa definição para n processos.

Portanto, se $X(w,n)$ é uma sequência de processos aleatórios independentes, e com média nula, pode-se inferir:

$$\begin{aligned} E [X(w,j) \cdot X(w,k)] &= E [X(w,j)] E [X(w,k)] = \\ &= \delta_{jk} \cdot E [X(w,j) \cdot X(w,k)] \end{aligned} \quad (1)$$

onde δ_{jk} é o delta de Kronecker.

Logo, uma sequência de processos aleatórios independentes com média nula é ortogonal em relação ao valor esperado.

2.2.5 - Espaço de Processos Estocásticos

Neste trabalho, a menos que expressamente dito, consideraremos somente processos estocásticos de média nula e variância finita, i.é.,

$$E [X(w,n)] = 0, \quad \forall n \quad (2)$$

$$E [X^2(w,n)] < \infty, \quad \forall n \quad (3)$$

Tenha-se em mente que a restrição imposta por (2) não implica em perda de generalidade, pois qualquer processo de média não nula poderá ser considerado por meio de $X - \mu_X$. Podemos considerar os processos que satisfazem (2) e (3) como vetores de um espaço de vetores aleatórios.

Considere-se o espaço de Hilbert^{10,14}, que notaremos por \mathcal{H} , formado por processos estocásticos, sendo o produto interno e a norma definidos como:

$$\langle X(w,n), Y(w,j) \rangle = E [X(w,n) \cdot Y(w,j)], \quad X(w,n) \text{ e } Y(w,j) \in \mathcal{H} \quad (4)$$

$$\|X(w,n)\| = \sqrt{E [X^2(w,n)]} \quad (5)$$

Isto forma um espaço de Hilbert, H , de todos os processos estocásticos $X(w,n)$, tal que:

$$\|X(w,n)\| < \infty, \quad \forall n \quad (6)$$

Considere-se os processos em que

$$E [X(w,n)] = 0 \quad (7)$$

A restrição (7) não implica em perda de generalidade porque os processos de média não nula também pertencem a este espaço.

Seja um processo $X(w,n)$ e \mathcal{H} ; vamos procurar representá-lo por meio de uma base ortogonal, se esta existir.

Sabe-se que os processos de ruído branco $W(w,i)$, ($i=1, \dots, n$) satisfazem:

$$\langle W(w,i), W(w,j) \rangle = 0, \quad (i,j = 1, \dots, n) \quad (8)$$

porque:

$$\langle W(w,i) \cdot W(w,j) \rangle = E[W(w,i) \cdot W(w,j)] = E[W(w,i)] \cdot E[W(w,j)] = 0$$

Portanto, $\{W(w,i)\}$ forma uma sequência ortogonal de processos de ruído branco.

Pode-se escrever $X(w,n)$ na base dos vetores ortogonais $\{W(w,i)\}$? Para obter-se esta representação consideraremos somente aqueles processos $X(w,n)$ para os quais $\{W(w,i)\}$ forma uma base. Isto implica em que $\{W(w,i)\}$ gera um sub-espaço do espaço de Hilbert \mathcal{H} e nos propomos escolher somente aqueles processos $X(w,n)$ pertencentes a este sub-espaço.

2.2.6 - Representação por uma base de um sub-espaço de Hilbert.

Como vimos anteriormente, o conjunto $\{W(w,n)\}$ é uma base de um sub-espaço H de \mathcal{H} , e por isso, qualquer vetor \dots $\{X(w,n)\} \in H$ pode ser representado por uma combinação linear de $\{W(w,n)\}$, i.é.,

$$X(w,n) = \lim_{N \rightarrow \infty} \sum_{j=-N}^n a_{nj} \cdot W(w,j) \quad (9)$$

Porém, a equação (9) é uma igualdade de variáveis aleatórias. Qual o significado de (9)? O significado de (9) é que $X(w,n)$ converge para $\sum_{j=-N}^n a_{nj} \cdot W(w,j)$ quando $N \rightarrow \infty$, no sentido dos mínimos quadrado, i.é.,

$$\lim_{N \rightarrow \infty} E \left[X(w,n) - \sum_{j=-N}^n a_{nj} \cdot W(w,j) \right]^2 \rightarrow 0 \quad (10)$$

É fácil demonstrar que $\{a_{nj}\}$ é tal que:

$$a_{nj} = \frac{E [X(w,n) \cdot W(w,j)]}{E [W^2(w,j)]} \quad (11)$$

Usando as equações (10) e (11) pode-se reescrever (10) como:

$$\lim_{N \rightarrow \infty} E [X^2(w,n)] - \sum_{j=-N}^n a_{nj}^2 \cdot E [W^2(w,j)] \rightarrow 0 \quad (12)$$

Então, desde que $E [X^2(w,n)] < \infty$, podemos satisfazer (6).

Um exame mais detalhado da eq. (9) mostra que $X(w,n)$ pode ser considerado como a resposta de um sistema linear discreto excitado por um processo de ruído branco $\{W(w,j)\}$, ($j = -\infty, \dots, n$), como na Figura 7.

Pelo que ficou dito, um processo estocástico $X(w,n)$ pode ser representado pela resposta de um sistema linear discreto excitado por um processo de ruído branco. Por conveniência, desse ponto em diante deixaremos de representar o parâmetro w na descrição dos processos estocásticos, de modo que:

$X(w,n)$ será notado por $X(n)$ e

$W(w,n)$ será notado por $W(n)$.

Usando esta notação, a equação (9) passará a ser:

$$X(n) = \sum_{j=-\infty}^n a_{nj} \cdot W(j) \quad (13)$$

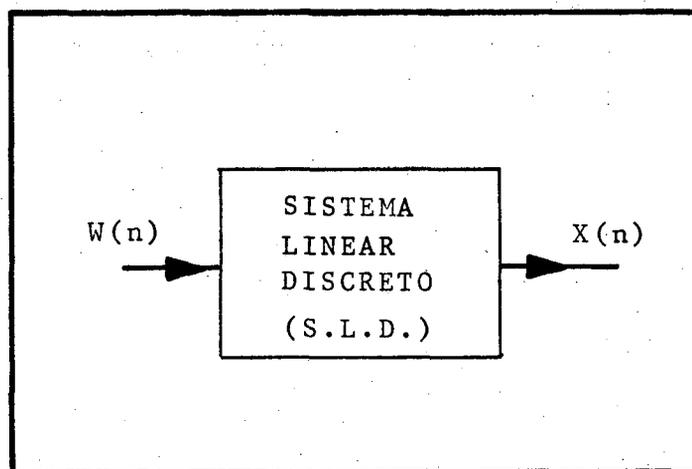


FIG. 7 - Sistema linear discreto excitado por um processo de ruído branco.

Se agora considerarmos $X(n)$ como a resposta de um sistema linear discreto a uma excitação $W(j)$, é fácil constatar que a equação (13) representa a solução da equação de diferenças' que descreve o sistema considerado. Portanto, como a equação do

sistema descreve o comportamento do sistema e também para maior simplicidade de representação, consideraremos, a partir desse ponto, as equações do sistema linear discreto como representação dos sistemas estocásticos considerados.

Um sistema linear discreto pode ser considerado, em sua forma mais geral, representado por uma equação e diferenças da forma:

$$\begin{aligned} X(n-m) + a_{n1}.X(n-m-1) + a_{n2}.X(n-m-2) + \dots + a_{nm}.X(n) &= \\ &= b_{n0}.W(n-m) + b_{n1}.W(n-m-1) + \dots + b_{nm}.W(n) \end{aligned} \quad (14)$$

Dependendo dos valores assumidos pelos coeficientes a_{nj} e b_{nj} , a eq. (14) representará diferentes sistemas, alguns muito conhecidos como:

Exemplo 1: Processo autoregressivo de ordem m . Quando os valores de $b_{nj} = 0$ se $j \neq 0$, i.é.,

$$X(n-m) + a_{n1}.X(n-m-1) + \dots + a_{nm}.X(n) = b_{n0}.W(n-m)$$

Exemplo 2: Se os coeficientes a_{nj} assumirem o valor zero, $a_{nj} = 0$, j , o processo é conhecido como processo de médias móveis de ordem m ; a equação (14) ficará:

$$X(n) = b_{n0}.W(n-m) + b_{n1}.W(n-m-1) + \dots + b_{nm}.W(n)$$

2.3 - Representação de um processo estocástico por variáveis de estado.

Como sabemos que um processo estocástico pode ser representado pela equação de um sistema linear discreto, nada mais natural que procurar expressá-lo através das variáveis de estado do sistema. Iremos agora mostrar, através de alguns exemplos simples, como uma equação de diferenças de ordem m , como a dada pela eq. (14), pode ser escrita como m equações de diferenças de primeira ordem, usando a técnica das variáveis de estado.

A representação por variáveis de estado é largamente utilizada em estimação, detecção, etc.^{7,8,16}.

2.3.1 - Exemplo 1:

Seja dada a equação:

$$Z(n+m) + a_{n1} \cdot Z(n+m-1) + a_{n2} \cdot Z(n+m-2) + \dots + a_{nm} \cdot Z(n) = U(n) \quad (15)$$

representativa de um sistema linear com excitação dada pela função $U(n)$ e com resposta dada por $Z(n+m)$, como na Fig. 7.

Define-se a variável de estado X , tal que

$$X = \{x_1, x_2, \dots, x_n\}$$

e

$$\begin{cases} x_1(n) = Z(n) \\ x_2(n) = Z(n+1) \\ x_3(n) = Z(n+2) \\ \vdots \\ x_{m-1}(n) = Z(n+m-2) \\ x_m(n) = Z(n+m-1) \end{cases} \quad (16)$$

Calculando-se as equações acima, no ponto $n+1$, e substituído o conjunto resultante na equação (15), obteremos:

$$x_m(n+1) + a_{n1} \cdot x_m(n) + a_{n2} \cdot x_{m-1}(n) + \dots + a_{nm} \cdot x_1(n) = U(n) \quad (17)$$

ou ainda

$$\begin{aligned} x_m(n+1) = & -a_{n1} \cdot x_m(n) - a_{n2} \cdot x_{m-1}(n) - a_{n3} \cdot x_{m-2}(n) - \\ & - \dots - a_{nm} \cdot x_1(n) + U(n) \end{aligned} \quad (18)$$

podemos então expressar as equações (16), (17) e (18) em sua forma matricial:

$$X(n+1) = A \cdot X(n) + B' \cdot U(n) \quad (19)$$

$$Z(n) = C' \cdot X(n) \quad (20)$$

onde,

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a_{n1} & -a_{n2} & & \dots & -a_{nm} \end{bmatrix}$$

e A é uma matriz quadrada de ordem m .

$$B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \quad e \quad C = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \text{ com } B \text{ e } C \text{ vetores colunas de } m \text{ linhas.}$$

Os vetores $X(n+1)$ e $X(n)$ são vetores de m linhas representativos dos m estados do sistema, e

$$U(n) = [u(n)]$$

$$Z(n) = [z(n)]$$

O sistema linear discreto representado pelas equações (19) e (20) descrevem um sistema como o da Fig. 8.

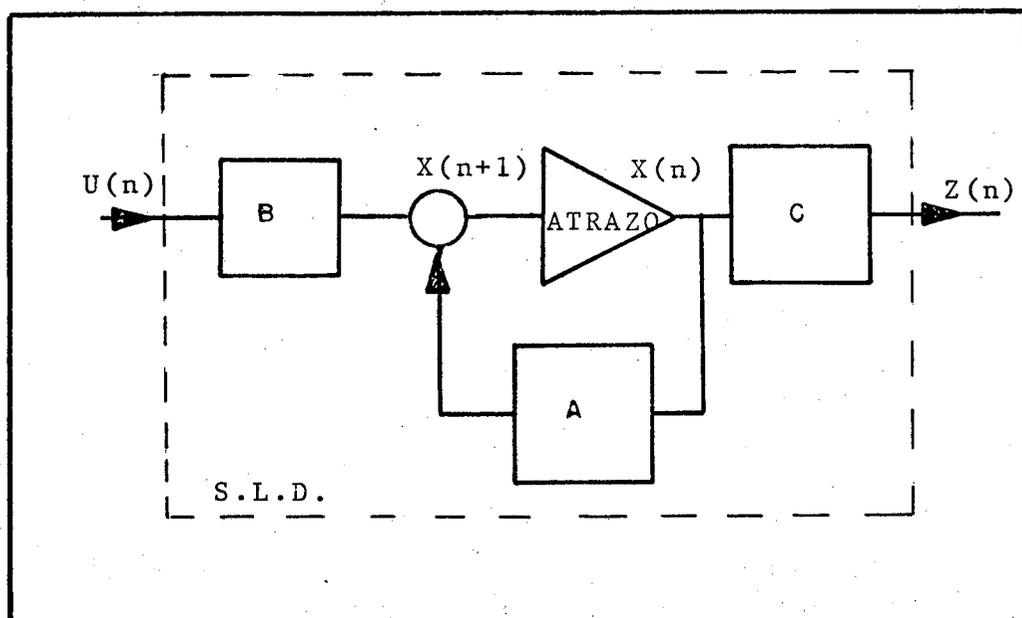


FIG. 8 - Diagrama esquemático do sistema linear discreto representado pelas equações (19) e (20).

Observação: O bloco  representa um operador linear que transfere uma variável $X(n+1)$ dada no tempo $n+1$ para o tempo n .

Observamos mais uma vez que o sistema linear discreto acima representado pode ser tomado como um sistema linear discreto excitado por uma função de ruído U e que responde com a variável Z .

2.3.2 - Exemplo 2

Consideremos uma equação de diferenças em sua forma mais geral. Seja:

$$\begin{aligned} Z(n+m) + a_{n1}Z(n+m-1) + a_{n2}Z(n+m-2) + \dots + a_{nm}Z(n) &= \\ = b_{n0}U(n+m) + b_{n1}U(n+m-1) + b_{n2}U(n+m-2) + \dots + b_{nm}U(n). \end{aligned} \quad (21)$$

Vamos assumir as seguintes mudanças de variáveis:

$$Z(n) = x_1(n) + \beta_{n0}U(n) \quad (22)$$

$$\left\{ \begin{aligned} x_1(n+1) &= x_2(n) + \beta_{n1}U(n) \\ x_2(n+1) &= x_3(n) + \beta_{n2}U(n) \\ &\vdots \\ x_{m-1}(n+1) &= x_m(n) + \beta_{nm-1}U(n) \\ x_m(n+1) &= -a_{nm}x_1(n) - a_{nm-1}x_2(n) - \dots - a_{n2}x_{m-1}(n) - \\ &\quad - a_{n1}x_m(n) + \beta_{nm}U(n) \end{aligned} \right. \quad (23)$$

Calculando-se sucessivamente a equação (21) nos pontos $n, n+1, n+2, \dots, n+m$, obtêm-se:

$$\left\{ \begin{aligned} Z(n) &= x_1(n) + \beta_{n0}U(n) \\ Z(n+1) &= x_2(n) + \beta_{n1}U(n) + \beta_{n0}U(n+1) \\ Z(n+2) &= x_3(n) + \beta_{n2}U(n) + \beta_{n1}U(n+1) + \beta_{n0}U(n+2) \\ &\vdots \\ Z(n+m-1) &= x_{m-1}(n) + \beta_{nm-1}U(n) + \beta_{nm-2}U(n+1) + \dots + \\ &\quad + \beta_{n1}U(n+m-2) + \beta_{n0}U(n+m-1) \\ Z(n+m) &= -x_1(n) a_{nm} - a_{nm-1}x_2(n) - \dots - a_{n1}x_m(n) + \\ &\quad + \beta_{nm}U(n) + \dots + \beta_{n0}U(n+m) \end{aligned} \right. \quad (24)$$

Substituindo (24) em (21) e agrupando os termos de mesma ordem de U , obteremos:

$$\left\{ \begin{array}{l} 1 \cdot \beta_{n0} \\ a_{n1}\beta_{n0} + \beta_{n1} \\ a_{n2}\beta_{n0} + a_{n1}\beta_{n1} + \beta_{n2} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{nm-1}\beta_{n0} + a_{nm-2}\beta_{n1} + \dots + \beta_{nm-1} \\ a_{nm}\beta_{n0} + a_{nm-1}\beta_{n1} + \dots + a_{n1}\beta_{nm-1} + \beta_{nm} \end{array} \right. \begin{array}{l} = b_{n0} \\ = b_{n1} \\ = b_{n2} \\ \vdots \\ = b_{nm-1} \\ = b_{nm} \end{array}$$

O sistema de equações acima permite-nos determinar facilmente os valores de β , pois o sistema é triangular. Colocando-o sob a forma matricial teremos:

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{n1} & 1 & 0 & \dots & 0 \\ a_{n2} & a_{n1} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{nm} & a_{nm-1} & \dots & \dots & 1 \end{bmatrix} \cdot \begin{bmatrix} \beta_{n0} \\ \beta_{n1} \\ \beta_{n2} \\ \vdots \\ \beta_{nm} \end{bmatrix} = \begin{bmatrix} b_{n0} \\ b_{n1} \\ b_{n2} \\ \vdots \\ b_{nm} \end{bmatrix} \quad (25)$$

ou então:

$$A\beta = B \quad \text{donde,}$$

$$\beta = A^{-1} \cdot B \quad \text{permite obter os valores de } \beta.$$

Por outro lado, o conjunto de eqs. (22) e (23) poderá também ser posto sob a forma matricial, conforme:

$$X(n+1) = A \cdot X(n) + \beta' \cdot U(n) \quad (26)$$

$$Z(n) = C' \cdot X(n) + D \cdot U(n)$$

onde:

$$A_{m \times m} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_{nm} & -a_{nm-1} & \dots & \dots & -a_{n1} \end{bmatrix}, \beta_{m \times 1} = \begin{bmatrix} \beta_{n0} \\ \beta_{n1} \\ \beta_{n2} \\ \vdots \\ \beta_{nm} \end{bmatrix}, C_{m \times 1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{e} \quad D_{1 \times 1} = [\beta_{n0}]$$

Então, pode-se representar o conjunto de equações (26) por um diagrama esquemático representativo do sistema linear que

tem estas equações como modelo, como na figura 9 .

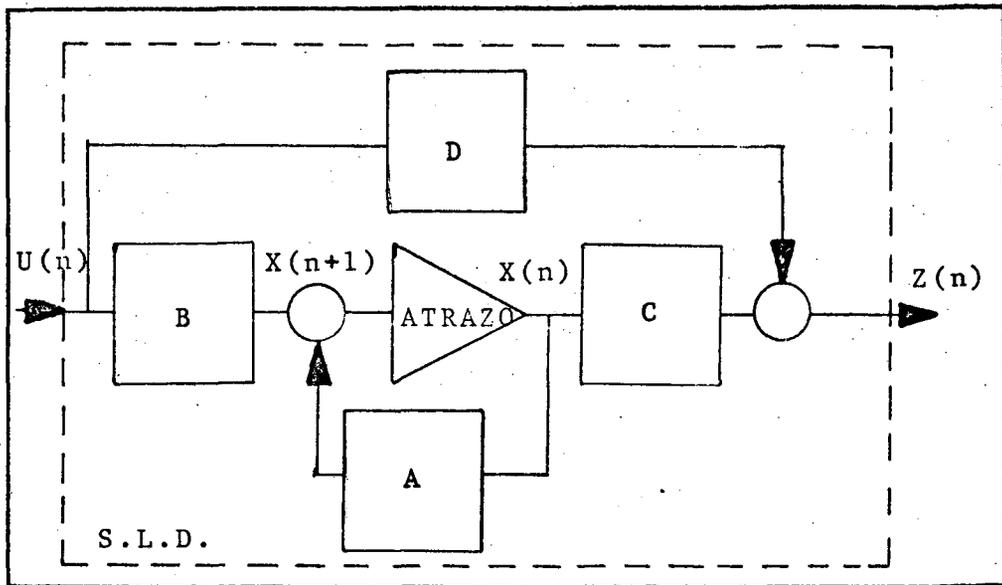


FIG. 9 - Sistema linear representativo do sistema de equações (41) e (42).

2.4 - Análise comparativa e aplicação do modelo

O sistema mostrado na figura 9 é uma representação do sistema linear discreto mais geral. Vamos efetuar algumas hipóteses sobre este modelo, de modo que possamos obter três dos mais conhecidos modelos lineares:

a) Modelo Dinâmico: Suponhamos que β_{n0} seja nulo, como ocorre normalmente nos sistemas físicos. Então:

$$b_{n0} = \beta_{n0} = 0$$

logo:

$$D = 0$$

e o sistema se reduz a:

$$X(n+1) = A \cdot X(n) + B' \cdot U(n)$$

$$Z(n) = C' \cdot X(n)$$

que é a forma do exemplo 1 da pág. 14 .

b) Modelo de Médias Móveis (Moving Average): Nesse caso vamos supor que:

$$a_{n2} = a_{n3} = \dots = a_{nm} = 0 \quad \text{e} \quad b_{n0} = 0$$

Com isso as equações (25) se tornarão:

$$\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \cdot \begin{bmatrix} \beta_{n0} \\ \beta_{n1} \\ \vdots \\ \beta_{nm} \end{bmatrix} = \begin{bmatrix} b_{n0} \\ b_{n1} \\ \vdots \\ b_{nm} \end{bmatrix} \quad \text{ou seja, } b_{ni} = \beta_{ni}, (i=1,2,\dots,m).$$

Da mesma forma as equações (24) se tornarão:

$$\left\{ \begin{array}{l} X(n+1) \\ Z(n) \end{array} \right. = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \cdot X(n) + \beta \cdot U(n) \quad (27)$$

$$Z(n) = C' \cdot X(n)$$

As equações (27) são representativa do método de previsão através de médias móveis.

c) Modelo de Exponencial Ponderada (Exponential Smoothing) :
Nesse modelo atribuímos mais peso às observações mais recentes e menos peso às observações mais remotas. Seja:

$$X(n+1) = \alpha [X(n) + (1-\alpha) X(n+1) + \dots + (1-\alpha)^n X(0)] + (1-\alpha)^{n+1} \cdot W(n) \quad (28)$$

a forma escalar das equações do método de suavização exponencial. Tomemos agora as equações (19) e (20) em sua forma escalar e calculemos seu desenvolvimento através do tempo

$$\left\{ \begin{array}{l} x(n+1) = A x(n) + B U(n) \\ x(n) = A x(n-1) + B U(n-1) \\ x(n-1) = A x(n-2) + B U(n-2) \\ \vdots \\ \vdots \\ \vdots \\ x(n-p) = A x(n-p-1) + B U(n-p-1) \end{array} \right.$$

Substituindo-se as $p+1$ equações acima na primeira, obtém-se sucessivamente:

$$\left\{ \begin{array}{l} x(n+1) = A x(n) + B.U(n) \\ x(n+1) = A^2 x(n-1) + AB.U(n-1) + B.U(n) \\ x(n+1) = A^3 x(n-2) + AB.U(n-2) + AB.U(n-1) + B.U(n) \\ \vdots \\ x(n+1) = A^{p-1} x(n-p) + AB.U(n-p) + \dots + AB.U(n-1) + B.U(n). \end{array} \right.$$

Somando membro a membro:

$$x(n+1) = 1/p [Ax(n) + A^2x(n-1) + \dots + A^{p-1}x(n-p)] + K_0.W(n) \quad (29)$$

onde $W(n)$ é uma função de $U(n)$ e K_0 é uma constante, função de A e B .

A equação (29) evidencia o fato de ser o modelo de suavização exponencial, um caso particular do modelo geral de um sistema linear discreto, como o expresso pela equação (28).

Como vimos, a técnica das variáveis de estado permite-nos obter uma representação unificada dos sistemas lineares discretos. Portanto, podemos afirmar que um processo estocástico $X(n)$ representa o comportamento de um sistema linear discreto excitado por um processo estocástico do tipo Ruido Branco, ou seja:

$$X(n+1) = A.X(n) + B.W(n) \quad (30)$$

$$x(n) = C.X(n) + D.W(n) \quad (31)$$

Como já dissemos em 2.4, ítem a., ocorre que na maioria dos sistemas físicos $D = 0$ e, portanto, desse ponto em diante consideraremos sempre a equação (31) como sendo:

$$x(n) = C.X(n) \quad (32)$$

Considere-se agora a equação (9) e a equação (30). Ambas são representativas da resposta de um sistema linear a uma excitação do tipo Ruido Branco, portanto, são equações equivalentes. Geralmente, na literatura ^{5,18} os autores assumem como modelo a equação (30). Como, porém, determinar quais são as classes de processos estocásticos que podem ser representados desta maneira? Em nosso caso, sabemos que somente aqueles processos que pertencem ao sub-espaço de Hilbert, em que as condições (6) e (7) são satisfeitas, admitem essa representação.

Outra grande vantagem da representação dos sistemas lineares pela equação (30) é a simplicidade que resulta de sua utiliza

ção como modelo para computação digital. Observe-se que o valor de $X(n+1)$ depende explicitamente somente de um dos valores anteriores da série de $X(n)$. Isso implica em que, para calcularmos $X(n+1)$ não necessitamos lembrar $X(0), X(1), \dots, X(n-1)$. Essa classe de modelos é chamada de modelos recursivos, enquanto que os modelos da forma da equação (3) são chamados de modelos lineares auto-regressivos.

Uma aplicação largamente difundida no campo da econometria dos modelos lineares auto-regressivos é o estudo dos processos de renda e consumo ou demanda ^{2,5,18}.

2.5 - Modelos de observação

Nos modelos até aqui estudados consideramos que os dados disponíveis fôsem representativos do exato comportamento dos sistemas. No entanto, é evidente que nos casos práticos isso não ocorre, pois os processos estocásticos raramente são passíveis de medida direta e sem erro. É importante, portanto, considerarmos as observações como portadoras de um certo erro.

De um modo geral, os valores observados Y_n são uma função g_n dos valores reais X_n e de uma variável aleatória V_n , i.é.,

$$Y_n = g_n(X_n, V_n)$$

Para mantermos a unidade de nosso trabalho, consideraremos que g_n seja uma função linear de X_n e V_n , embora seja evidente que não é esse o caso geral. Portanto, vamos admitir que:

$$Y_n = C_n \cdot X_n + V_n \quad (33)$$

No entanto, não podemos, a priori, assumir qual seja o comportamento estatístico de V_n .

Podemos classificar V_n , em relação à média estatística, em duas classes, a saber:

- a) V_n tem média nula
- b) V_n tem média não nula

A classe dos valores observados de média nula define o con

junto das observações não tendenciosas. É o caso em que mais satisfatoriamente se obtêm valores de Y_n . Quando, porém, V_n apresenta média não nula, ocorre um erro sistemático nas medidas, que, se detectado, deve ser eliminado.

Em nosso trabalho vamos admitir três hipóteses a respeito de V_n :

- a) V_n é normalmente distribuída com média nula e variância unitária.
- b) V_n é Laplacianamente distribuída com média zero.
- c) V_n distribui-se conforme a distribuição de Cauchy.

Conseqüentemente V_n poderá assumir as seguintes distribuições:

- 1) distribuição de Gauss:

$$f(v) = (2\pi)^{-1/2} \cdot e^{-v^2/2}, \quad \mu = 0 \quad e \quad \sigma^2 = 1$$

- 2) distribuição de Laplace:

$$f(v) = \alpha/2 \cdot e^{-\alpha|v|}, \quad \mu = 0 \quad e \quad \sigma^2 = 2/\alpha^2$$

- 3) distribuição de Cauchy:

$$f(v) = \frac{\alpha/\pi}{\sigma^2 + v^2}, \quad \mu = 0 \quad e \quad \sigma^2 = \infty$$

Façamos um estudo comparativo das três hipóteses sobre V_n considerando a figura 10.

É fácil verificar que a probabilidade dos valores de V_n diferirem da média aumenta a medida que a distribuição de V_n é sucessivamente gaussiana, laplaciana ou de Cauchy. Isto ocorre porque, para valores grandes de v , estas distribuições são respectivamente proporcionais a $\text{Exp}\{-v^2\}$, $\text{Exp}\{-v\}$, e $1/v^2$. Tal comportamento assintótico permite que os valores de v se afastem mais da média na distribuição de Cauchy que nas demais, e na distribuição de Laplace mais que na de distribuição de Gauss. Isso implica que a confiabilidade dos dados coletados de

pende fundamentalmente da função densidade de probabilidade do fator de erro V_n . Por isso, Y_n é a mais confiável se V_n for distribuída normalmente. Caso a distribuição de V_n for Laplaciana, os dados não serão tão confiáveis como anteriormente, embora $E[V^2] < \infty$. Porém, quando V_n for distribuída segundo Cauchy, teremos um conjunto de observações muito ruins, pois $E[V^2] = \infty$.

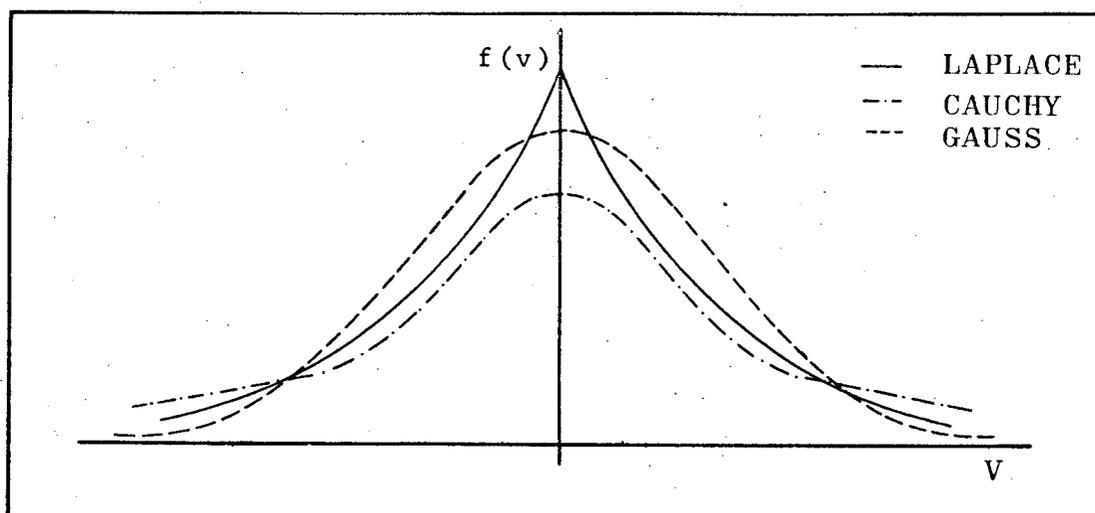


FIG. 10 - Representação comparativa das distribuições de Gauss, Laplace e Cauchy.

2.6 - Conclusão do Segundo Capítulo

Nesse capítulo expressamos processos estocásticos em função de uma base ortogonal de processos de Ruído Branco. Após isso, vimos como representar um processo estocástico como a resposta de um sistema linear discreto à excitação de um processo de Ruído Branco, e o expressamos usando o modelo das variáveis de estado.

A representação por meio das variáveis de estado é mostrada como sendo a forma mais geral de se representar um processo estocástico. Consideramos também as diferenças entre os modelos ideais e os modelos resultantes das observações, examinando-se três hipóteses sobre o comportamento do erro.

No entanto, o ponto mais importante de nosso trabalho, a

nosso ver, se refere à determinação das condições de validade dos modelos, as quais ficam perfeitamente definidas quando do estudo de carácter microscópico do sistema, que procuramos fazer através da definição de um processo estocástico por meio de uma base ortogonal de um sub-espaço de Hilbert.

C A P Í T U L O I I I

3. SIMULAÇÃO

3.1 - INTRODUÇÃO

A simulação é largamente utilizada no estudo de sistemas - ou organismos complexos. No estudo desses sistemas ou organismos, ocorre, em alguns casos, muita dificuldade de especificar a forma matemática exata do modelo do sistema, ou em outros casos, o sistema tem equações que não têm solução através de métodos analíticos. Nesse caso a simulação do sistema em computadores fornece dados sobre o comportamento do sistema que serão de grande valor.

Neste capítulo apresentamos a simulação que efetuamos com os modelos desenvolvidos no segundo capítulo. Nesse caso, a utilização da simulação se justifica pela estocasticidade dos modelos.

São apresentadas as técnicas que usamos para simular nosso modelo, usando linguagem FORTRAN em um computador IBM-1130 . São apresentados esquemas geradores de números aleatórios com distribuições de probabilidade uniforme, gaussiana, laplaciana e de Cauchy.

Mostramos gráficos, obtidos através dessas técnicas , que simulam o comportamento dos modelos dinâmicos e de observação.

O modelo de observação é discutido em função de três hipóteses assumidas com relação à distribuição de probabilidade da variável de erro, simulando os casos em que a confiabilidade dos dados é boa, muito boa e má.

3.2 - TESTES DE COMPORTAMENTO ESTATÍSTICO DOS GERADORES DE NÚMEROS ALEATÓRIOS.

Para confirmação das propriedades esperadas dos geradores de números aleatórios e suas distribuições de probabilidade, efetuamos três testes; os dois primeiros visam verificar a exatidão dos valores esperados das médias e variâncias das distribuições.

3.2.1 - Testes de Média e Variância

Como esses testes são amostrais, podemos escrever:

$$E [x] = \frac{1}{n} \sum_{i=1}^n x_i = \hat{\mu} = \bar{x}$$

$$E [x^2] = n \sigma_{\bar{x}}^2 = n \sum_{i=1}^n (x_i - \mu)^2 = \hat{\sigma}^2$$

Utilizamos estes testes em todas as distribuições de probabilidade que usamos e obtivemos resultados muito bons.

Para o teste de independência da sequência dos números aleatórios usamos o teste Chi-Quadrado, que é talvez o melhor teste estatístico dentre todos os conhecidos ^{4, 9} e é largamente utilizado por ser um teste básico e que pode ser usado em conexão com muitos outros.

3.2.2 - Teste de Chi-Quadrado

Para executarmos este teste tomamos uma amostra de n números. Divide-se o intervalo de variação dos números em K sub-intervalos iguais. A medida que os números aleatórios forem sendo gerados, verifica-se a qual dos K sub-intervalos ele pertence e anota-se a frequência de ocorrência de números em cada sub-intervalo. Adotamos essa frequência como sendo a frequência observada de ocorrência, f_{oi} . Calcula-se a frequência esperada de ocorrência de números aleatórios em cada sub-intervalo, e definimos:

$$V = \sum_{i=1}^K \frac{(f_{oi} - f_{ei})^2}{f_{ei}}$$

em que V é uma variável aleatória com distribuição aproximadamen-

te chi-quadrado, com $K-1$ graus de liberdade.

A frequência esperada é calculada por:

$$f_e = n \cdot p_i$$

onde n é o tamanho da amostra e p_i é a probabilidade de ocorrência de um número aleatório no sub-intervalo i , ou ainda:

$$p_i = P [a \leq x \leq b] = F(b) - F(a)$$

onde $F(a)$ é o valor da função densidade de probabilidade acumulada no ponto a ; e " a " e " b " são os valores extremos dos sub-intervalos.

Então, de uma forma geral, o valor da frequência esperada f_e para distribuições uniformes, será:

$$f_{e u} = \frac{n}{K}$$

Usamos este teste no gerador de números aleatórios uniformes RALEA ¹, mostrada na figura (13) e obtivemos resultados positivos com um nível de significância de 95%, conforme mostrado no exemplo 1.

Exemplo 1:

Seja uma população de 30.000 números aleatórios uniformemente distribuídos entre (0, 1). Tomamos 30 amostras de 1.000 elementos cada. Vamos determinar com 95% de confiabilidade se esses números são independentes e realmente uniformemente distribuídos. Para isso dividimos o intervalo (0, 1) em 100 sub-intervalos iguais. Então:

$$f_{eu} = \frac{n}{K} = \frac{1.000}{100} = 10$$

então, os valores s de V poderão estar entre:

$$V_{\max} = 128,422$$

e

$$V_{\min} = 73,361$$

pois V tem distribuição chi-quadrado com 99 graus de liberdade.

A figura 11 mostra um programa FORTRAN ¹³ para executar o exemplo 1.

```

      INTEGER F(100)
      DATA IX,JX,K,PI/1,3,100,3.14159/
      WRITE(3,12)
12    FORMAT(38X,'***  RELACAO DAS ESTADISTICAS CHI-QUI'
1'ADRADO  ***',//,33X,'***  VERIFICACAO DA UNIFOR'
2'MIDADE DO GERADOR RALEA  ***',//,41X,'***  NIVE'
3'L DE SIGNIFICANCIA 95 P.C.',//,33X,54('-',//)
      DO 5 M=1,33
      DO 10 N=1,K
10    F(N)=0.
      DO 40 J=1,1000
      U1=RALEA(IX)
      U2=PALEA(JX)
      THETA=2.*PI*U2
      PFL=((( -2)*ALOG(U1))**0.5)*COS(THETA)
      I=PFL*K+1
40    F(I)=F(I)+1
      QUI2=0.
      DO 50 I=1,K
50    QUI2=QUI2+F(I)**2
      QUI2=QUI2/(1000./FLOAT(K))-1000.
      IF(QUI2-73.361)25,26,26
26    IF(QUI2-128.422)27,27,28
25    WRITE(3,13)M,QUI2
13    FORMAT(52X,I6,F8.3,'  *')
      GO TO 5
28    WRITE(3,14)M,QUI2
14    FORMAT(52X,I6,F8.3,'  **')
      GO TO 5
27    WRITE(3,15)M,QUI2
15    FORMAT(52X,I6,F8.3)
5     CONTINUE
      CALL EXIT
      END

```

FIG. 11 - Programa teste Chi-Quadrado.

O gerador RALEA está programado em ASSEMBLER e foi gentilmente cedido para este trabalho pela autora, Professora Márcia Ra buske.

A função RALEA pode ser utilizada através dos seguintes comandos:

X = RALEA (IX)

ou

I = IALEA (IX)

conforme se queira valores respectivamente, reais ou inteiros.

```

// ASM
*LIST
0000      19053141 00001      ENT      RALEA
0023      09053141 00002      ENT      IALEA
0000 0    0000      00003      RALEA DC  *-*
0001 01  6E00001C 00004      STX     L2  SAVE+1
0003 01  6D00001E 00005      STX     L1  SAVI+1
0005 01  66800000 00006      LDX     I2  RALEA
0007 00  C6800000 00007      LD      I2  0
0009 0    A02D      00008      M       IRAIZ
000A 01  AC000036 00009      D       L   MOD
000C 0    1090      00010      SLT     16
000D 00  D6800000 00011      STO     I2  0
000F 0    1890      00012      SRT     16
0010 0    108F      00013      SLT     15
0011 01  AC000036 00014      D       L   MOD
0013 0    6110      00015      LDX     1  16
0014 0    1140      00016      SLCA    1
0015 0    1801      00017      SRA     1
0016 0    D021      00018      STO     TEMP
0017 0    7171      00019      MDX     1  113
0018 0    6920      00020      STX     1  TEMP+1
0019 01  74010000 00021      MDM     L   RALEA,1
001B 00  66000000 00022      SAVE   LDX  L2  *-*
001D 00  65000000 00023      SAVI   LDX  L1  *-*
001F 20  064C4000 00024      LIBF   FLD
0020 1    0038      00025      DC     TEMP
0021 01  4C800000 00026      B       I   RALEA
0023 0    0000      00027      IALEA DC  *-*
0024 00  6E000000 00028      STX     L2  SAVA+1
0026 01  66800023 00029      LDX     I2  IALEA
0028 00  C6800000 00030      LD      I2  0
002A 0    A00C      00031      M       IRAIZ
002B 01  AC000036 00032      D       L   MOD
002D 0    1090      00033      SLT     16
002E 00  D6800000 00034      STO     I2  0
0030 01  74010023 00035      MDM     L   IALEA,1
0032 00  66000000 00036      SAVA   LDX  L2  *-*
0034 01  4C800023 00037      B       I   IALEA
0036 0    7FED      00038      MOD    DC   32749
0037 0    03E5      00039      IRAIZ  DC   997
0038      0001      00040      TEMP  BSS  E   1
0039      0001      00041      BSS   1
003A      00042      END

```

FIG. 12 - Função RALEA

O programa mostrado na figura 11 executa esse teste para o gerador RALEA, obtendo-se os seguintes resultados:

*** RELACAO DAS ESTATISTICAS CHI QUADRADO ***
*** VERIFICACAO DA UNIFORMIDADE DO GERADOR RALEA ***
*** NIVEL DE SIGNIFICANCIA 95 P.C.

1	109.000
2	85.600
3	93.000
4	101.200
5	85.400
6	108.400
7	99.600
8	79.600
9	93.800
10	103.200
11	82.000
12	100.200
13	98.200
14	92.600
15	90.600
16	86.000
17	98.800
18	88.600
19	93.400
20	99.000
21	75.400
22	100.000
23	95.600
24	97.600
25	97.000
26	75.400
27	90.200
28	101.600
29	95.800
30	96.600
31	106.600
32	78.000
33	104.600

FIG. 13 - Resultado do teste Chi-Quadrado aplicado ao gerador RALEA.

3.3 - GERAÇÃO DE NÚMEROS ALEATÓRIOS COM DISTRIBUIÇÃO NORMAL(0,1):

Para obtenção de números aleatórios com distribuições não uniformes, usamos a técnica baseada na função inversa da função distribuição de probabilidade acumulada. Para o nosso gerador, arbitramos a média como zero e a variância unitária; portanto:

$$f(x) = \frac{1}{\sqrt{2\pi}} \text{Exp} \left\{ \frac{-x^2}{2} \right\}$$

A função distribuição de probabilidade acumulada $F(x)$ é dada por:

$$F(x) = \int_{-\infty}^x f(t) \cdot dt.$$

Como pode-se observar na figura 14, se dispormos da função inversa $F^{-1}(x)$, poderemos facilmente gerar um conjunto de números aleatórios com distribuição normal através de uma variável aleatória independente uniformemente distribuída entre (0, 1).

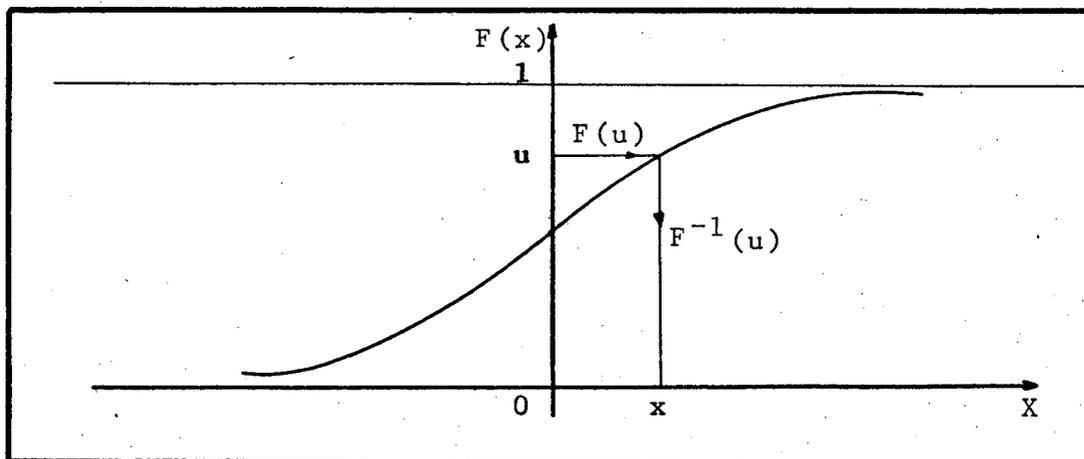


FIG. 14 - Método gráfico para obtenção de um gerador de números aleatórios normalmente distribuídos.

Ocorre, no entanto, que $F^{-1}(x)$ é muito difícil de ser obtida analiticamente. Existem muitos métodos aproximados para determinar $F^{-1}(x)$. Escolhemos o algoritmo de Box, Müller e Marsaglia⁹. O algoritmo de Box-Müller & Marsaglia se origina da integração da função densidade de probabilidade conjunta de duas variáveis gaussianas. Em vista disto necessita de duas sequências de números aleatórios uniformemente distribuídos. Para maior comodidade, programamos um gerador em tudo igual ao gerador RALEA, que chama

mos PALEA, para conjuntamente formarem as duas sequências de números aleatórios requeridas. Mostra-se na figura 15 um fluxograma - simplificado do algoritmo de Box-Müller e Marsaglia, e na figura 16 a sub-rotina GAUSS que programamos para gerar números aleatórios normalmente distribuídos com média zero e variância unitária.

A sub-rotina Gauss pode ser utilizada através do comando:

```
CALL GAUSS (IX, JX, U, V)
```

em que: IX e JX são variáveis inteiras positivas menores que 32748, que inicializam os geradores uniformes, e U e V são duas variáveis reais, estatisticamente independentes com distribuição normal (0, 1).

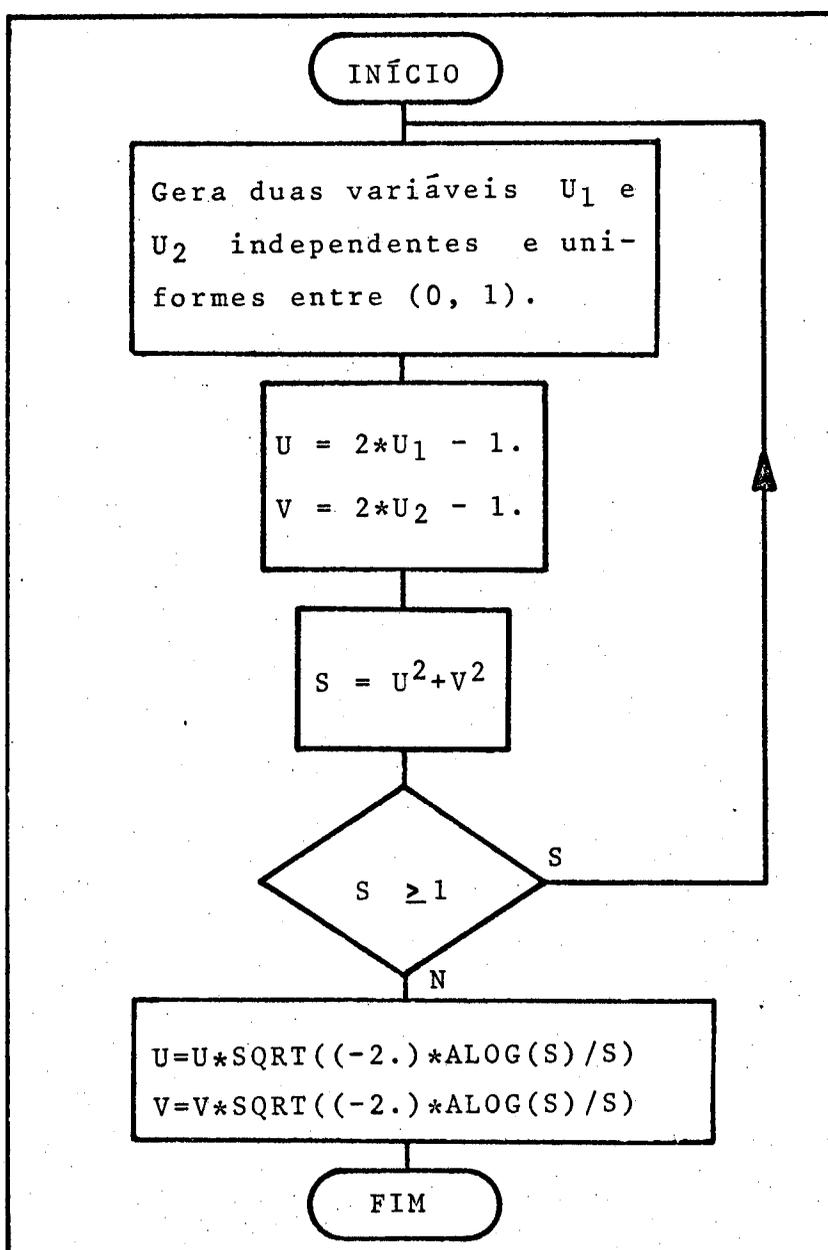


FIG. 15 - Algoritmo de Box, Müller e Marsaglia

```

// FOR
*LIST SOURCE PROGRAM
SUBROUTINE GAUSS(IX,JX,X1,X2)
1  U1=RALEA(IX)
   U2=PALEA(JX)
   V1=2.*U1-1.
   V2=2.*U2-1.
   S=V1**2.+V2**2.
   IF(S-1.)2,1,1
2  X1=V1*SQRT((-2.)*ALOG(S)/S)
   X2=V2*SQRT((-2.)*ALOG(S)/S)
   RETURN
   END

```

FIG. 16 - Sub-rotina GAUSS

3.4 - GERADOR DE NÚMEROS ALEATÓRIOS INDEPENDENTES COM DISTRIBUIÇÃO DE LAPLACE.

A função densidade de probabilidade de Laplace é:

$$f(x) = \frac{\alpha}{2} \cdot \text{Exp} \{-\alpha |x|\}$$

e tem a forma mostrada na figura 17.

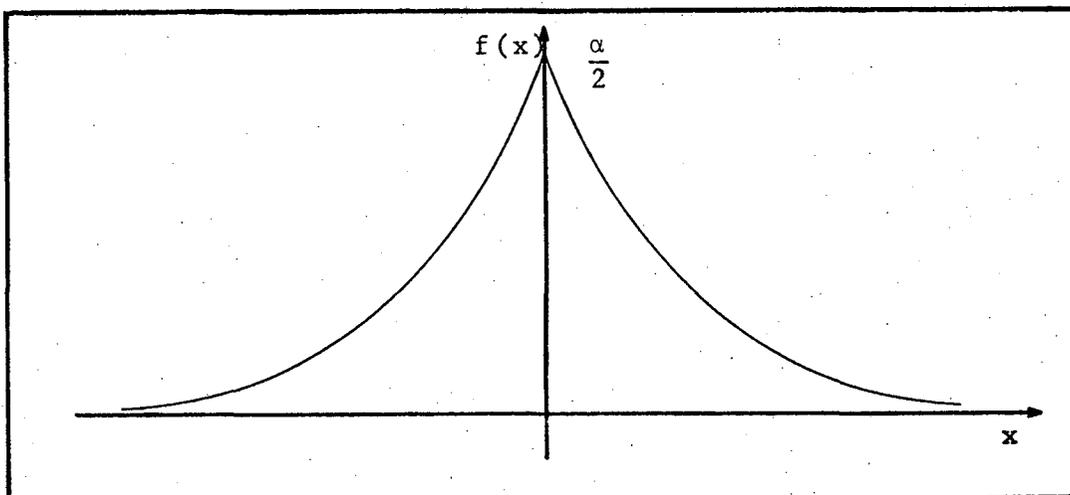


FIG. 17 - Distribuição de Laplace.

Adotamos $\alpha = 1$ e então teremos:

$$E [x] = 0$$

$$E [x^2] = \frac{2}{\alpha^2} = \sigma^2 ,$$

então com $\alpha = 1$, $\sigma^2 = 2$.

A função distribuição de probabilidade acumulada será:

$$F(x) \begin{cases} \int_{-\infty}^x \frac{\alpha}{2} e^{\alpha x} \cdot dx & , x \leq 0 \\ \int_{-\infty}^0 \frac{\alpha}{2} e^{\alpha x} \cdot dx + \int_0^x \frac{\alpha}{2} e^{-\alpha x} \cdot dx & , x > 0 \end{cases}$$

logo:

$$F(x) \begin{cases} 1/2 \cdot e^{\alpha x} & , x \leq 0 \\ 1 - 1/2 \cdot e^{-\alpha x} & , x > 0 \end{cases}$$

donde a função inversa $F^{-1}(x)$ será:

$$F^{-1}(x) \begin{cases} \frac{1}{\alpha} \ln(2 \cdot U) & , U \leq 1/2 \\ \frac{-1}{\alpha} \ln(-2 \cdot (U-1)) & , U > 1/2 \end{cases} \quad (34)$$

onde $U = F(x)$.

Usando as equações (34), programamos a sub-rotina LAPCE para gerar números aleatórios independentes com distribuição de Laplace. Para utilizar a sub-rotina LAPCE usa-se o comando:

CALL LAPCE (IX, X, ALFA)

em que IX é uma variável inteira positiva, menor que 32748, que inicializa o gerador uniforme. ALFA é o parâmetro da distribuição e X é a variável aleatória com distribuição de Laplace . O programa que usamos está na figura 18.

```

SUBROUTINE LAPCE(IX,X,ALFA)
U=RALEA(IX)
IF(U-0.5)1,1,2
1  X=ALOG(2.*U)/ALFA
   GO TO 10
2  X=-ALOG((-2.)*(U-1.))/ALFA
10 RETURN
END

```

FIG. 18 - Sub-rotina LAPCE.

3.5 - GERADOR DE NÚMEROS ALEATÓRIOS INDEPENDENTES COM DISTRIBUIÇÃO DE CAUCHY.

A distribuição de probabilidade de Cauchy tem a seguinte função densidade de probabilidade:

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{(1 + (x-\mu)^2)} \quad , \quad -\infty < x < \infty$$

com:

$$E [x] = \mu = 0$$

$$E [x^2] = \sigma^2 = \infty$$

A função distribuição de probabilidade acumulada é:

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \text{arc tg } (x-\mu)$$

Tomando-se $F(x) = U$ e $\mu = 0$, obtêm-se:

$$F^{-1}(x) = \text{tg} \left(\pi (U-1/2) \right)$$

A sub-rotina Cauch, figura 19, utiliza a equação (35) para gerar números aleatórios independentes com distribuição de Cauchy. Esse gerador pode ser chamado pelo comando:

```
CALL CAUCH (IX, X)
```

onde IX é uma variável inteira positiva menor que 32748 que inicializa o gerador de distribuição uniforme, e X é o número aleatório gerado.

```
// FOR
*LIST SOURCE PROGRAM
SUBROUTINE CAUCH(IX,X)
PI=3.14159
U=PALEA(IX)
ARCO=PI*(U-0.5)
X=SIN(ARCO)/COS(ARCO)
RETURN
END
```

FIG. 19 - Sub-rotina CAUCH.

Assim ficam definidos os quatros geradores de números aleatórios que usaremos em nosso trabalho de simulação.

3.6 - SIMULAÇÃO DOS MODELOS

Nessa simulação usamos os modelos fornecidos pela equação (38), desenvolvidos em forma escalar. Consideramos somente dois modelos diferentes:

$$x(n+1) = a \cdot x(n) + b \cdot w(n)$$

$$x(n+1) = a_1 \cdot x(n) + a_2 \cdot x(n-1) + b_1 \cdot w(n)$$

onde a , a_1 , a_2 , b e b_1 são constantes e w é uma variável aleatória que responde pelo comportamento estocástico dos modelos.

Simulamos o primeiro modelo através da sub-rotina PARPO (figura 22), e o segundo modelo através da sub-rotina PARSO (figura 23). Essas sub-rotinas usam o Plotter⁶ do sistema IBM-1130 para traçar o comportamento dos modelos.

O princípio básico das sub-rotinas está mostrado no diagrama de blocos mostrado na figura 20. Ambas as sub-rotinas PARPO e PARSO usam duas outras sub-rotinas que são as sub-rotinas GRAFC e GRADE (figuras 23 e 24), como auxiliares no traçado. A sub-rotina GRAFC define as escalas e traça os eixos de referência dos gráficos. A sub-rotina GRADE traça as margens e escreve os parâmetros do modelo. Evidentemente, além dessas sub-rotinas, fizemos largo uso das sub-rotinas do plotter do sistema IBM-1130.

A sub-rotina PARPO pode ser usada através do comando:

```
CALL PARPO (A1, B, XO, IX, JX, VET)
```

onde:

A1 = é o coeficiente de $x(n)$

B = é o coeficiente de $w(n)$

XO = é o valor inicial do processo, fornecido pelo programa principal.

IX, JX = são variáveis inteiras positivas menores que \dots 32748, que inicializa as funções geradoras de números aleatórios.

VET = é um vetor de dimensão igual ao tamanho da amostra mais um. Deve ser dimensionado no programa principal. É usado para simular os modelos de observação.

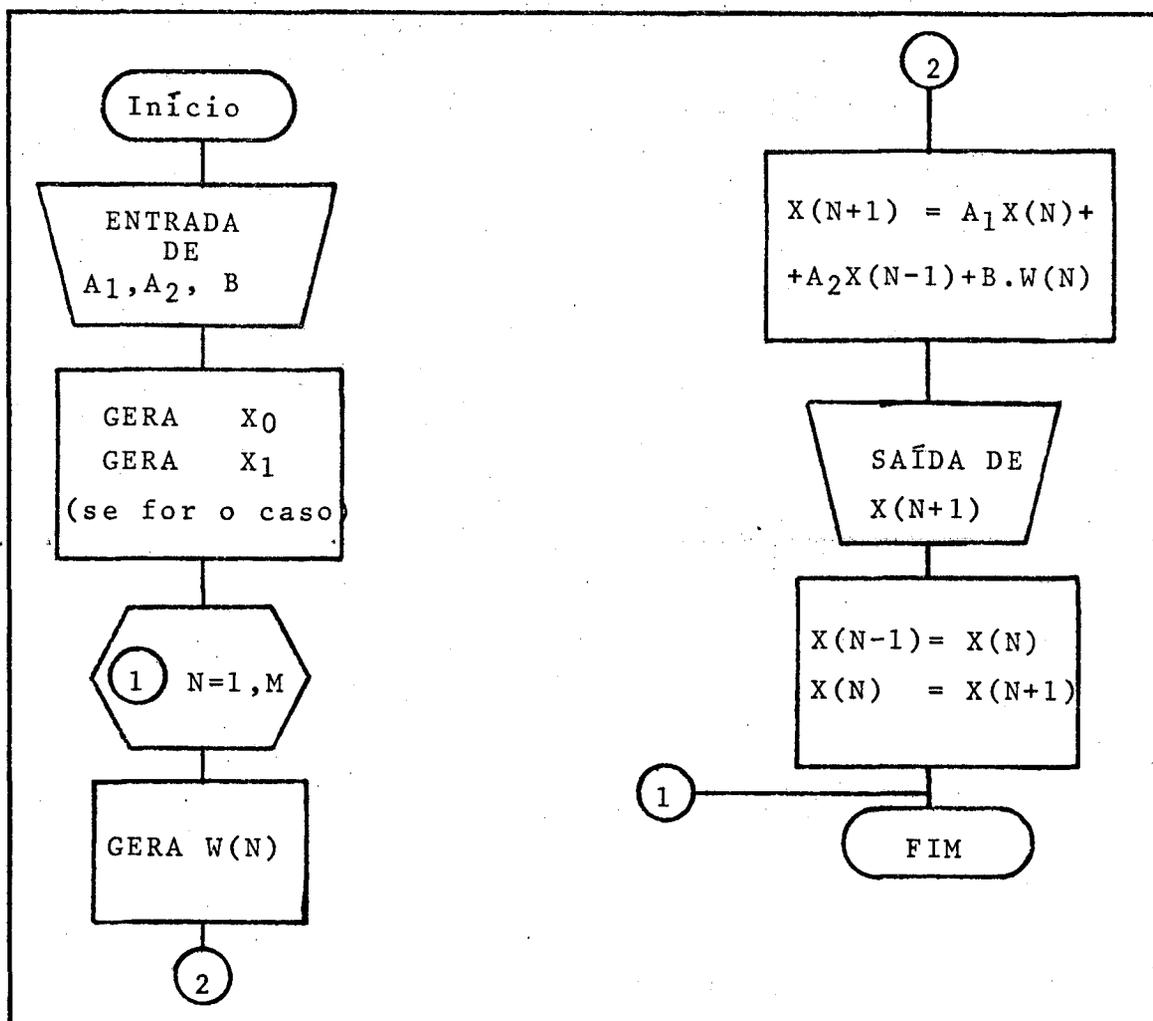


FIG. 20 - Esquema de simulação dos modelos.

A sub-rotina PARSO é chamada por:

CALL PARSO (A1, A2, B, X0, X1, IX, JX, VET)

em que os parâmetros são os mesmos da anterior, com exceção de:

A2 = é o coeficiente de $x(n-1)$

```

SUBROUTINE PARPO(A1,B,XO,IX,JX,VET)
DIMENSION VET(101)
VET(1)=XO
M=0
N=100
WRITE(3,11)A1,B,XO
WRITE(3,12)
CALL GRAFC
CALL FPLOT(-2,0.,XO)
WRITE(3,43)M,WO,XO
43  FORMAT(39X,I3,20X,F6.3,3X,F6.3)
DO 100 I=1,N
CALL GAUSS(IX,JX,XV1,XV2)
WO=XV1
X1=A1*XO+B*WO
VET(I+1)=X1
T=FLOAT(I+1)
WRITE(3,44)I,WO,XO,X1
51  CALL FPLOT(2,T,X1)
XO=X1
100 CONTINUE
CALL FPLOT(1,108.,-4.)
11  FORMAT(1H1,50X,'VALOR DO PARAMETRO A',F6.3,/,51X
1,'VALOR DO PARAMETRO B',F6.3,/,51X,'VALOR INICI'
2'AL XO',4X,F6.3,/)
12  FORMAT(45X,'DADOS PARA O TRACADO DO GRAFICO',/,3
18X,'TEMPO',20X,'W(N)      X(N)      X(N+1)',/)
44  FORMAT(39X,I3,9X,8X,3X,F6.3,3X,F6.3,3X,F6.3)
RETURN
END

```

FIG. 21 - Sub-rotina PARPO.

```

SUBROUTINE PARSO(A1,A2,B,X0,X1,IX,JX,VET)
DIMENSION VET(101)
WRITE(3,110)A1,A2,B,X0,X1
WRITE(3,120)
VET(1)=X1
M=0
N=100
CALL GRAFC
CALL FPLOT(-2,0.,X1)
WRITE(3,40)M,X0,X1
DO 100 I=1,N
CALL GAUSS(IX,JX,XV1,XV2)
WO=XV1
XN=A1*X1+A2*X0+B*WO
VET(I+1)=XN
WRITE(3,39)I,WO,X0,X1,XN
T=FLOAT(I)
CALL FPLOT(2,T,XN)
XO=X1
100 X1=XN
CALL FPLOT(1,108.,-4.)
39  FORMAT(31X,I3,20X,F6.3,4X,F6.3,4X,F6.3,3X,F6.3)
40  FORMAT(31X,I3,40X,F6.3,3X,F6.3)
110 FORMAT(1H1,50X,'VALOR DO PARAMETRO A1',F6.3,/,/,51
1X'VALOR DO PARAMETRO A2',F6.3,/,/,51X,'VALOR DO P'
2'ARAMETRO B',F7.3,/,/,51X,'VALOR INICIAL X0',7X,F6
3.3,/,/,51X,'VALOR INICIAL X1',7X,F6.3,/,/,)
120 FORMAT(45X,'DADOS PARA O TRACADO DO GRAFICO',/,/,3
10X,'TEMPO',19X,'W(N)      X(N-2)      X(N-1)      X(N)'
2,/)
RETURN
END

```

FIG. 22 - Sub-rotina PARSO.

```

SUBROUTINE GRAFC
PI=3.14159
SX=13./100.
SY=7.2/6.
CALL SCALF(SX,SY,0.,0.)
CALL FGRID(0,0.,0.,1.,100)
CALL FGRID(1,0.,0.,1.,3)
CALL FGRID(3,0.,0.,1.,3)
T=3.*PI/2.
XD=0.050/SX
Y=-(1./6.)
X=10.
DO 10 I=10,100,10
CALL FCHAR(X-XD,Y,.10,.10,T)
WRITE(7,4)I
X=X+10.
10 CONTINUE
T=0.
Y=-3.
X=-0.55/SX
XD=.05/SY
DEL=1.
DO 20 I=1,7
CALL FCHAR(X,Y-XD,.10,.10,T)
WRITE(7,5)Y
Y=Y+DEL
20 CONTINUE
5 FORMAT(F5.1)
4 FORMAT(I3)
RETURN
END

```

FIG. 23 - Sub-rotina GRAFC.

```

SUBROUTINE GRADE(A1,A2,B)
XO=3./25.4
YO=5./25.4
CALL SCALF(1.,1.,XO,YO)
CALL FPLLOT(-2,0.,0.)
CALL FPLLOT(0,0.,10.)
CALL FPLLOT(0,14.,10.)
CALL FPLLOT(0,14.,0.)
CALL FPLLOT(0,0.,0.)
XD=1./5.
XS=12.-XD
YS=XD
XR=14.-XD
YR=2.+XD
CALL FPLLOT(1,XS,YS)
CALL FPLLOT(2,XS,YR)
CALL FPLLOT(0,XR,YR)
CALL FPLLOT(0,XR,YS)
CALL FPLLOT(0,XS,YS)
XN=XS+XD
YN=YR-XD
CALL FCHAR(XN,YN,0.15,0.15,0.)
WRITE(7,9)
9 FORMAT(1X,'PARAMETROS')
XN=XN+3.*.15
YN=YN-2.*XD
CALL FCHAR(XN,YN,0.10,0.10,0.)
WRITE(7,8)A1
8 FORMAT('A1 = ',F4.2)
YN=YN-XD
CALL FCHAR(XN,YN,0.10,0.10,0.)
WRITE(7,7)A2
7 FORMAT('A2 = ',F4.2)
YN=YN-XD
CALL FCHAR(XN,YN,0.10,0.10,0.)
WRITE(7,6)B
6 FORMAT(' B = ',F4.2)
CALL FPLLOT(1,0.5,5.)
CALL FPLLOT(2,0.5,5.)
RETURN
END

```

FIG. 24 - Sub-rotina GRADE.

3.7 - SIMULAÇÃO DOS MODELOS DE OBSERVAÇÃO.

Simulamos os modelos de observação segundo a equação (33)-na sua forma escalar. A simulação dos modelos de observação é feita com a sub-rotina OBVS, figura 25; a chamada da sub-rotina OBVS é executada pelo comando:

```
CALL OBVS (VET, K, A1, A2, B, IX, JX)
```

com

VET= o mesmo vetor referenciado em PARPO e PARSO.

K = constante inteira que define a distribuição da variável de erro conforme o código:

```
K = 1 - distribuição de Gauss
K = 2 - distribuição de Laplace
K = 3 - distribuição de Cauchy
```

Os demais parâmetros já foram anteriormente definidos.

```

SUBROUTINE OBSV(VET,K,A1,A2,B,IX,JX)
DIMENSION VET(101)
CALL GRADE(A1,A2,B)
CALL FCHAR(12.,0.75,0.15,0.15,0.0)
GO TO (10,11,12),K
10 WRITE(7,13)
GO TO 30
11 WRITE(7,14)
GO TO 30
12 WRITE(7,17)
30 CALL FPLLOT(1,0.5,5.)
CALL FPLLOT(2,0.5,5.)
CALL GRAFC
DO 15 I=1,101
GO TO (1,2,3),K
1 CALL GAUSS(IX,JX,Y,X)
GO TO 9
2 CALL LAPCE(IX,Y,1.0)
GO TO 9
3 CALL CAUCH(IX,Y)
9 OBS=VET(I)+Y
T=FLOAT(I-1)
IF(I-1)31,31,15
31 CALL FPLLOT(-2,T,OBS)
15 CALL FPLLOT(0,T,OBS)
17 FORMAT(1X,'CAUCHY')
14 FORMAT('LAPLACE')
13 FORMAT(1X,'GAUSS')
RETURN
END

```

FIG. 25 - Sub-rotina OBVS.

Nas figuras que se seguem mostramos os resultados da simulação dos modelos. Estas figuras representam três amostras de um processo de Ruído Branco, de processos de auto-regressão de primeira e segunda ordens, correspondendo respectivamente às figuras (27) a (29), (30) a (32) e (33) a (35).

Por último estão as figuras representativas dos modelos de observação, em que cada conjunto de duas figuras mostra o modelo ideal e o modelo com erro. São três conjuntos mostrando o modelo de observação com função distribuição de probabilidade da variável de erro sucessivamente, gaussiana, figuras (36) e (37), laplaciana, figuras (38) e (39), e de Cauchy, figuras (40) e (41).

Para simular o processo de Ruído Branco plotamos um processo auto-regressivo de primeira ordem com coeficiente A_1 nulo e $B = 1,0$. Para a simulação dos modelos de primeira ordem adotamos como parâmetros:

$$A_1 = 0,50$$

$$B = 1,00$$

enquanto que para os modelos de segunda ordem adotamos:

$$A_1 = 0,50$$

$$A_2 = 0,25$$

$$B = 1,00$$

O programa principal que executa toda a simulação está reproduzido na figura 26. A simulação do processo de Ruído Branco é feita tomando-se A_1 e A_2 nulos.

3.8 - CONSIDERAÇÕES E OBSERVAÇÕES SOBRE ESTA SIMULAÇÃO.

Nos modelos dinâmicos somente os modelos de primeira e segunda ordem foram considerados, porque queríamos mostrar a vinculação, como efetivamente observa-se nos gráficos, existente entre um processo estocástico e um modelo estocástico.

Como era de se esperar, a vinculação entre os valores mais antigos da série com o comportamento atual do processo, a cada instante, aumenta à medida que mais dados entram na formação do valor atual do processo. Esta vinculação se apresenta sob a forma de uma "suavização" no comportamento dos processos. Lembrando

a atuação do sistema linear sobre um processo estocástico, visto no capítulo anterior, este comportamento se justifica por que a solução da equação do sistema é uma exponencial decrescente.

Nos modelos de observação a simulação mostra a influência da distribuição de probabilidade da variável de erro no comportamento do modelo com dados incorretos.

```

// FOR
*LIST SOURCE PROGRAM
*IOCS(CARD,TYPEWRITER,KEYBOARD,1132 PRINTER,PLOTTER)
  DIMENSION VET(101)
C   INICIA A SIMULACAO DO MODELO DINAMICO
    A1=0.50
    A2=0.0
    B=1.0
    IX=2
    JX=7
    DO 22 II=1,3
      CALL GAUSS(IX,JX,V1,V2)
      XO=V1
C   GERA O VALOR INICIAL XO
      WRITE(1,1)
      PAUSE
C   ESCRIVE MENSAGEM E ESPERA
      CALL GRADE(A1,A2,B)
C   TRACA MARGEM
      CALL PARPO(A1,B,XO,IX,JX,VET)
C   EXECUTA SIMULACAO DO PROCESSO DE OBSERVACAO
      K=II
      CALL OBSV(VET,K,A1,A2,B,IX,JX)
C   EXECUTA SIMULACAO DO MODELO DE OBSERVACAO DEPENDENDO
C   DO VALOR DE K CONFORME O CODIGO
C   SE K=1, VR TEM DISTRIBUICAO DE GAUSS
C   SE K=2, VR TEM DISTRIBUICAO DE LAPLACE
C   SE K=3, VR TEM DISTRIBUICAO DE CAUCHY
22  CONTINUE
C   INICIA SIMULACAO DE MODELO DINAMICO DE SEGUNDA ORDEM
    A1=0.50
    A2=0.25
    B=1.0
    IX=11
    JX=13
    DO 90 JJ=1,3
      CALL GAUSS(IX,JX,V1,V2)
      XO=V1
      X1=V2
C   GERA VALORES INICIAIS XO E X1
      WRITE(1,1)
      PAUSE
C   ESCRIVE MENSAGEM E ESPERA
      CALL GRADE(0.50,0.25,1.0)
      CALL PARPO(A1,A2,B,XO,X1,IX,JX,VET)
      K=JJ
      CALL OBSV(VET,K,A1,A2,B,IX,JX)
90  CONTINUE
1   FORMAT(1X,'POSICIONE A PENA DO PLOTTER'
1'NO CANTO INFERIOR DIREITO DO PAPEL',/,
21X,'APORTE A TECLA START PARA PROSSEGU'
3'IR O PROGRAMA')
      CALL EXIT
      END

```

FIG. 26 - Programa principal de simulação.

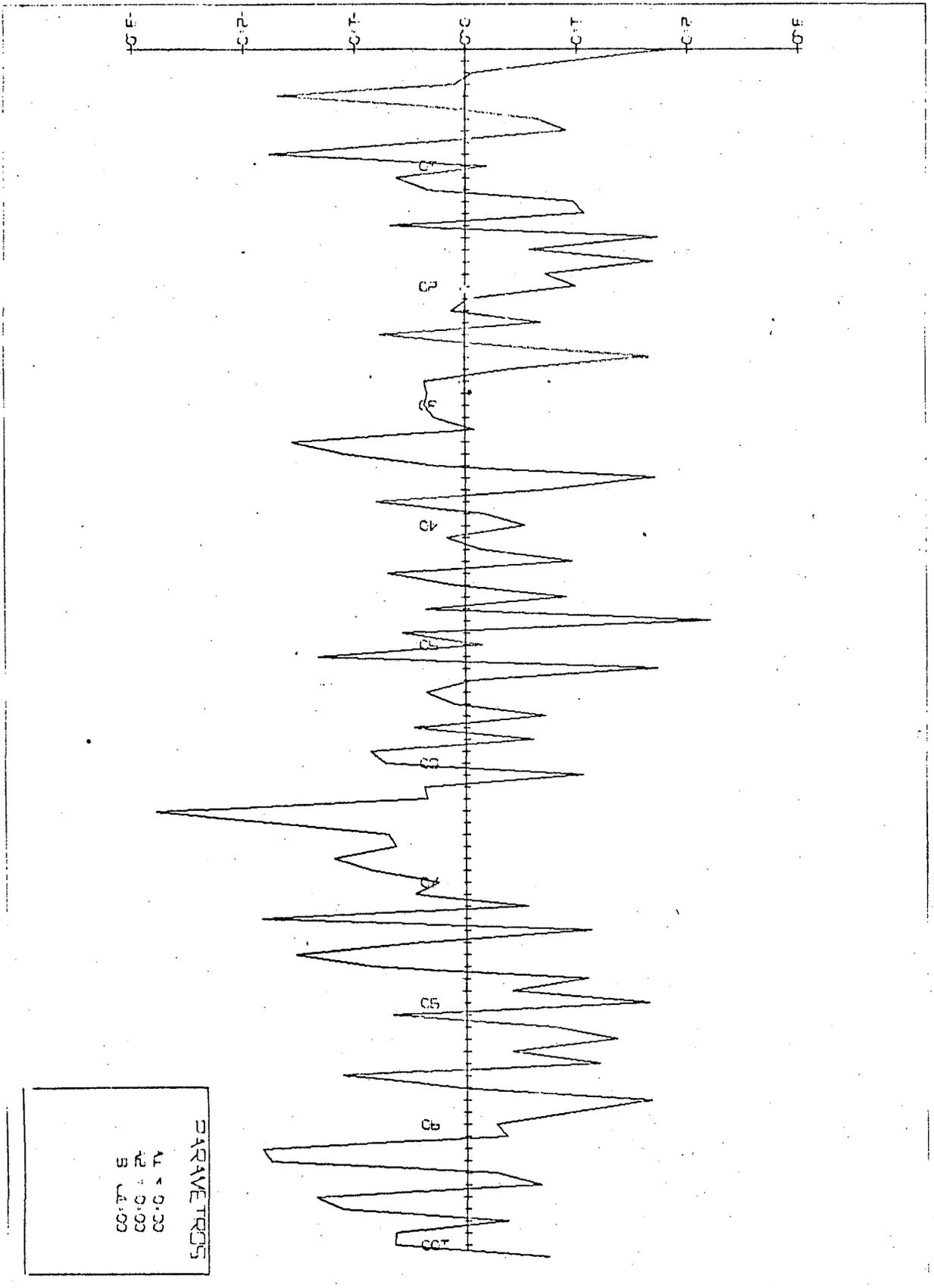


FIG. 28 - Simulação de um processo de Ruido Branco (Experiência 2).

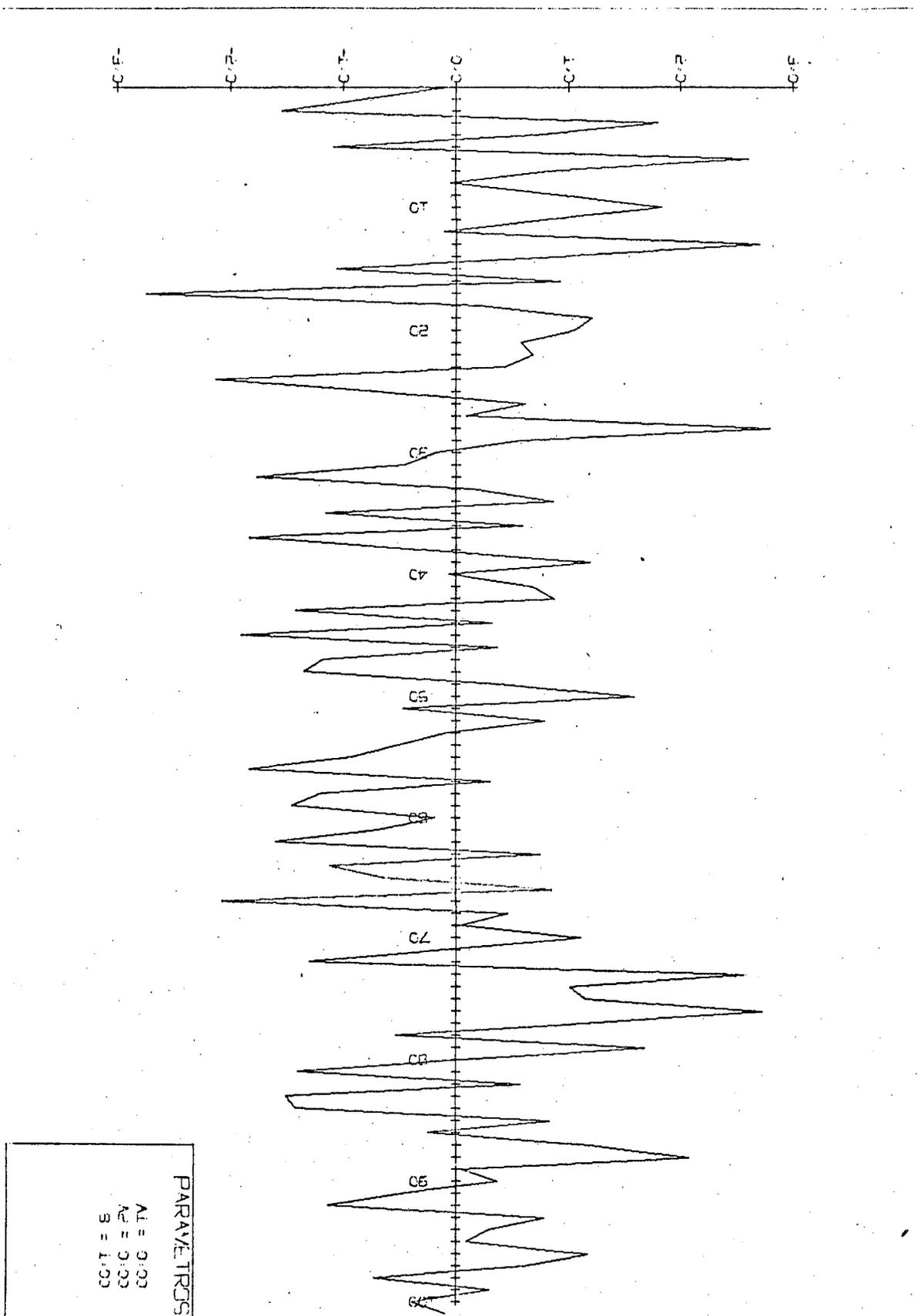


FIG. 29 - Simulação de um processo de Ruído Branco (Experiência 3).

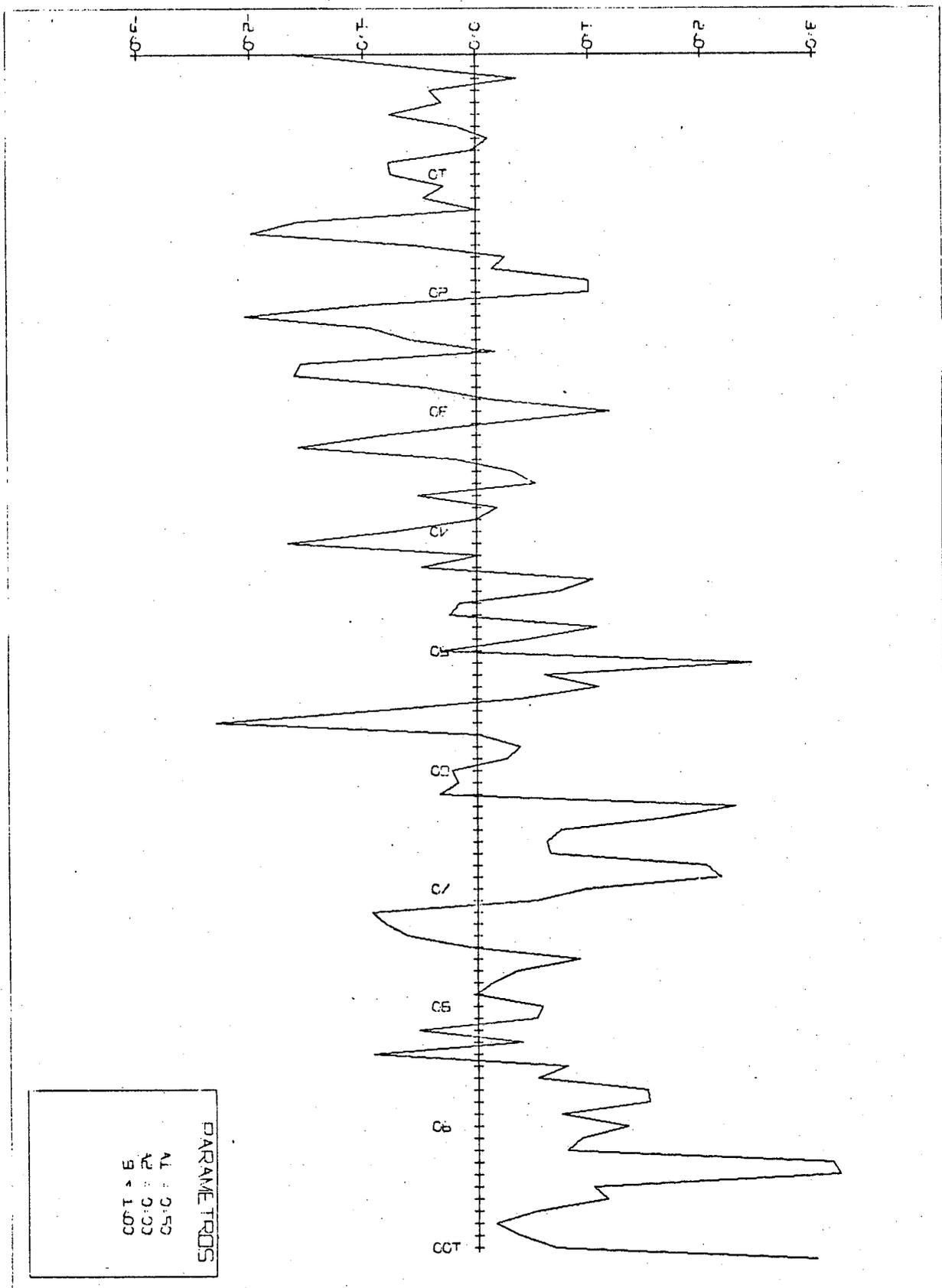


FIG. 30 - Simulação do modelo de primeira ordem (Experiência 1).

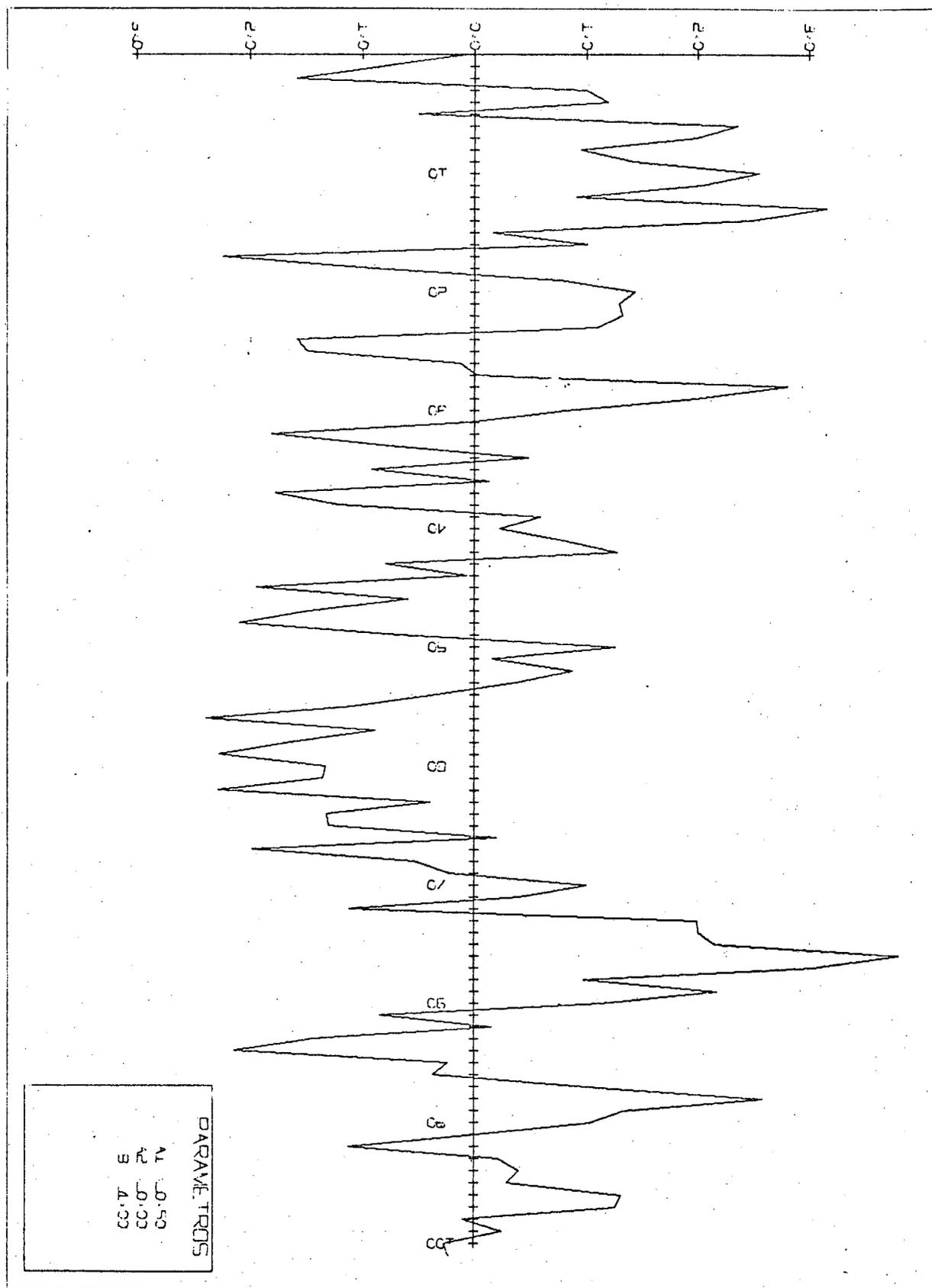


FIG. 31 - Simulação do modelo de primeira ordem (Experiência 2).

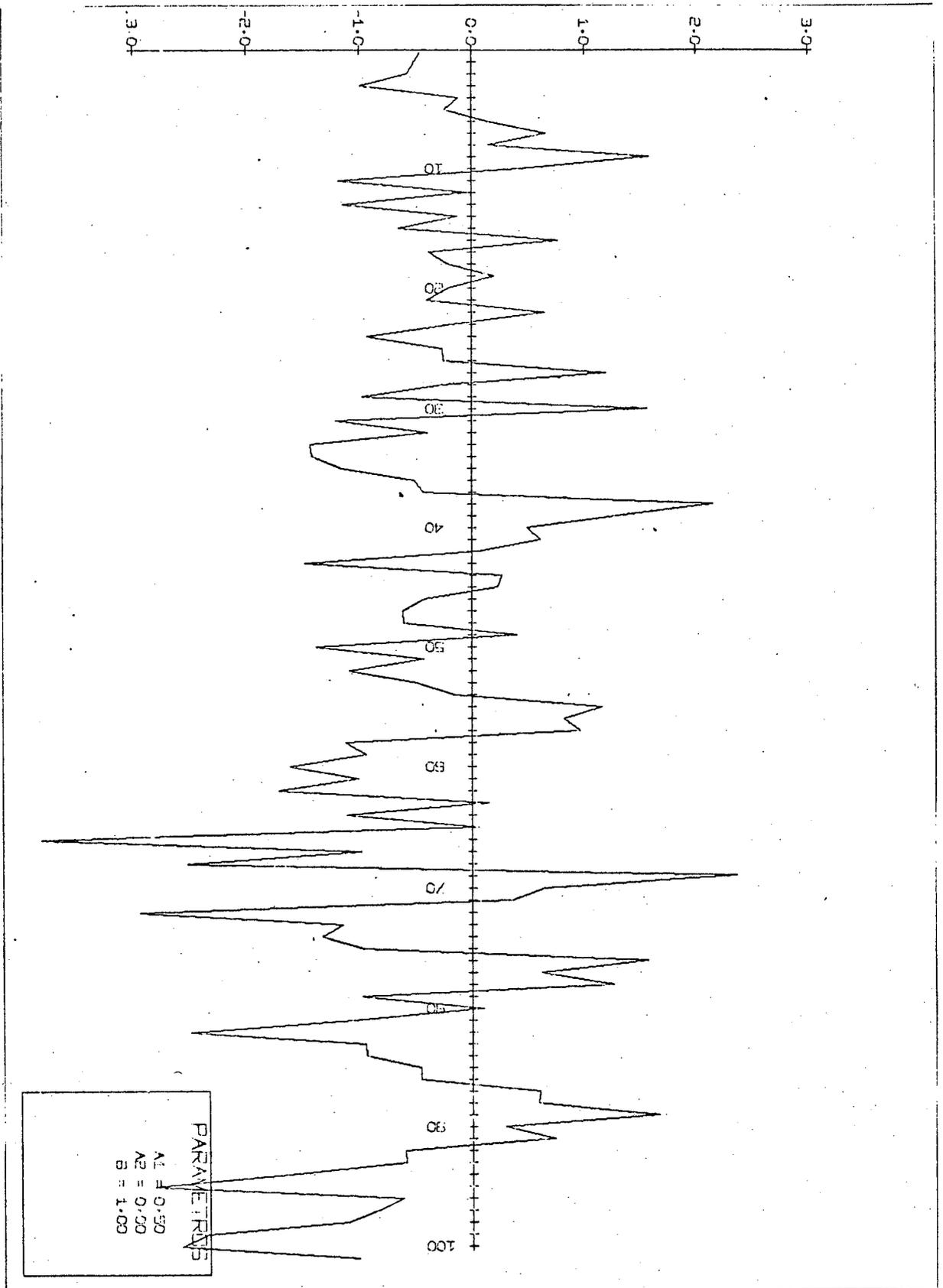


FIG. 32 - Simulação do modelo de primeira ordem (Experiência 3).

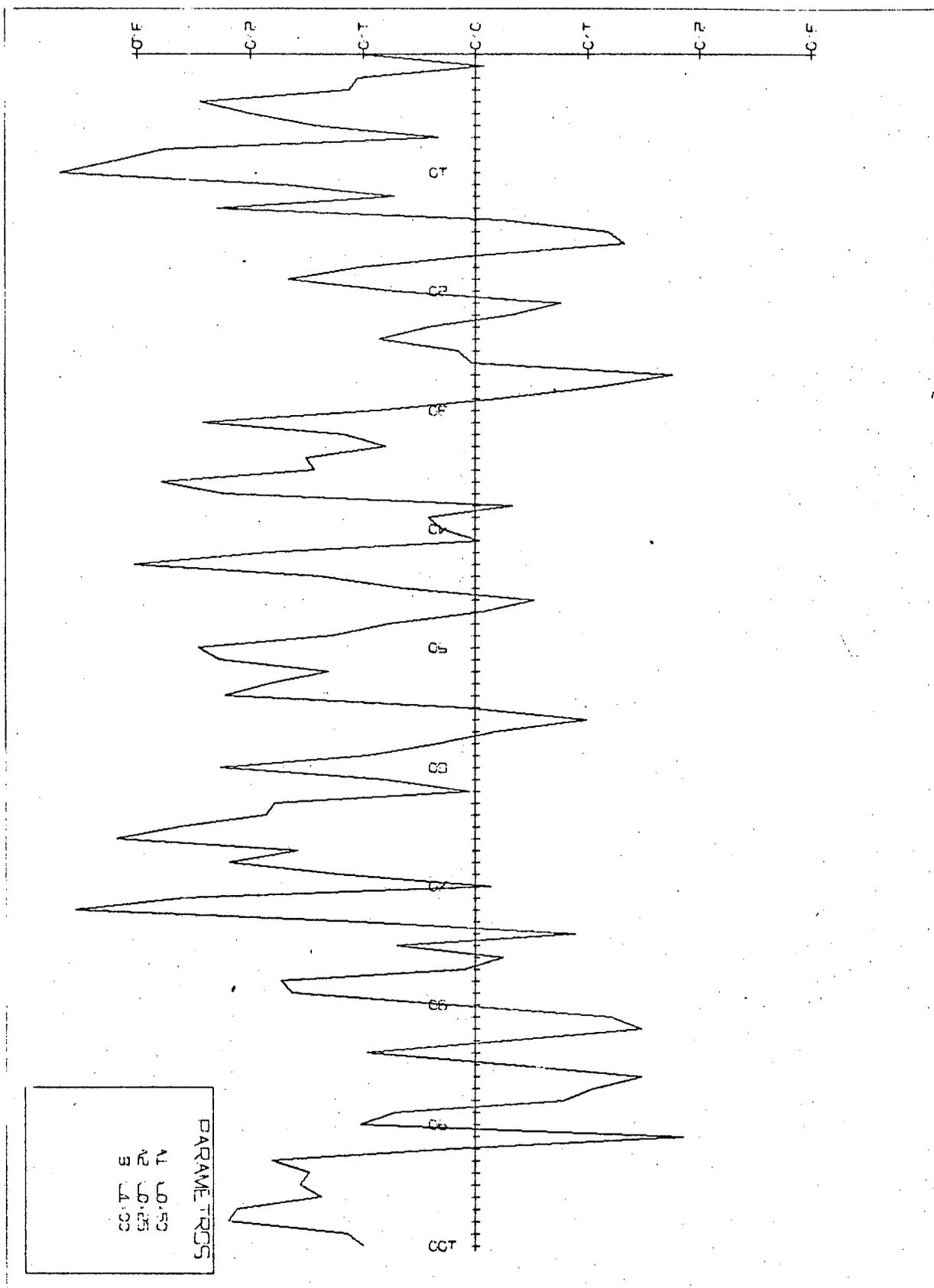


FIG. 33 - Simulação do modelo de segunda ordem (Experiência 1).

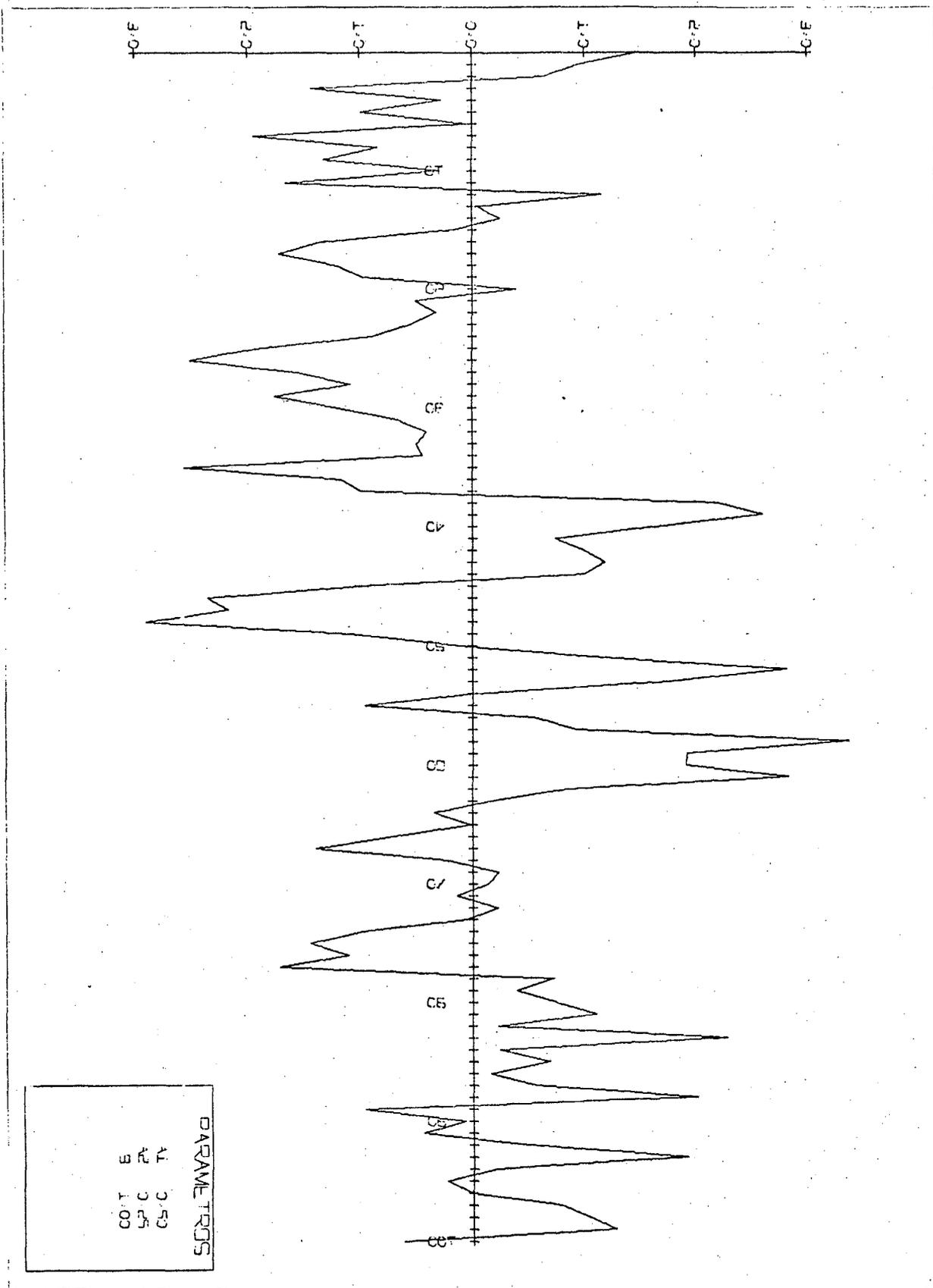


FIG. 34 - Simulação do modelo de segunda ordem (Experiência 2).

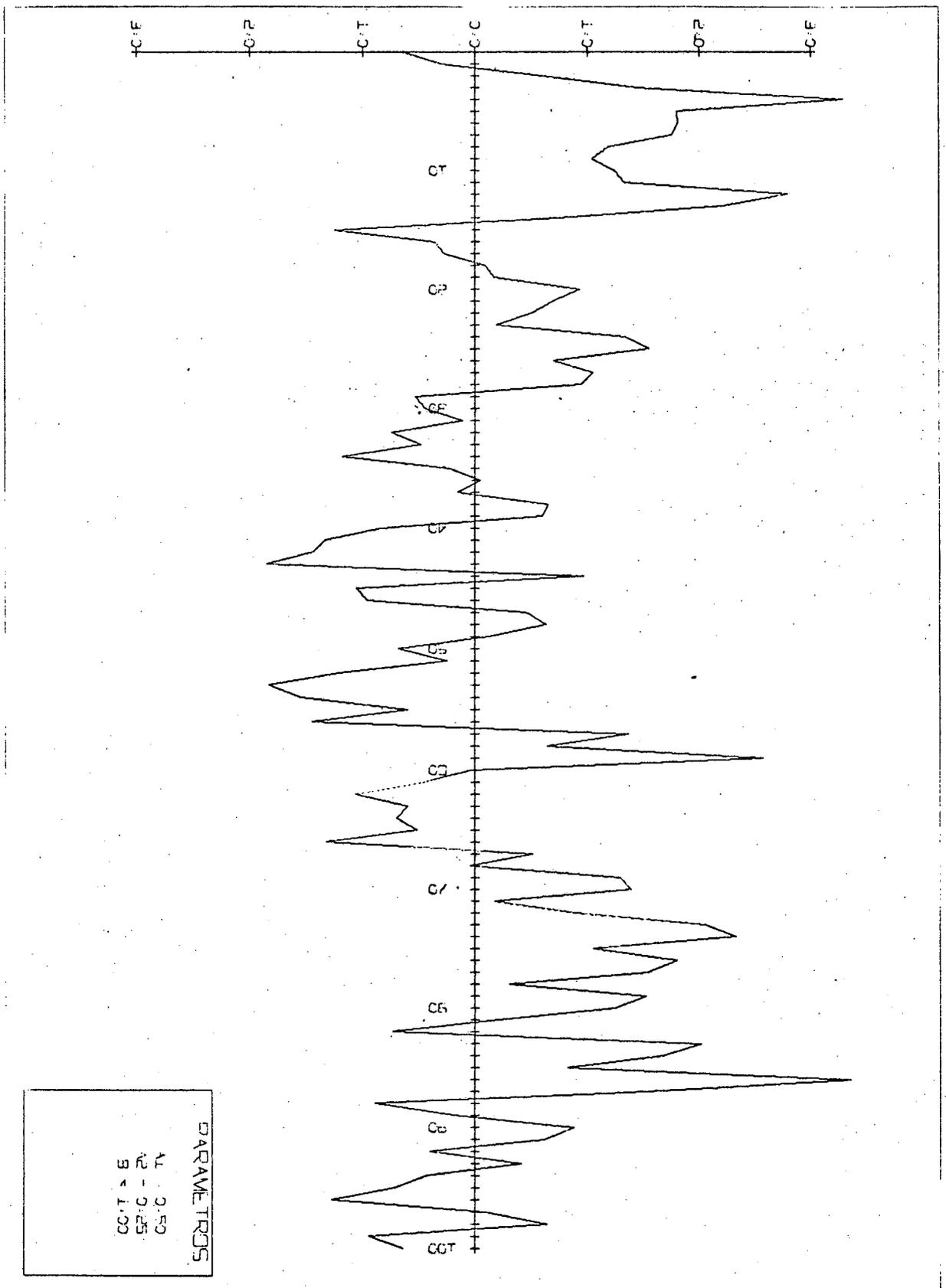


FIG. 35 - Simulação do modelo de segunda ordem (Experiência 3).

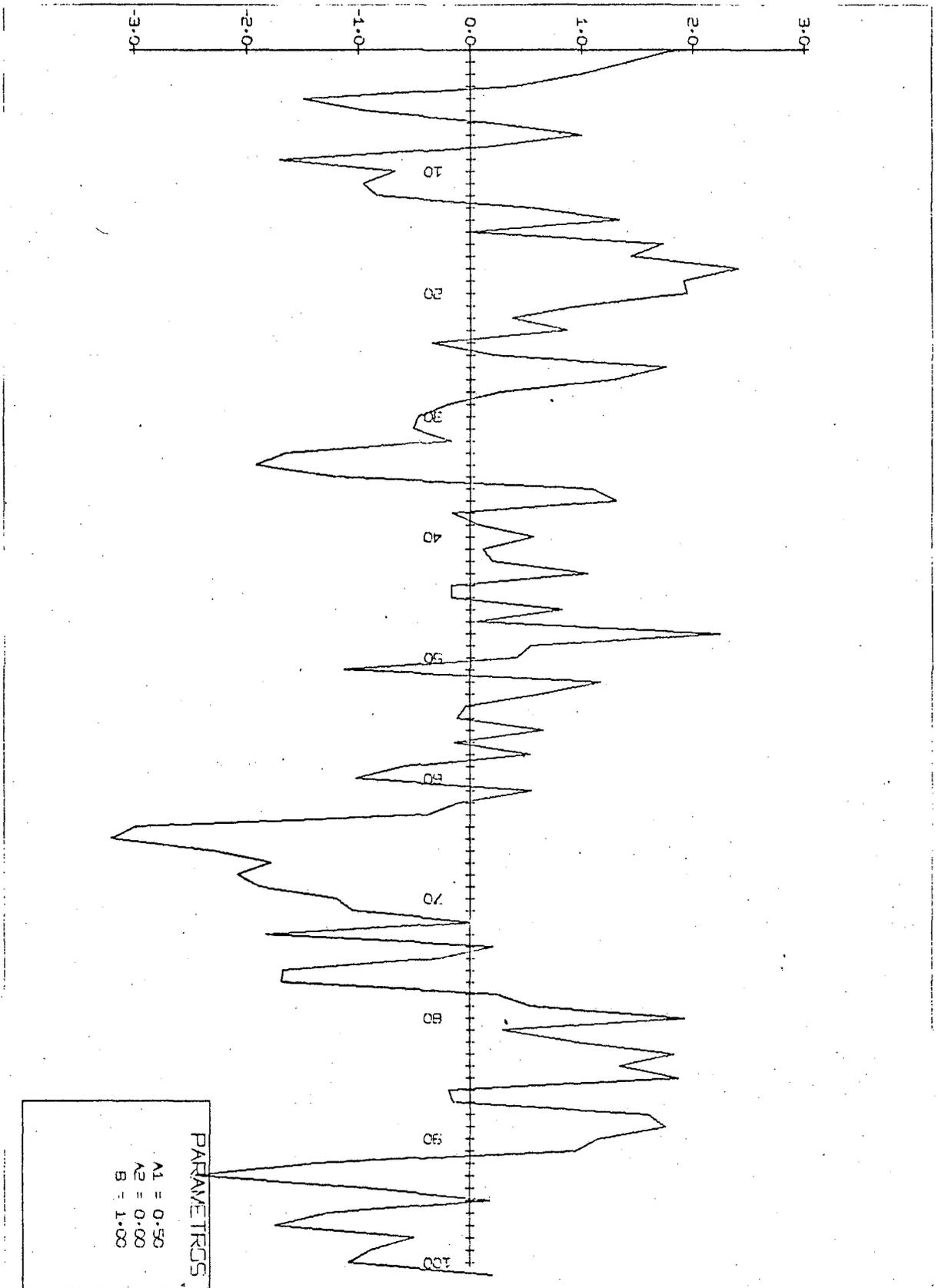


FIG. 36 - Simulação do modelo de observação
(Valor exato - Experimento 1)

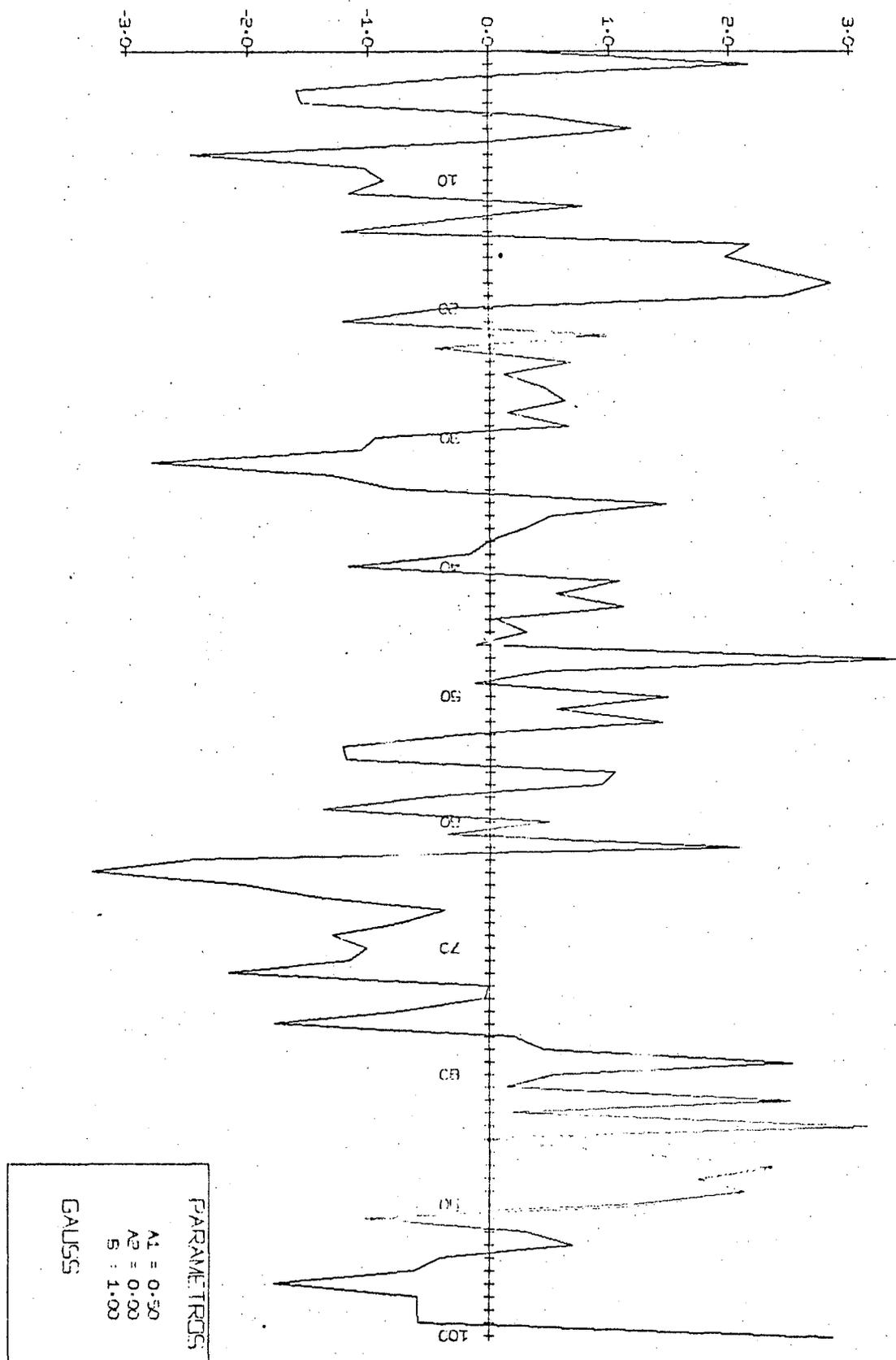


FIG. 37 - Simulação do modelo de observação
(Erro Gaussiano - Experimento 1)

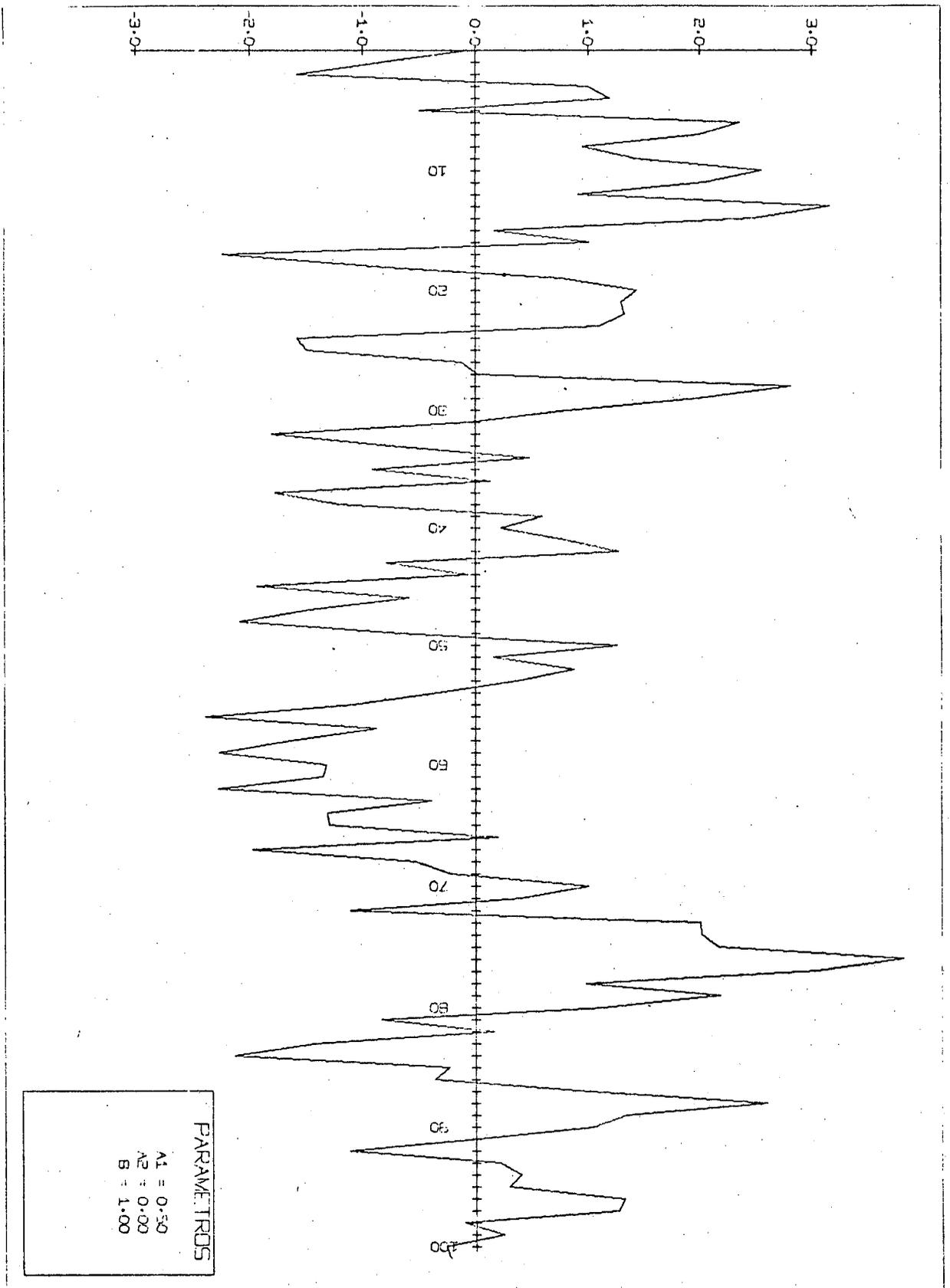


FIG. 38 - Simulação do modelo de observação
 (Valor exato - Experimento 2)

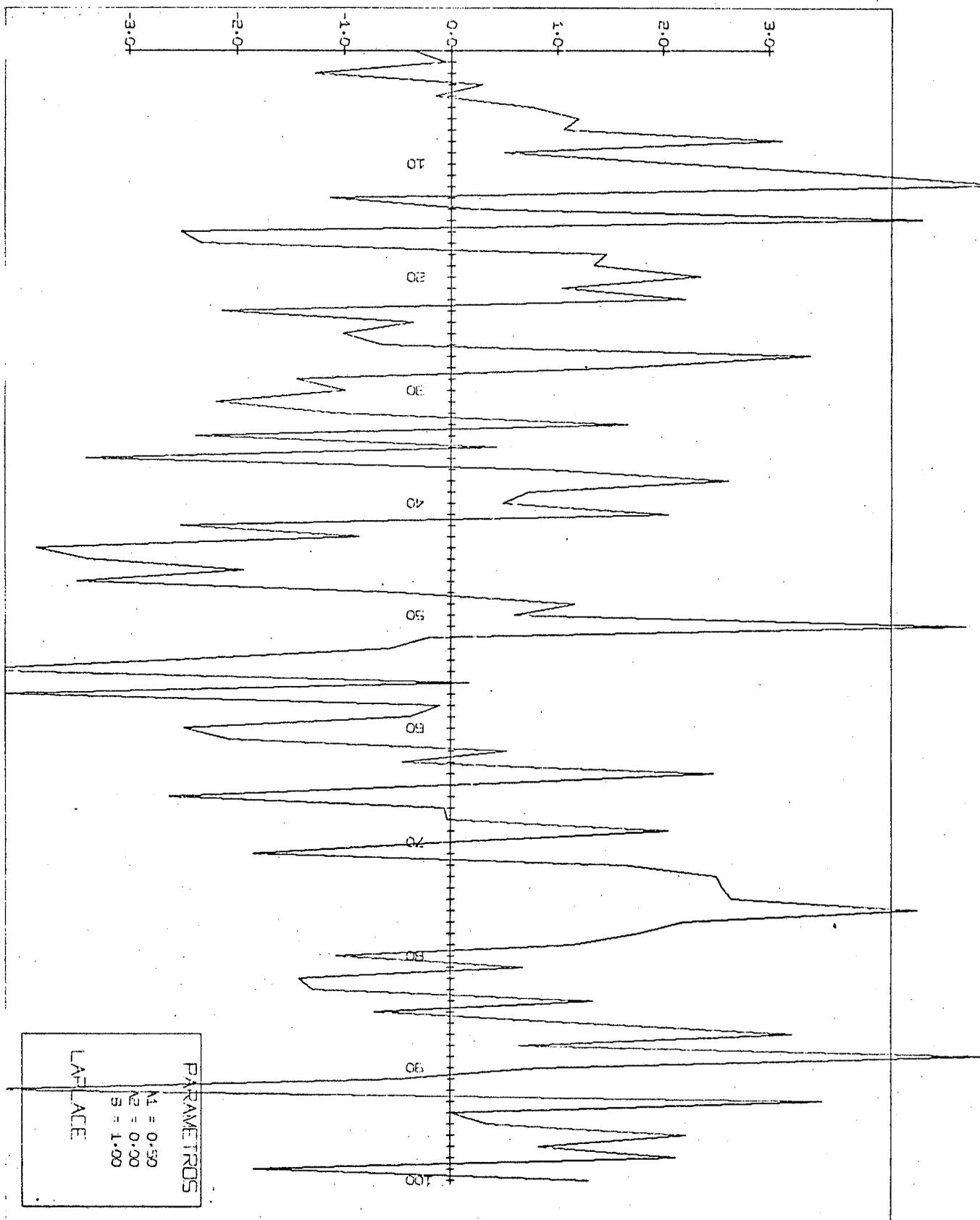


FIG. 39 - Simulação do modelo de observação.
 (Erro Laplaciano - Experimento 2)

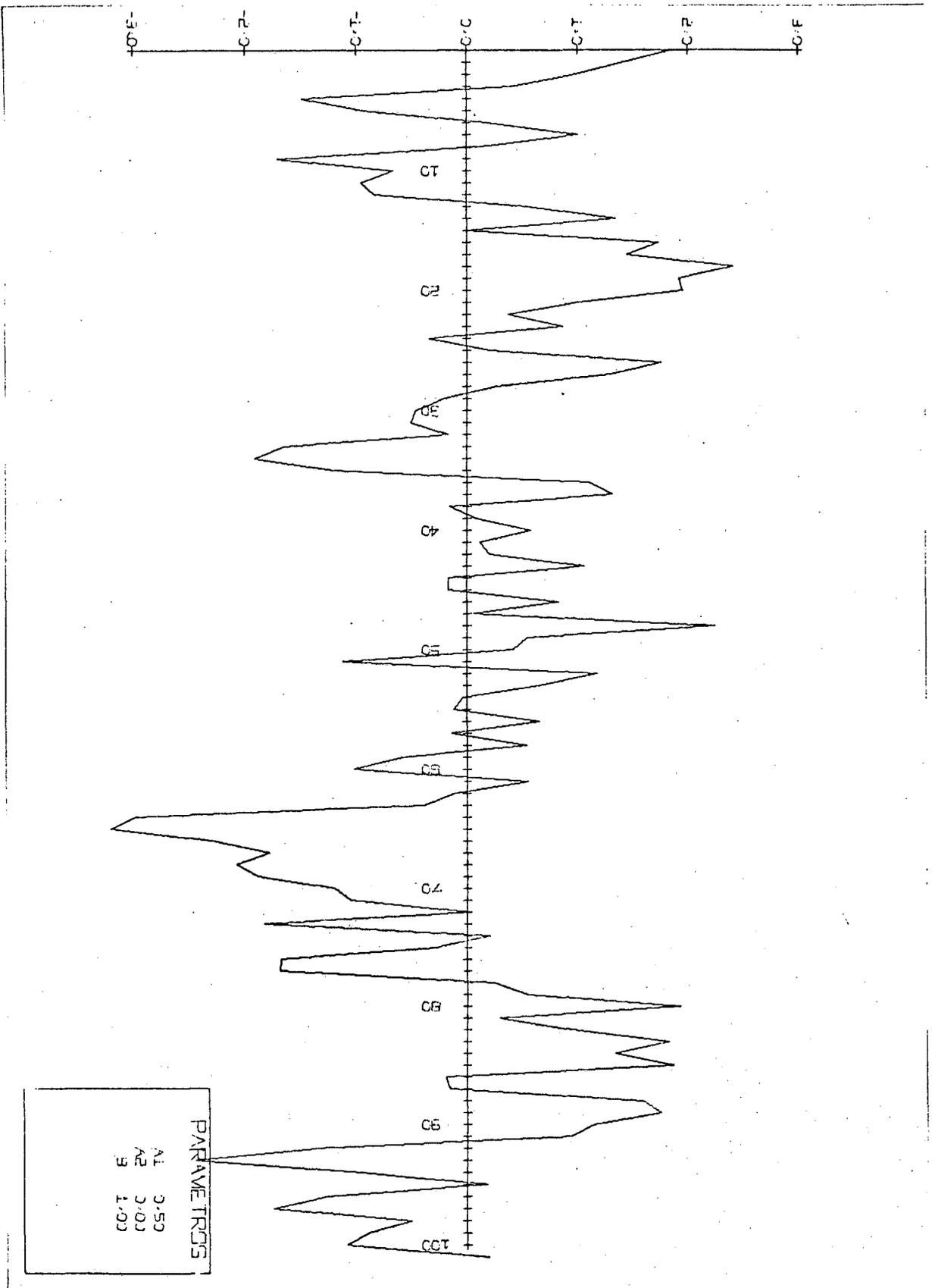


FIG. 40 - Simulação do modelo de observação
(Valor exato - Experimento 3)

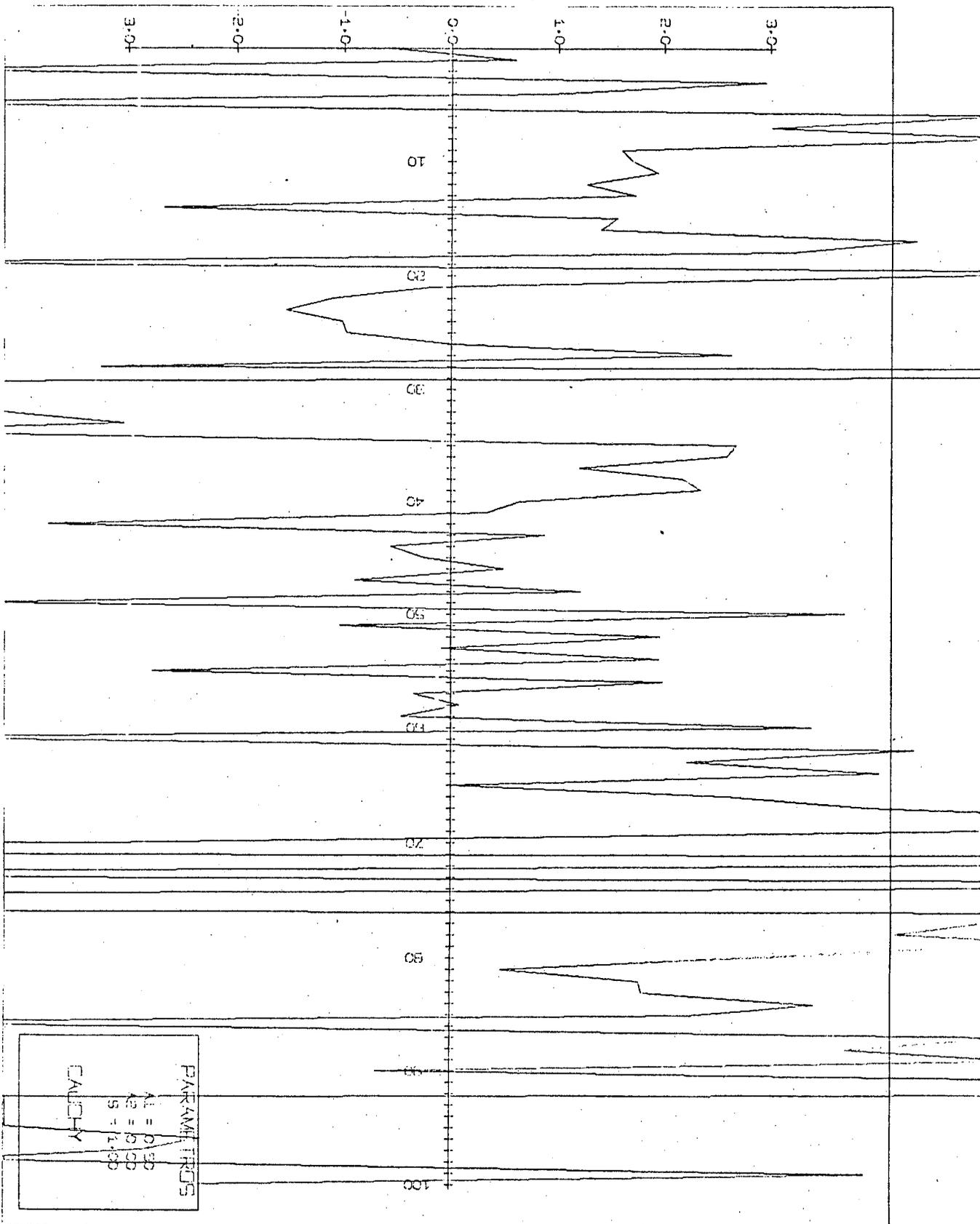


FIG. 41 - Simulação do modelo de observação
 (Erro com distribuição de
 Cauchy - Experimento 3)

C A P Í T U L O I V

4 - UM ESTUDO SOBRE IDENTIFICAÇÃO DE MODELOS

4.1 - INTRODUÇÃO:

Nesse capítulo consideraremos a determinação dos parâmetros do modelo de um sistema linear conhecendo-se um conjunto de dados. Consideraremos três casos:

- I - A forma funcional do modelo é conhecida.
- II - A forma funcional do modelo é desconhecida, mas sabe-se que é linear.
- III - O modelo é desconhecido totalmente, embora saiba-se ser um dentre um conjunto finito.

Na estimação dos parâmetros usamos o critério de avaliação de máxima probabilidade (maximum likelihood estimator).

No caso I a forma funcional do modelo é conhecida, e, dispondo-se de um conjunto de dados, usamos o estimador de máxima probabilidade para determinar os parâmetros desconhecidos.

No caso II, já que o modelo é completamente desconhecido, mas linear, assumimos a forma geral dos modelos lineares como modelo e determinam-se os parâmetros.

No último caso, usamos testes de hipóteses para determinar qual o modelo mais provável dentre um conjunto de modelos possíveis.

4.2 - O PROBLEMA GERAL DE ESTIMAÇÃO

O problema geral consiste na estimação de parâmetros desconhecidos do modelo de um sistema de modo que se otimize um critério de avaliação escolhido.

Seja θ , $\theta \in R^m$, um vetor representativo dos parâmetros desconhecidos e X , $X \in R^n$, um vetor representativo dos valores dados. A transformação do espaço dos θ para o espaço dos X pode ser probabilisticamente descrita por sua função densidade de probabilidade $f(X, \theta)$, ou por uma representação funcional do tipo:

$$g_i(X_i, \theta) = \epsilon_i \quad (36)$$

onde ϵ_i são variáveis aleatórias independentes e identicamente distribuídas, representativas do comportamento estocástico do modelo.

O vetor θ poderá ser determinístico ou aleatório, mas em nosso trabalho trataremos os parâmetros sempre como quantidades determinísticas.

4.2.1 - Tipos de Estimadores

Os estimadores podem ser divididos em estimadores de parâmetros determinísticos ou aleatórios.

a) Estimadores de parâmetros determinísticos:

1. Método dos mínimos quadrados: é uma aproximação não probabilística do problema e consiste em:

$$\min ||g_i(X, \theta)||^2$$

2. Estimador de mínima variância não tendencioso: é um estimador probabilístico, porém prescinde do conhecimento da função densidade de probabilidade. O objetivo é:

$$\min E [||\theta - \hat{\theta}||]^2$$

3. Estimador de máxima probabilidade: (maximum likelihood) é um estimador probabilístico que utiliza o conhecimento completo da função densidade de probabilidade. O critério de performance pode ser reduzido a:

$$\max_{\theta} L(X_i, \theta), \quad (i = 1, 2, \dots, n)$$

ou ainda:

$$\frac{\partial}{\partial \theta} \log (L (X, \theta)) = 0$$

onde L é a função de probabilidade de X (Likelihood function of X)

b) Estimadores de parâmetros aleatórios:

1. Estimador de Bayes: esse estimador usa o conhecimento - de probabilidades a priori sobre o comportamento do sistema.
2. Estimador de máxima probabilidade: é usado quando a função densidade de probabilidade a priori é desconhecida.
3. Critério Minimax: como o próprio nome diz não é exatamente um estimador e sim um critério para estabelecimento de políticas de procedimento ótimo, baseando-se no conhecimento das distribuições de probabilidade do sistema.

4.3 - OBTENÇÃO DOS PARÂMETROS DO MODELO A PARTIR DOS DADOS:

Vamos apresentar os diferentes estimadores e critérios de avaliação de modo que, estudando-se as características esperadas dos estimadores, possamos efetuar uma seleção adequada do método de estimação que usaremos.

4.3.1 - Estimadores e critérios de avaliação

Seja um problema da forma:

$$g_i (X_i, \theta) = \varepsilon_i \quad , (i = 1, 2, \dots, n)$$

onde $g_i(\theta)$ é conhecida a menos de seus parâmetros θ . Portanto, o problema consiste em determinar estimadores $\hat{\theta}$ de θ e de $g_i(\theta)$, de modo a otimizar o critério de avaliação escolhido.

As características esperadas de um bom estimador são:

1. Não tendenciosidade
2. Consistência
3. Eficiência
4. Ser uma função de estatísticas suficientes

Em vista disso, definiremos estas características e ire-

mos analisá-las:

1) Não tendenciosidade: um estimador $\hat{\theta}$ é dito não tendencioso se $E[\hat{\theta}] = \theta$, pois sendo θ uma função das observações Y , $\hat{\theta}$ é uma variável aleatória. Juntamente com a não tendenciosidade é preciso que a variância de $\hat{\theta}$ seja pequena, pois caso contrário um estimador tendencioso poderá fornecer melhores resultados, como se pode observar na figura 42.

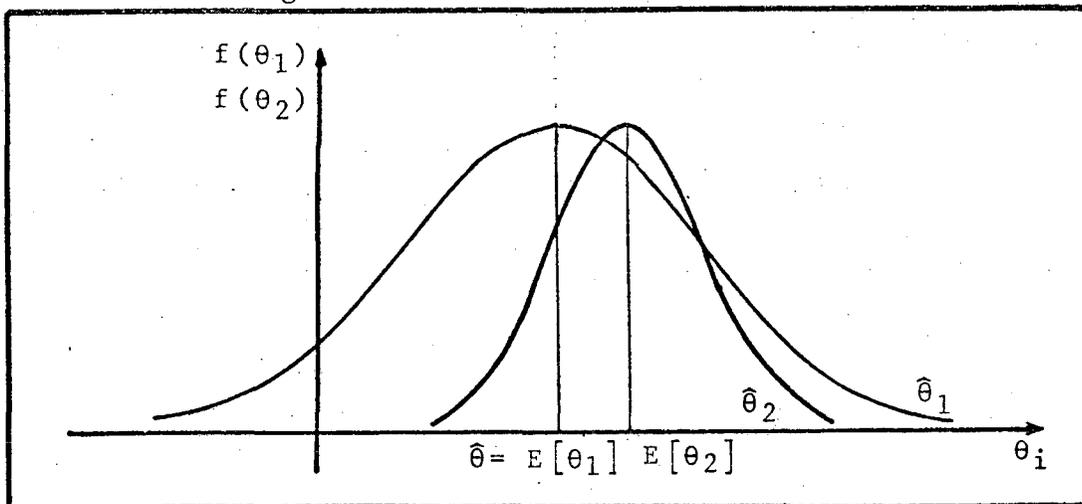


FIG. 42 - Tendenciosidade dos estimadores.

2) Estimador consistente: seja $\hat{\theta}(x_1, x_2, \dots, x_n)$ um estimador de θ , onde x_1, x_2, \dots, x_n é uma sequência de dados. Então $\hat{\theta}$ é dito um estimador consistente se:

$$\lim_{n \rightarrow \infty} \hat{\theta} \xrightarrow{p} \theta$$

ou seja, $\lim_{n \rightarrow \infty} \hat{\theta}$ converge probabilisticamente para θ quando o número de observações cresce.

3) Estimador eficiente: um estimador $\hat{\theta}$ é dito estimador eficiente de θ se ele fornece o menor erro quadrático possível, isto é, $\hat{\theta}$ é um estimador eficiente se minimiza:

$$E[\theta - \hat{\theta}]^2$$

4) Estimador suficiente: $\hat{\theta}$ é um estimador suficiente de θ se é uma função de estatísticas suficientes, ou ainda $\hat{\theta}(x_1, x_2, \dots)$ é um estimador suficiente de θ se e somente se:

$$f(x_1, x_2, \dots, x_n, \theta) = g(\hat{\theta}, \theta) \cdot h(x_1, x_2, \dots, x_n)$$

onde $g(\hat{\theta}, \theta)$ é a densidade ou função de probabilidade de $\hat{\theta}$ e

$h(x_1, \dots, x_n)$ não depende de θ .

Definidas as características esperadas de um estimador, é preciso analisar os diferentes esquemas de estimação, tendo em vista tais características, e com isso escolher o método mais adequado.

4.3.2 - Propriedades do estimador de máxima probabilidade ¹¹

1) Se um estimador eficiente para θ existir, ele é dado pelo estimador de máxima probabilidade*.

2) Se $\hat{\theta}$ é o estimador de máxima probabilidade de θ , então $V(\hat{\theta})$ é o estimador de máxima probabilidade de $V(\theta)$, onde $V(\theta)$ é qualquer função que admite inversa.

3) O estimador de máxima probabilidade sob condições gerais de regularidade é um estimador consistente.

4) O estimador de máxima probabilidade não precisa ser não tendencioso.

5) É uma função de estatísticas suficientes.

6) O estimador de máxima probabilidade é assintoticamente normal e assintoticamente eficiente.

O método de máxima probabilidade, por suas características acima relacionadas, foi o escolhido para a estimação dos parâmetros neste trabalho. Por outro lado, o estimador de máxima probabilidade é largamente utilizado em estimação, quando a função densidade de probabilidade é conhecida ².

Escolhemos a função de probabilidade como sendo:

$$L = \prod_{i=1}^n f(\epsilon_i)$$

então:

$$L = \prod_{i=1}^n f(g_i(X, \theta))$$

onde $f(\epsilon_i)$ é a função densidade de probabilidade de ϵ_i .

* - Se um estimador eficiente não existir, então poderão existir estimadores de menor variância; no entanto a dificuldade reside no fato de não haver maneira de encontrá-los.

Tem-se agora definidos o método de estimação e o respectivo critério de avaliação. Resta-nos agora analisar a situação usual de estimação.

Desde que a premissa de linearidade do modelo permaneça, podemos dividir os problemas usuais de estimação em duas classes distintas: o conjunto dos problemas em que a forma funcional do modelo é conhecida e a classe dos problemas em que se desconhece a forma funcional do modelo.

Procurando uma forma simples e coerente de apresentação para tais problemas, abordaremos inicialmente os problemas em que a forma funcional é conhecida.

4.3.3 - Determinação dos parâmetros de um modelo cuja forma funcional é conhecida.

Esse é o caso mais comum de problemas de estimação linear, caracterizado pela disponibilidade de uma série de dados representativos do passado de um sistema linear e pelo conhecimento, arbitrário ou não, da forma funcional do modelo.

Suponhamos que se tenha uma sequência de dados $\{X_i\}$, ($i = 1, 2, \dots, N$), e o modelo representativo do sistema* seja

$$X(n) = A.X(n-1) + B.W(n)$$

onde:

$$E [W(n)] = 0 \quad (38)$$

$$E [W(n) \cdot W(n)'] = \Lambda_n \quad (39)$$

são conhecidos. Temos que determinar A e B. Suponha-se que W(n) tenha distribuição normal. Então pode-se afirmar que $(X(n) - A.X(n-1))$ é uma sequência de variáveis aleatórias independentes, normalmente distribuídas com média nula e matriz de covariância

* - O modelo de primeira ordem $X(n) = A.X(n-1) + B.W(n)$ pode ser considerado sem perda de generalidade porque pode ser estendido a um modelo de ordem p qualquer; evitamos, assim procedendo, a problemas de equações simultâneas.

riância dada por:

$$\begin{aligned} E\{[B \cdot W(n)] [(B \cdot W(n))']\} &= E\{[B] [W(n)] [(W(n))'] [B']\} = \\ &= [B] E\{[W(n)] [(W(n))']\} [B'] = B \Lambda B' \end{aligned}$$

Assim sendo, a função densidade de probabilidade conjunta de $W(n)$, $\forall n$ será:

$$\begin{aligned} f(W(1), W(2), \dots, W(n)) &= \frac{1}{(2\pi)^{n/2} (\det. \Lambda)^{1/2}} \cdot \\ &\cdot \text{Exp.}\left\{-\frac{1}{2} \sum_{n=1}^N (W(n))' \Lambda^{-1} W(n)\right\} \end{aligned} \quad (40)$$

Usando o estimador de máxima probabilidade determinamos A e B tais que maximizem a função densidade de probabilidade, isto é, objetiva-se maximizar (40) com respeito a A e B .

Tomando-se $L = \log f(X(1), \dots, X(n))$, podemos afirmar que

$$\begin{aligned} \max_{A, B} f &= \max_{A, B} \log(f) = \min_{A, B} \frac{1}{2} \sum_{n=1}^N (X(n) - AX(n-1))' (BAB')^{-1} \cdot \\ &\cdot (X(n) - AX(n-1)) \end{aligned}$$

ou seja:

$$\frac{\partial L}{\partial A} = \frac{\partial}{\partial A} \left[\sum_{n=1}^N (X(n) - AX(n-1))' (BAB')^{-1} (X(n) - AX(n-1)) \right] = 0 \quad (41)$$

e

$$\frac{\partial L}{\partial B} = \frac{\partial}{\partial B} \left[\sum_{n=1}^N (X(n) - AX(n-1))' (BAB')^{-1} (X(n) - AX(n-1)) \right] = 0 \quad (42)$$

De (41) obtêm-se, após manipulações algébricas:

$$\sum_{n=1}^N (X(n-1))' (\hat{B} \hat{\Lambda} \hat{B}')^{-1} \cdot X(n) = \sum_{n=1}^N (X(n-1))' (\hat{B} \hat{\Lambda} \hat{B}')^{-1} \cdot \hat{A} X(n-1) \quad (43)$$

e de (42) obtêm-se:

$$\begin{aligned} \sum_{n=1}^N (X(n) - AX(n-1))' (BAB')^{-1} \cdot B \Lambda (BAB')^{-1} \cdot (X(n) - AX(n-1)) + \\ + \sum_{n=1}^N (X(n) - AX(n-1))' (BAB')^{-1} \cdot \Lambda B' (BAB')^{-1} (X(n) - AX(n-1)) \end{aligned} \quad (44)$$

donde então é fácil obter-se \hat{A} e \hat{B} .

Exemplos de aplicação:

1) Seja um sistema cujo modelo escalar seja:

$$X(n+1) = A.X(n) + B.W(n) \quad , \quad (n = 1, 2, \dots, N)$$

sendo conhecido um conjunto de dados $\{X(n)\}$. Admita-se que A e B sejam escalares e que, como o parâmetro B atua somente sobre a covariância da variável $W(n)$, assumimos $B = 1$. Então, podemos escrever:

$$f(W(n)) = \frac{1}{(2\pi)^{1/2} \cdot \sigma^2} \text{Exp.} \left\{ \frac{-1}{2\sigma^2} \cdot [W(n) - \mu_w]^2 \right\} \quad , \quad (n=1, 2, \dots, N)$$

desde que $W(n)$ seja normal e independentemente distribuída com média e variância dadas por μ_w e σ^2 .

Vamos adotar $\mu_w = 0$ e $\sigma_w^2 = 1$. Assim, tem-se:

$$f(X(1), \dots, X(N)) = \left(\frac{1}{2\pi} \right)^{N/2} \cdot \text{Exp.} \left\{ -\frac{1}{2} \sum_{n=1}^N [X(n) - A.X(n-1)]^2 \right\}$$

donde, aplicando o estimador de máxima probabilidade, ter-se-á:

$$\frac{\partial \log}{\partial A} \left[f(X(1), \dots, X(N)) \right] = - \sum_{n=1}^N [X(n) - A.X(n-1) (-X(n-1))] = 0$$

e finalmente:

$$\hat{A}_1 = \frac{\sum_{n=1}^N [X(n) \cdot X(n-1)]}{\sum_{n=1}^N [X^2(n-1)]}$$

2) Analizemos agora o modelo de segunda ordem em sua forma escalar com $B = 1$. Inicialmente têm-se:

$$W(n) = X(n+1) - A_1 X(n) - A_2 X(n-1) \quad , \quad (n = 1, 2, \dots, N)$$

Assumindo-se que $W(n)$ é normal e independente com média zero e variância unitária, escreve-se:

$$f(X(1), \dots, X(N)) = \frac{1}{(2\pi)^{N/2}} \cdot \text{Exp.} \left\{ \left[\left(-\frac{1}{2} \sum_{n=1}^N [X(n+1) - A_1 X(n) - A_2 X(n-1)]^2 \right) \right] \right\}$$

Maximizando-se em relação a A_1 e A_2 ,

$$\frac{\partial \log f}{\partial A_1} = \frac{-1}{2\pi} \sum_{n=1}^N [X(n+1) - A_1 X(n) - A_2 X(n-1)] [-X(n)] = 0 \quad (45)$$

$$\frac{\partial \log f}{\partial A_2} = \frac{-1}{2\pi} \sum_{n=1}^N [X(n+1) - A_1 X(n) - A_2 X(n-1)] [-X(n-1)] = 0 \quad (46)$$

Eliminando-se A_1 e A_2 de (45) e (46), resulta:

$$\hat{A}_1 = \frac{\sum_{n=1}^N X^2(n) \sum_{n=1}^N X(n+1) \cdot X(n-1) - \sum_{n=1}^N X(n) \cdot X(n-1) \cdot \sum_{n=1}^N X(n+1) X(n)}{\sum_{n=1}^N X^2(n) \cdot \sum_{n=1}^N X^2(n-1) - \left[\sum_{n=1}^N X(n-1) \cdot X(n) \right]^2}$$

$$\hat{A}_2 = \frac{\sum_{n=1}^N X(n-1) \cdot X(n) \sum_{n=1}^N X^2(n-1) - \sum_{n=1}^N X(n+1) \cdot X(n-1) \sum_{n=1}^N X(n-1) \cdot X(n)}{\sum_{n=1}^N X^2(n) \cdot \sum_{n=1}^N X^2(n-1) - \left[\sum_{n=1}^N X(n-1) \cdot X(n) \right]^2}$$

Como podemos observar, é enorme a importância da escolha da ordem do modelo, pois a complexidade e o custo da estimação cresce com o aumento da ordem do modelo. Em alguns casos, a melhoria da estimação com o aumento da ordem do modelo não compensa o maior trabalho se o número de dados não for grande.

n	X(n)	X(n-1)	
1	0,366	-1,619	$\sum_{n=1}^{10} X^2(n-1) = 4,807$
2	-0,406	0,366	
3	-0,299	-0,406	
4	-0,762	-0,299	$\sum_{n=1}^{10} X(n-1) \cdot X(n) = 0,546$
5	-0,166	-0,762	
6	-0,112	-0,166	Valor Estimado
7	-0,038	0,112	$\hat{A}_1 = 0,114$
8	-0,772	-0,038	
9	-0,759	-0,772	Valor Real
10	-0,283	-0,759	$A_1 = 0.500$

QUADRO 2 - Estimação do Parâmetro de Modelo de Primeira Ordem.

n	X(n+1)	X(n)	X(n-1)	
1	0,941	1,481	-1,432	$\sum_{n=1}^{10} X^2(n) = 12,821$
2	0,643	0,941	1,481	
3	-1,432	0,643	0,941	$\sum_{n=1}^{10} X^2(n-1) = 13,145$
4	-0,271	-1,432	0,643	
5	-0,990	-0,271	-1,432	$\sum_{n=1}^{10} X(n+1) \cdot X(n-1) = 4,283$
6	-0,052	-0,990	-0,271	
7	-1,946	-0,052	-0,990	$\sum_{n=1}^{10} X(n+1) \cdot X(n) = 4,965$
8	-0,839	-1,946	-0,052	
9	-1,314	-0,839	-1,946	$\sum_{n=1}^{10} X(n) \cdot X(n-1) = 2,504$
10	-0,287	-1,314	-0,839	

Valores Estimados

$$\hat{A}_1 = 0,298$$

$$\hat{A}_2 = 0,156$$

Valores Reais

$$A_1 = 0,500$$

$$A_2 = 0,250$$

QUADRO 3 - Estimaco dos Parmetros do Modelo de Segunda Ordem.

Nos quadros 2 e 3 mostramos resultados que obtemos na estimaco dos parmetros. O quadro 2 se refere a um modelo de primeira ordem e o quadro 3 a um modelo de segunda ordem. Estimamos os parmetros com dados obtidos da simulaco dos modelos.

4.3.4 - O Problema de Estimaco dos Parmetros de um Modelo cuja Forma Funcional  Desconhecida.

A teoria da informaco nos assegura que a qualidade de uma estimaco  proporcional  quantidade de informaco (nmero de dados) utilizada em sua obtenço. Ora, o caso em que o modelo  desconhecido poderia, a primeira vista, significar uma perda de

qualidade na estimativa; no entanto, o problema é, teoricamente, facilmente solucionado sem perda de qualidade na estimativa.

É fácil compreender que utilizando-se o modelo geral dos sistemas lineares não estaremos incorrendo em erro por omissão de informação, desde que o sistema seja linear, pois qualquer que seja o modelo real do sistema este será um caso particular do modelo geral.

Consequentemente, um sistema linear admite como modelo:

$$X(n+1) = A_1X(n) + A_2X(n-1) + \dots + A_pX(n-p+1) + \\ + B_1W(n) + \dots + B_pW(n-p+1) \quad , \quad p \leq N \quad (47)$$

onde N é o número de dados disponíveis.

Nesse caso a simulação poderá ser utilizada para determinação do melhor valor de p .

Evidentemente que com a equação (47) o modelo fica determinado e cai-se no caso anterior de estimação dos parâmetros. Por outro lado, o emprego da simulação para determinação da forma funcional do modelo poderá ser um trabalho demorado e de custo apreciável. Em vista do exposto, discute-se a seguir um modo mais simples para escolher o modelo mais adequado a cada situação, usando a teoria dos testes de hipóteses.

4.4 - ESCOLHA DO MODELO MAIS ADEQUADO.

Devido à dificuldade de determinação do modelo no caso anterior, consideraremos a seguir o caso em que é sabido que o modelo pertence a um conjunto finito de modelos ¹⁵.

Vamos abordar o problema de decidir dentre dois modelos, denominados modelo 0 e modelo I, qual o mais adequado a um conjunto de dados preliminarmente obtidos. Para isso vamos supor que possamos dicotomizar o espaço amostral Ω em espaços mutuamente exclusivos Ω_0 e Ω_1 , respectivamente, representativos dos espaços de resultados $\{X\}$ dos modelos 0 e I. A cada conjunto de observações $X = \{x_1, x_2, \dots, x_n\}$ pode-se associar um ponto no espaço amostral. Assim podemos formular duas hipóteses:

- 1) - Hipótese H_0 - Modelo 0 é verdadeiro.

2) - Hipótese H_1 - Modelo I é verdadeiro.

Como critério de escolha, adotaremos H_0 se $\{X\} \in \Omega_0$ e H_1 se $\{X\} \in \Omega_1$, como esquematizado na figura 43.

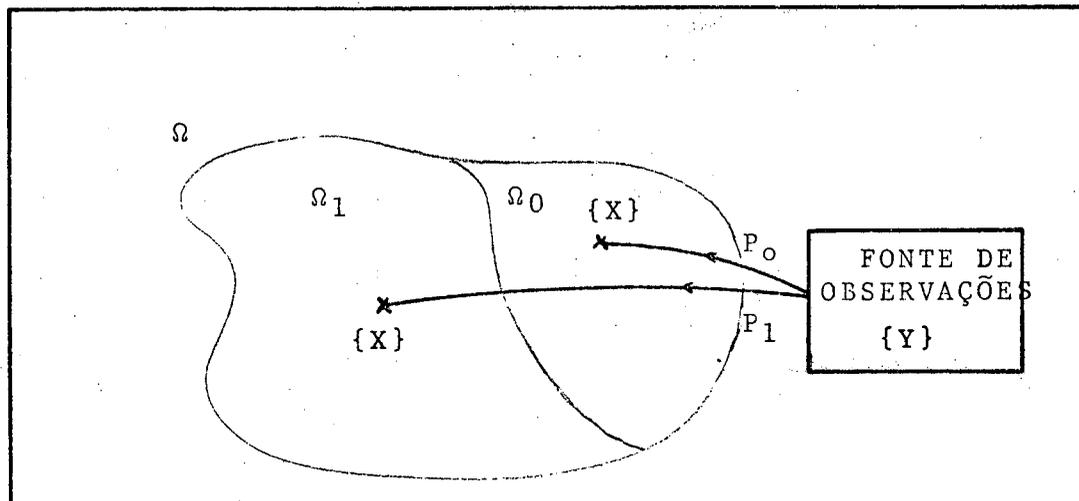


FIG. 43 - Critério de escolha de modelos.

Nosso objetivo reside na determinação de qual hipótese, H_0 ou H_1 , é verdadeira. Usando o critério de decisão esboçado acima, ter-se-á, em média, a decisão acertada.

Nessa situação poderão ocorrer os seguintes casos:

- a) H_0 é a hipótese verdadeira e decide-se por H_0
- b) H_0 é a hipótese correta e adotamos H_1
- c) H_1 é a hipótese verdadeira e adotamos H_1
- d) H_1 é a hipótese verdadeira e adotamos H_0

A cada uma destas situações associa-se um peso ou custo, C_{ij} , de modo que, minimizando o custo médio total, possamos obter uma regra de decisão.

Sejam C_{10} , C_{01} , C_{00} e C_{11} esses pesos, em que o primeiro índice indica a hipótese escolhida e o segundo índice indica a hipótese que ocorreu. É preciso assumir, ainda, que a ocorrência de X em Ω_1 , ou Ω_0 seja probabilística, com probabilidades p_1 e p_0 , respectivamente. Tais probabilidades são chamadas probabilidades a priori de ocorrência. Estas probabilidades representam a informação do observador a respeito da fonte ou do sistema antes de que o experimento seja efetuado.

Agora pode-se calcular o valor esperado do custo de toma-

da de decisão em cada situação. Assim tem-se:

Custo de decidir por $H_0 | H_0$ é verdade = $C_{00}p_0P$ (decidir $H_0 | H_0$ é certo)

Custo de decidir por $H_0 | H_1$ é verdade = $C_{01}p_1P$ (decidir $H_0 | H_1$ é certo)

Custo de decidir por $H_1 | H_1$ é verdade = $C_{11}p_1P$ (decidir $H_1 | H_1$ é certo)

Custo de decidir por $H_1 | H_0$ é verdade = $C_{10}p_0P$ (decidir $H_1 | H_0$ é certo)

Denotaremos o valor esperado do custo total como R , o risco esperado, logo:

$$R = C_{00}p_0P (\text{dec. } H_0 | H_0 \text{ é vd.}) + C_{01}p_1P (\text{dec. } H_0 | H_1 \text{ é vd.}) + \\ + C_{11}p_1P (\text{dec. } H_1 | H_1 \text{ é vd.}) + C_{10}p_0P (\text{dec. } H_1 | H_0 \text{ é vd.}) \quad (48)$$

Reescrevendo (48) em t\u00e9rmos das probabilidades de ocorr\u00eancia de $\{X\}$ em cada espa\u00e7o Ω_0 ou Ω_1 , obtem-se:

$$R = C_{00}p_0 \int_{\Omega_0} f(X|H_0) dx + C_{10}p_0 \int_{\Omega_1} f(X|H_0) dx + \\ + C_{01}p_1 \int_{\Omega_0} f(X|H_1) dx + C_{11}p_1 \int_{\Omega_1} f(X|H_1) dx . \quad (49)$$

Para que os pesos C_{ij} possam ser encarados como custos e R assumo o significado de risco esperado, devemos atribuir \u00e0s decis\u00f5es erradas um custo maior que \u00e0s decis\u00f5es corretas. Assim:

$$C_{10} > C_{00} \\ C_{01} > C_{11} \quad (50)$$

Levando em conta que $\Omega = \Omega_1 \cup \Omega_0$, e que

$$\int_{\Omega} f(X|H_0) dx = \int_{\Omega} f(X|H_1) dx = 1 \quad (51)$$

pode-se reescrever (49) assim:

$$R = p_0C_{10} + p_1C_{11} + \int_{\Omega_0} \{ [p_1(C_{01}-C_{11})] \cdot f(X|H_1) - \\ - [p_0(C_{10}-C_{00})] \cdot f(X|H_0) \} dx . \quad (52)$$

A equa\u00e7\u00e3o (52) nos permite obter uma interessante an\u00e1lise sobre o problema: o primeiro termo de (52) representa o custo fixo, enquanto a integral representa o custo control\u00e1vel pela ocorr\u00eancia de $\{X\}$ em Ω_0 . De acordo com (50), os t\u00e9rmos entre pa-

Admita-se o desconhecimento total sobre a probabilidade a priori da ocorrência de H_0 ou H_1 . Assumimos então $p_0 = p_1$; quanto aos custos, atribuímos:

$$C_{00} = C_{11}$$

$$C_{10} = C_{01}$$

de modo que

$$\gamma(X) \underset{H_0}{\overset{H_1}{>}} 1 \quad (56)$$

Supondo que $W(n)$ seja um ruído branco gaussiano, pode-se escrever:

$$f(W(n)) = \frac{1}{(2\pi)^{N/2}} \cdot \text{Exp} \left[-\frac{1}{2} \cdot W(n)' \Lambda^{-1} \cdot W(n) \right] \quad , (n=1, \dots, N)$$

logo:

$$f(X(n) - A_0 X(n-1)) = \frac{1}{(2\pi)^{N/2} \det(\Lambda)} \cdot \text{Exp} \left[-\frac{1}{2} (X(n) - A_0 X(n-1))' \Lambda^{-1} (X(n) - A_0 X(n-1)) \right]$$

e

$$f(X(n) - A_1 X(n-1) - A_2 X(n-2)) = \frac{1}{(2\pi)^{N/2} \det(\Lambda)} \cdot \text{Exp} \left[-\frac{1}{2} (X(n) - A_1 X(n-1) - A_2 X(n-2))' \Lambda^{-1} (X(n) - A_1 X(n-1) - A_2 X(n-2)) \right]$$

Como a função $\ln(x)$ é uma função monotônica, pode-se escrever que se

$$\gamma(X) \geq 1 \quad , \quad \ln \gamma(X) \geq 0 \quad ,$$

portanto, pode-se modificar (56) para:

$$\ln \gamma(X) \underset{H_0}{\overset{H_1}{>}} 0 \quad (57)$$

Portanto, aplicando-se (57) pode-se obter facilmente:

$$\begin{aligned} \ln \gamma = & - \left[(X(n) - A_0 X(n-1))' \cdot \Lambda^{-1} \cdot (X(n) - A_0 X(n-1)) \right] + \\ & + \left[(X(n) - A_1 X(n-1) - A_2 X(n-2))' \cdot \Lambda^{-1} (X(n) - A_1 X(n-1) - A_2 X(n-2)) \right] . \end{aligned}$$

Então, teremos como critério final de escolha:

$$(X(n) - A_1 X(n-1) - A_2 X(n-2))' \cdot \Lambda^{-1} \cdot (X(n) - A_1 X(n-1) - A_2 X(n-2)) \stackrel{H_1}{>} \stackrel{H_0}{}$$

$$(X(n) - A_0 X(n-1))' \cdot \Lambda^{-1} \cdot (X(n) - A_0 X(n-1))$$

O quadro 4 mostra a utilização do critério de decisão.

Sejam $X(n)$ e $X(n-1)$ vetores com 10 componentes cada um, gerados pelo modelo:

$$X(n+1) = 0,500 X(n) + W(n)$$

Simulemos a tomada de decisão entre os modelos

$$X(n+1) = 0,5 X(n) + W(n) \rightarrow \text{Hipótese } H_0$$

$$X(n+1) = 1,0 X(n) + W(n) \rightarrow \text{Hipótese } H_1$$

Então o critério de decisão será:

$$\sum_{n=1}^{10} (X(n) - 0,50 X(n-1))^2 \stackrel{H_1}{>} \stackrel{H_0}{<} \sum_{n=1}^{10} (X(n) - 1,0 X(n-1))^2$$

n	X(n-1)	X(n)	
1	-1,619	0,366	$\sum_{n=1}^{10} (X(n) - 0,50 X(n-1))^2 = 2,925 \rightarrow H_0$
2	0,366	-0,406	
3	-0,406	-0,299	
4	-0,299	-0,762	$\sum_{n=1}^{10} (X(n) - X(n-1))^2 = 5,945 \rightarrow H_1$
5	-0,762	-0,166	
6	-0,166	0,112	
7	0,112	-0,038	
8	-0,038	-0,772	
9	-0,772	-0,759	
10	-0,759	-0,283	

QUADRO 4 - Uso do critério de decisão.

Escolhemos então H_0 como a hipótese correta, pois

$$\sum_{n=1}^{10} (X(n) - X(n-1))^2 > \sum_{n=1}^{10} (X(n) - 0,50 X(n-1))^2$$

o que confirma a validade do critério de decisão.

4.5 - OBSERVAÇÕES SOBRE ESTE CAPÍTULO

Da mesma forma que no trabalho de Zellner e Geisel ¹⁸, em todo nosso trabalho adotamos, embora não seja o caso geral, a suposição de serem os parâmetros dos modelos quantidades determinísticas. Isso porque o desenvolvimento da estimação de parâmetros estocásticos diverge do nosso objetivo.

Por outro lado, neste capítulo mostramos uma forma unificada e geral para obtenção dos parâmetros, através da representação matricial, evitando desta maneira, as conseqüentes variações de técnicas de estimação como no trabalho supra citado.

Observado-se as técnicas de obtenção de um estimador, podemos agora tecer alguns comentários sobre os modelos de observação: Em primeiro lugar deve-se tomar cuidado na determinação da função densidade de probabilidade conjunta da variável estocástica, pois caso não seja gaussiana, a função não será a mesma que determinamos. Por isso, a suposição, frequentemente adotada, de que as variáveis aleatórias tem distribuição gaussiana nos processos físicos ou econômicos, pode ser a causa de grandes erros.

Deve-se levar em conta que os valores dos parâmetros obtidos por estimação podem diferir dos valores por ventura esperados.

Outro fator a considerar, fruto da estocasticidade do processo, é a propriedade de um modelo estocástico não seguir exatamente, a cada momento, o comportamento do sistema. No entanto, o modelo estocástico representará o comportamento médio do sistema, estando o valor fornecido pelo modelo na vizinhança de seu valor exato. Esta característica pode não ser a esperada em muitos processos, mas nos parece que na área da Econometria este comportamento já está assimilado e raramente ocasiona perdas.

C A P Í T U L O V

5 - UMA ABORDAGEM À PREVISÃO

5.1 - INTRODUÇÃO

A estimação estatística é talvez uma das mais ricas e excitantes áreas do conhecimento científico. Aparece quase que na totalidade dos ramos das ciências; nas ciências sociais, no controle de sistemas submetidos a distúrbios aleatórios ou na tomada de decisões sob incerteza.

Existem um grande número de técnicas que são utilizadas na teoria da estimação. A diferença entre os diversos métodos de estimação comumente usados, reside fundamentalmente na escolha do critério de avaliação dos estimadores e das suposições estatísticas assumidas.

Geralmente os métodos de previsão existentes são lineares, no entanto os métodos lineares de previsão podem nem sempre dar bons resultados.

Nesse capítulo proporemos um novo método de previsões. Este método explora o comportamento estatístico dos processos em consideração. Aqui mostra-se que as previsões lineares são boas - sempre que os processos envolvidos são gaussianos. Nos casos em que os processos envolvidos não são gaussianos os estimadores lineares são falhos, pois mostra-se que não é garantido que o valor ótimo (mínimo erro), seja obtido.

Mais uma vez consideraremos os modelos dinâmicos e de observação desenvolvidos no Capítulo II, agora com o objetivo de se obter previsões.

Consideraremos somente esquemas recursivos de previsão, para assim evitarmos dificuldades de computação.

5.2 - O PROBLEMA DE PREVISÃO ESTOCÁSTICA.

Consideremos os modelos dados por (30) e (32), aqui repetidas para maior clareza:

$$\begin{aligned} X(n+1) &= A.X(n) + B.W(n) \\ Y(n) &= X(n) + V(n) \end{aligned} \quad (58)$$

Assumimos que todos os processos envolvidos são de média nula, que $\{W(n)\}$ é um processo de ruído branco gaussiano e que $\{V(n)\}$ é um processo estocástico com distribuição de probabilidade geralmente não-gaussiana.

Nosso objetivo reside em obter previsões para os valores de $X(n+1)$, conhecendo-se um conjunto de dados $\{Y(n)\}$ e uma previsão prévia, $\hat{X}(n)$. Matematicamente, o problema pode ser expresso por:

$$\hat{X}(n+1) = \psi(Y(n), \hat{X}(n))$$

em que $\psi(Y(n), \hat{X}(n))$ é a função de previsão que deveremos determinar. Determinaremos ψ de modo que se tenha em média o mínimo erro, isto é:

$$\min E [(X(n+1) - \hat{X}(n+1))^2 | Y(n), \hat{X}(n)] \quad (59)$$

Para melhor compreensão dividimos o problema em três partes:

- I - Estimação com uma única observação.
- II - Estimação com um conjunto de observações.
- III - Predição.

Em que assumimos os termos estimacão como a obtenção de $X(n)$ conhecendo-se um conjunto de dados $\{Y(n)\}$ e como predição, a determinação de $X(n+m)$ sabendo-se somente n dados, isto é, $\{Y(n)\}$.

5.3 - ESTIMAÇÃO COM UMA ÚNICA OBSERVAÇÃO.

Seja

$$Y = X + V$$

o modelo de observação.

Suponhamos que $\{X\}$ seja um processo gaussiano, V um pro

cesso não-gaussiano e que X e V sejam processos independentes.

Adotemos \hat{X} como estimativa de X ; evidentemente que \hat{X} é uma função de Y , i.é.,

$$\hat{X} = \phi(Y)$$

A estimativa \hat{X} de X , é obtida de:

$$\min_{\hat{X}} E [(X - \hat{X})^2]$$

Como conhecemos Y , podemos reescrever a relação acima como:

$$\min_{\hat{X}} E [(X - \hat{X})^2 | Y] \quad (60)$$

Por outro lado, sabemos que³:

$$\hat{X} = E [X | Y] \quad (61)$$

ou da mesma forma:

$$\begin{aligned} \hat{X} &= \int_{-\infty}^{\infty} X \cdot f(X|Y) dx = \int_{-\infty}^{\infty} X \frac{f(X, Y)}{f_Y(Y)} dx = \\ &= \frac{1}{f_Y(Y)} \cdot \int_{-\infty}^{\infty} X \cdot f(Y|X) \cdot f_X(X) dx \end{aligned} \quad (62)$$

Porém; como $Y = X + V$, resulta:

$$f_Y(Y|X) = f_V(Y-X) \quad (63)$$

Então, (63) substituída em (62), resulta:

$$\hat{X} = \frac{\int_{-\infty}^{\infty} X \cdot f_V(Y-X) \cdot f_X(X) \cdot dx}{f_Y(Y)} \quad (64)$$

Lembrando que $f(X)$ é a função densidade de probabilidade de Gauss, podemos escrever:

$$\frac{d}{dx} f_X(X) = -\frac{X}{\sigma^2} f_X(X) \quad (65)$$

De (64) e (65) obteremos:

$$\begin{aligned} \hat{X} &= \frac{\int_{-\infty}^{\infty} f_V(Y-X) \cdot (-\sigma^2 \frac{d}{dx} f_X(X)) dx}{f_Y(Y)} = \\ &= \frac{-\sigma^2}{f_Y(Y)} \int_{-\infty}^{\infty} f_V(Y-X) \frac{d}{dx} f_X(X) dx \end{aligned} \quad (66)$$

Integrando (66) por partes tem-se:

$$\hat{X} = \frac{-\sigma^2}{f_Y(Y)} \left[f_X(X) \cdot f_V(Y-X) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f_X(X) \cdot \frac{d}{dx} \cdot f_V(Y-X) dx \right]$$

Mas, o primeiro termo é nulo porque $f_X(-\infty) = f_X(\infty) = 0$ e então:

$$\begin{aligned} \hat{X} &= \sigma^2 \frac{\int_{-\infty}^{\infty} f_X(X) \frac{d}{dx} f_V(Y-X) dx}{f_Y(Y)} \\ &= \frac{-\sigma^2}{f_Y(Y)} \int_{-\infty}^{\infty} f_X(X) \cdot \frac{d}{dy} f_V(Y-X) dx \end{aligned}$$

porque

$$\frac{d}{dx} f_V(Y-X) = - \frac{d}{dy} f_V(Y-X)$$

Como $f_X(X)$ independe de Y , pode-se escrever:

$$\hat{X} = \frac{-\sigma^2}{f_Y(Y)} \cdot \frac{d}{dy} \int_{-\infty}^{\infty} f_X(X) \cdot f_V(Y-X) dx \quad (67)$$

Mas sabemos que:

$$f_Y(Y) = \int_{-\infty}^{\infty} f_X(X) \cdot f_V(Y-X) dx$$

porque $Y = X + V$. Dessa maneira (67) pode ser finalmente posta sob a forma:

$$\hat{X} = -\sigma^2 \frac{\frac{d}{dy} \cdot f_Y(Y)}{f_Y(Y)} \quad (68)$$

5.3.1 - Análise do comportamento do estimador proposto

Analizaremos agora alguns sistemas estocásticos. Procuraremos determinar qual o comportamento do estimador dado por (68) para alguns sistemas.

CASO 1 - Suponhamos que X e Y sejam processos gaussianos independentes. Consequentemente Y é gaussiano. Esse caso é o mais simples, e como sabemos, resultará um estimador ótimo linear. Assim sendo:

$$f_y(Y) = \frac{1}{\sqrt{2\pi} \sigma_y} \cdot \text{Exp} - \left\{ \frac{1}{2} \frac{Y^2}{\sigma_y^2} \right\}$$

Como $Y = X + V$ e X e Y são independentes:

$$\sigma_y^2 = E [Y^2] = E [X^2] + E [V^2]$$

vamos adotar: $E [V^2] = \mu^2$ e como $E [X^2] = \sigma^2$ pode-se escrever:

$$f_y(Y) = \frac{1}{\sqrt{2\pi} (\sigma^2 + \mu^2)} \cdot \text{Exp} - \left\{ \frac{Y^2}{2(\mu^2 + \sigma^2)} \right\}$$

De acordo com (68) teremos

$$\hat{X} = -\sigma^2 \cdot \frac{(-1)}{(\mu^2 + \sigma^2)} Y = \frac{\sigma^2}{\sigma^2 + \mu^2} Y \quad (69)$$

Observemos a equação (69), mostrada na figura 44.

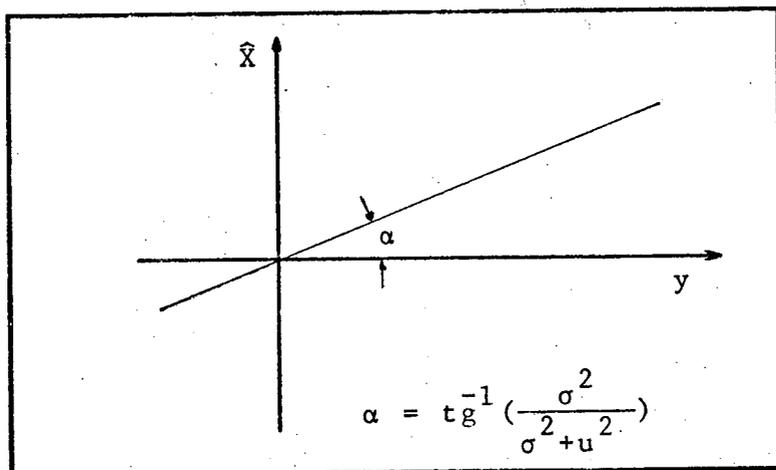


FIG. 44 - Estimador Linear

Como esperávamos, o estimador ótimo é linear. No caso - em que a variância de X é zero, o estimador ótimo é nulo, porque, $\sigma^2 = 0$ implica no total conhecimento de X , sendo portanto desnecessário um estimador. Nesse caso o ângulo $\alpha = 0$ e a reta da figura 44 se reduz ao eixo dos Y .

Se a variância σ^2 de X for infinita, implicando na falta absoluta de informação,

$$\alpha = \lim_{\sigma^2 \rightarrow \infty} \text{tg}^{-1} \left(\frac{\sigma^2}{\mu + \sigma^2} \right) = \text{tg}^{-1}(1) = \pi/4.$$

Então a máxima inclinação da reta de estimação será $\pi/4$

e então, $\hat{X} = Y$.

Podemos interpretar este resultado como a atribuição de pesos iguais para todas as observações $Y(n)$ na obtenção de $\hat{X}(n)$.

CASO 2 - Seja um sistema estocástico em que X seja um processo gaussiano, com $E[X^2] = \sigma^2$, V um processo não gaussiano tal que $E[V^2] = \mu^2$ e Y seja processo exponencial. Portanto:

$$f_y(Y) = \begin{cases} \alpha e^{-\alpha X} & , X \geq 0 \\ \alpha e^{\alpha X} & , X < 0 \end{cases} \quad (70)$$

Como:

$$E[Y^2] = \sigma_y^2 = \frac{2}{\alpha^2} = \sigma^2 + \mu^2 \quad \dots \quad \alpha = \sqrt{\frac{2}{(\sigma^2 + \mu^2)}} \quad (71)$$

Aplicando-se em (70) as relações (68) e (71), tem-se:

$$\hat{X} = \begin{cases} \frac{\sigma^2}{\sqrt{\frac{2}{\sigma^2 + \mu^2}}} = \sigma^2 \sqrt{\frac{\sigma^2 + \mu^2}{2}} & , Y \geq 0 \\ -\frac{\sigma^2}{\sqrt{\frac{2}{\sigma^2 + \mu^2}}} = -\sigma^2 \sqrt{\frac{\sigma^2 + \mu^2}{2}} & , Y < 0 \end{cases} \quad (72)$$

Plotamos a função (72) na figura 45.

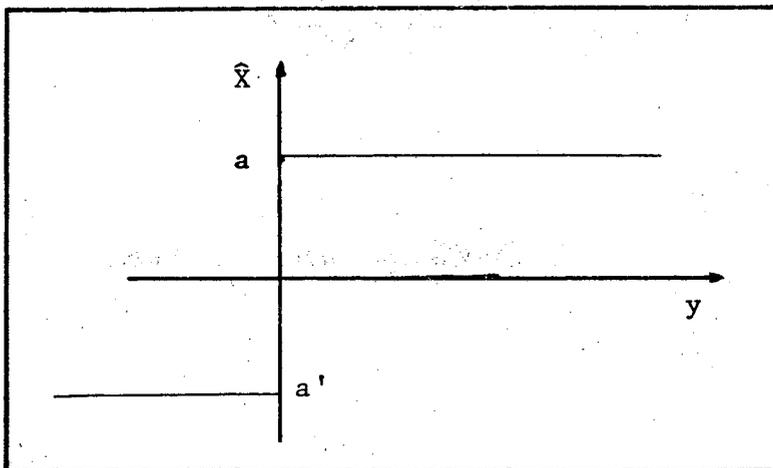


FIG. 45 - Estimador duplo salto.

Analizando-se a eq. (72) nota-se que novamente quando a informação em X for máxima, implicando em $\sigma^2 = 0$, o estimador \hat{X} se reduz a zero. Quando a informação a respeito de X for mínima,

i.é., se $\sigma^2 = \infty$, o estimador \hat{X} perde seu significado e os pontos a e a' da figura 45 se afastam infinitamente, resultando no eixo dos \hat{X} .

Qual o significado de tal comportamento do estimador \hat{X} , nesse sistema? Nos parece que a explicação reside em que a tentativa de minimização do erro de estimativa leva o estimador a desprezar os valores de X que se afastam substancialmente da média, o que é permitido pela própria distribuição de probabilidade do processo de observações. Quando a informação sobre o sistema é mínima, o estimador perde seu significado pela ação do comportamento assintótico da função densidade de probabilidade de Y .

CASO 3 - Vamos assumir um sistema em que X é um processo gaussiano como os anteriores, V um processo qualquer com $E[V^2] = \mu^2$, que resulte em ser Y um processo de Cauchy.

Então:

$$f_Y(Y) = \frac{1}{\pi(1+Y^2)}$$

em que $E[Y^2] = \infty$

Como:

$$E[Y^2] = E[X^2] + E[V^2] \quad \text{e} \quad E[X^2] = \sigma^2 < \infty$$

$$E[Y^2] = \infty$$

Usando-se novamente a eq. (68), teremos:

$$\hat{X} = \frac{2Y\sigma^2}{1 + Y^2} \quad (73)$$

A eq. (73) está mostrada na figura 46.

Na hipótese da máxima informação disponível ocorrer, este sistema, no que se refere ao seu estimador, comporta-se como os anteriores, reduzindo-se a zero. No caso contrário, i.é., supondo-se mínima a informação disponível, o estimador perde novamente o significado. Na curva isso implica na aproximação dos dois ramos da curva para o eixo dos \hat{X} .

O comportamento deste estimador (Eq.(73)), prevê perfeitamente a possibilidade de ocorrência de valores muito distanciados de seu valor esperado. Isso justifica o comportamento assintótico

tico da eq. (73), atribuindo a grandes valores* de Y um peso mínimo. Para os valores de Y próximo da média o estimador prevê pesos quase iguais.

Esse sistema então, do ponto de vista de estimação, é uma combinação dos dois anteriores.

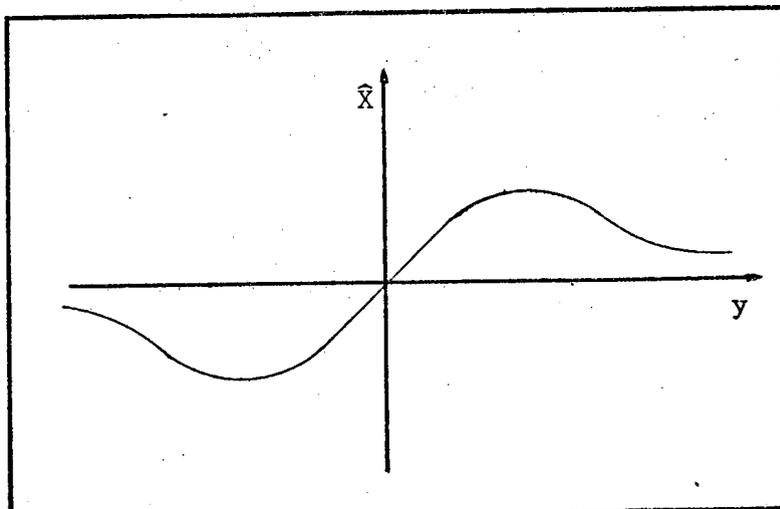


FIG. 46 - Estimador para processos de Cauchy.

CASO 4 - Nesse caso vamos analisar mais profundamente o comportamento assintótico de um sistema como o estudado no terceiro caso.

Admita-se Y como um processo estocástico, tal que sua função densidade de probabilidade tenha o comportamento assintótico de:

$$\lim_{Y \rightarrow \infty} f_y(Y) \rightarrow \frac{1}{Y^n}, \quad n > 2$$

Então, assintoticamente tem-se:

$$\frac{d}{dy} f_y(Y) \rightarrow \frac{-n}{Y^{n+1}}$$

Portanto, o estimador ótimo desse sistema, quando $Y \rightarrow \infty$, de acordo com (68), será da forma:

$$\hat{X} \rightarrow \frac{\sigma^2 n}{Y}$$

Essa forma de comportamento assintótico mais uma vez e-

* - OBS.: Assumimos como sempre que todos os processos tem média nula. Dessa forma, "Grandes Valores" significa grandemente afastados da média.

videncia-se. Como no caso 3, o estimador não permite a ocorrência de grandes desvios da média.

Por outro lado, nos processos encontrados na vida prática, a função densidade dos processos estocásticos X , Y , V são muito suaves, i.é., pelo menos a primeira derivada da função densidade de probabilidade é contínua. Processos com densidade de Laplace são muito incomuns.

Na figura 47 mostramos os diferentes tipos de distribuições de probabilidade encontráveis na prática e também os tipos de estimadores correspondentes.

Também pode-se assumir que na prática a função densidade de probabilidade não é assintoticamente maior que a densidade exponencial.

Em outras palavras, o ruído $\{V(n)\}$ não mascara completamente os valores de X , pois densidades de probabilidades como a de Cauchy são muito irrealis.

Assim podemos concluir que na prática o comportamento assintótico dos estimadores estará entre linear e duplo-salto, como os mostrados na figura 47 .

5.4 - PREVISÃO DISPONDO-SE DE UM CONJUNTO DE OBSERVAÇÕES.

Nesta secção propomos uma solução para o problema de prever o valor $X(n)$, sabendo-se $\{Y(n)\}$ e $\hat{X}(n-1)$, onde $\hat{X}(n-1)$ é o valor previamente estimado de $X(n-1)$.

Lembrando a eq. (59), e adaptando-a à presente situação ficamos com:

$$\min E [(X(n) - \hat{X}(n))^2 | Y(n), \hat{X}(n-1)] \quad (74)$$

Dessa forma, a forma funcional do estimador de $X(n)$ será:

$$\hat{X}(n) = \phi_n [\hat{X}(n-1), Y(n)] \quad (75)$$

Analizemos o problema por partes. Imagine-se a situação em que desconhecemos as informações $\{Y(n)\}$, dispondo-se somente de $\hat{X}(n-1)$.

Se o comportamento do sistema é dado por (58), o melhor valor que se pode obter por estimação de $X(n)$ será:

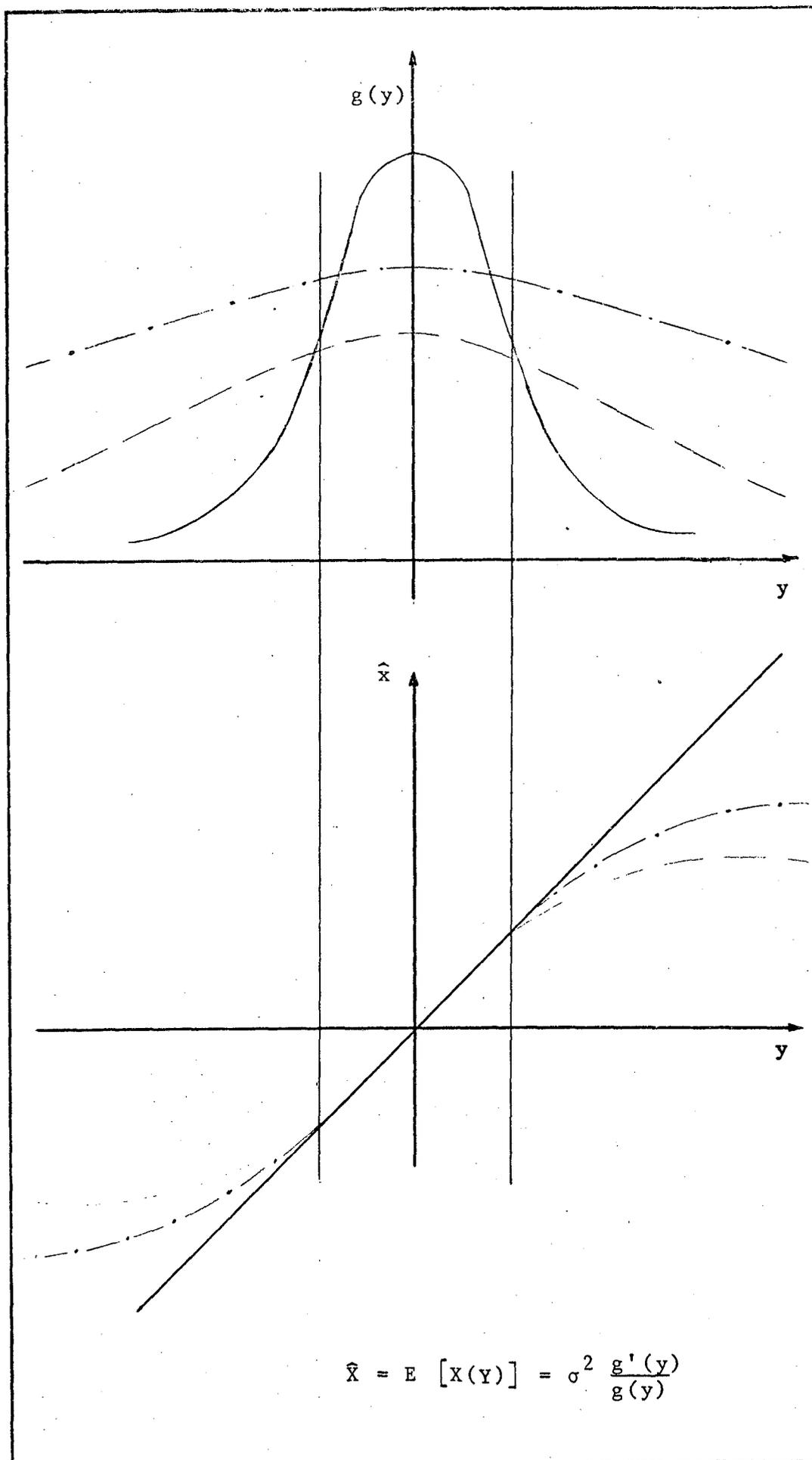


FIG. 47 - Os processos práticos e seus estimadores.

$$\hat{X}(n) = A \cdot \hat{X}(n-1)$$

Pois $W(n)$ e $V(n)$ são processos de média nula.

Da mesma forma o valor mais próximo da realidade que podemos obter por estimação para $Y(n)$ será:

$$Y(n) = \hat{X}(n) = A \cdot \hat{X}(n-1)$$

Porém, voltando ao nosso problema, dispomos das informações $\{Y(n)\}$, e é lógico que estas informações adicionais modificarão a forma da (74) para:

$$\hat{X}(n) = A \cdot \hat{X}(n-1) + \phi_n [Y(n) - A \cdot \hat{X}(n-1)] \quad (76)$$

onde ϕ_n é desconhecida e geralmente não linear.

A equação (76) tem uma estrutura conhecida como preditor-corretor, referindo-se aos dois termos do lado direito. O primeiro termo é chamado de preditor e o segundo de corretor.

Assim, usando (76) e (59), obtemos:

$$\min_{\phi_n} E [(X(n) - A \cdot \hat{X}(n-1) - \phi_n [Y(n) - A \cdot \hat{X}(n-1)])^2 | Y(n), \hat{X}(n-1)] \quad (77)$$

Definindo:

$$\xi(n) = X(n) - A \cdot \hat{X}(n-1) \quad (78)$$

e

$$\phi(n) = Y(n) - A \cdot \hat{X}(n-1) \quad (79)$$

e, usando a (77), obtemos:

$$\min_{\phi_n} E \{ [\xi(n) - \phi_n [\phi(n)]]^2 | \phi(n), \hat{X}(n-1) \} \quad (80)$$

Mas então, lembrando (59), pode-se escrever:

$$\hat{\xi}(n) = \phi_n [\phi(n)] \quad (81)$$

pois este problema é equivalente à determinação de $\hat{\xi}(n)$ como função de $\phi(n)$.

Apresentamos alguns exemplos a seguir da utilização da eq. (81):

Exemplo 1: Suponhamos que $\xi(n)$ e $\phi(n)$ sejam processos independentes e gaussianos. Então, de acordo com (69),

$$\hat{\xi}(n) = \alpha \phi(n)$$

onde

$$\alpha = \frac{\frac{\sigma_{\xi}^2}{2}}{\sigma_{\phi}}$$

e $\xi = X(n) - A.\hat{X}(n-1)$

conforme (78), então:

$$\hat{\xi} = \hat{X}(n) - A.\hat{X}(n-1)$$

Assim, a (81) fica:

$$\hat{X}(n) - A.\hat{X}(n-1) = \alpha(Y(n) - A.\hat{X}(n-1))$$

e finalmente:

$$\hat{X}(n) = A.\hat{X}(n-1) + \alpha(Y(n) - A.\hat{X}(n-1)) \quad (82)$$

Comparando (82) e (76) verificamos que:

$$\phi_n = \alpha(Y(n) - A.\hat{X}(n-1))$$

ou seja, ϕ_n é linear se os processos envolvidos são gaussianos.

Exemplo 2: Vamos considerar $\xi(n)$ e $\phi(n)$ processos independentes, $\xi(n)$ gaussiano, e $\phi(n)$ laplaciano.

De acordo com (80) e (70) escrevemos:

$$\hat{\xi}(n) = \begin{cases} K & , \quad \phi(n) \geq 0 \\ -K & , \quad \phi(n) < 0 \end{cases}$$

onde

$$K = \sigma_{\xi}^2 \sqrt{\frac{\sigma_{\phi}^2}{2}}$$

Então resulta para o estimador de $X(n)$:

$$\hat{X}(n) = A.\hat{X}(n-1) + \begin{cases} K & , \quad Y(n) - A.\hat{X}(n-1) \geq 0 \\ -K & , \quad Y(n) - A.\hat{X}(n-1) < 0 \end{cases}$$

Concluindo, considere-se o caso prático em que as distribuições de probabilidade de $\xi(n)$ e $\phi(n)$ são suaves como as mostradas na figura 47.

Então, as análises anteriormente feitas permitem - nos afirmar que:

$$\hat{X}(n) = A.\hat{X}(n-1) + \phi_n(Y(n) - A.\hat{X}(n-1))$$

em que ϕ_n é uma função cujo comportamento situa-se entre as fun

ções estimadoras mostradas na figura 47, e cuja forma algébrica - depende exclusivamente das distribuições de probabilidade de X , Y e V .

5.5 - PREDIÇÃO

Nesse caso o problema difere um pouco em sua estrutura e reside em determinar $\hat{X}(n+m)$, $m \geq 1$, sabendo-se um conjunto de n dados, $\{Y(n)\}$. Conhecemos o comportamento dinâmico do sistema, portanto sabemos que:

$$X(n+1) = A \cdot X(n) + W(n)$$

O único problema que podemos resolver com segurança consiste em estimar $\hat{X}(n)$, pois sabemos $Y(n)$.

Como então determinar $\hat{X}(n+m)$, $m > 1$?

A única solução racional, admitindo que $W(n)$ tem média nula, será fazermos:

$$\hat{X}(n+1) = A \cdot \hat{X}(n)$$

$$\hat{X}(n+2) = A \cdot \hat{X}(n+1) = A^2 \hat{X}(n)$$

e assim sucessivamente até:

$$\hat{X}(n+m) = A^m \hat{X}(n) .$$

É importante considerar que em cada estágio da previsão acima introduzimos um erro por desprezarmos o comportamento estocástico do sistema. Portanto, somente obteremos bons valores para $\hat{X}(n+m)$ se m for pequeno. Portanto este é um esquema de predição de curto prazo.

5.6 - DETERMINAÇÃO DA FUNÇÃO DE CORREÇÃO

Resta-nos solucionar somente um problema. Como determinar o comportamento estatístico de X , Y e V , e assim obtermos $\hat{\theta}_n$ em cada caso?

Nesse caso, somente podemos dispor das observações $Y(n)$. A única sugestão que podemos fazer é determinar se $Y(n)$ é gaussiana, o que podemos fazer aproximadamente através de um teste chi-quadrado.

Suponha-se que $\{Y(n)\}$ é gaussiano, então

$$V = \sum_{i=1}^n Y_i^2$$

tem distribuição χ^2 , com 1 grau de liberdade, pois Y_{i+1} depende de Y_i , ($i = 1, \dots, n$).

Com este teste poderemos, estabelecendo-se um grau de confiabilidade, determinar se a distribuição de probabilidade $Y(n)$ é gaussiana, e conseqüentemente, $X(n)$, $V(n)$ também o serão.

Caso o teste não seja plenamente satisfatório, i.é., muitos valores de V não pertencerem ao intervalo $\{\chi_{\max}^2, \chi_{\min}^2\}$, ter-se-á ao menos um julgamento do quanto a função difere da distribuição de Gauss.

Como preliminar do capítulo de simulação testamos o gerador de números gaussianos através de um teste de chi-quadrado e obtivemos bons resultados. Mostramos na figura 48 o resultado desse teste em 20 amostras de 1000 dados cada uma.

OBSERVAÇÃO

O VALOR DE QUI2 SERÁ CONSIDERADO SATISFATORIO SE ESTIVER NO INTERVALO 4.575-19.690

1	9.504	
2	7.174	
3	40.045	**
4	5.763	
5	28.674	**
6	15.112	
7	11.456	
8	7.250	
9	32.200	**
10	2.723	*
11	7.148	
12	9.455	
13	20.409	**
14	5.201	
15	5.921	
16	5.400	
17	6.067	
18	4.672	
19	5.377	
20	2.182	

FIG. 48 - Resultado do teste chi-quadrado aplicado ao gerador Gauss.

Outra técnica de grande valor prático pode ser utilizada - baseando-nos no trabalho de Wheelwright e Makridakis ¹⁷, para obtermos previsões adaptativas.

Esta técnica consiste em sempre atualizar as funções de - previsão, fazendo uso das novas observações a medida que estas possam ser obtidas.

Dessa forma, as funções e seus parâmetros são sequencial - mente ajustadas, de maneira que se reduzirão os erros para a próxima previsão.

Esta técnica é repetida até que um conjunto de parâmetros ótimos sejam obtidos.

C A P Í T U L O V I
=====6. CONCLUSÃO

Nosso trabalho não tem a pretensão de ser um trabalho definitivo sobre previsões. No entanto, esperamos ter alcançado nosso objetivo ao tratarmos o assunto de uma maneira unificada e abordando todos os problemas relacionados.

O método que propomos para obter-se previsões, dentro de sua característica inerente de recursividade, é bastante geral, pois explora as propriedades estatísticas dos processos estocásticos envolvidos.

A linearidade dos modelos que assumimos não limita a utilização do processo proposto quanto às possíveis distribuições de probabilidades envolvidas no sistema objeto.

Os sistemas, modelos e processos estocásticos que usamos são, mais do que se pode imaginar a primeira vista, parte do cotidiano daqueles que administram e/ou controlam sistemas sujeitos a estímulos aleatórios.

O uso dos modelos e do método de previsão para cálculo de estimativas é simples, e sua programação para computadores não requer sofisticação desde que o problema seja conhecido perfeitamente e dados sejam disponíveis.

Evidentemente que, se desconhecermos as distribuições de probabilidade dos processos estocásticos envolvidos, mesmo que se tenham dados, o problema se complica e nos parece que, até onde vai nosso conhecimento, não há solução conhecida.

A solução que sugerimos para determinarmos se um conjunto de dados tem distribuição gaussiana ou não, através de um teste chi-quadrado, somente tem sentido se o comportamento físico do sistema sugerir algo sobre a distribuição. Aí então o teste poderá somente confirmar a suspeita.

Através deste trabalho procuramos sempre manter uma linha

de ação de modo que os resultados pudessem sempre ser obtidos em computadores grandes ou pequenos. Por causa dessa característica' desejada optamos sempre por esquemas ou procedimentos recursivos.

Dentre o conjunto de resultados e conclusões que obtivemos, podemos enumerar:

1 - A análise de caráter microscópico dos modelos estocásticos nos levou à obtenção, embora de um modo muito amplo, dos limites de utilização dos modelos estocásticos lineares. Assim sendo, concluimos que os processos estocásticos considerados deverão pertencer aos sub-espacos de Hilbert em que o processo referido possa ser obtido de uma base ortogonal de processos de Ruído Branco Gaussiano. Como consequência, os processos estocásticos envolvidos de verão ter variância finita.

2 - A simulação é uma ferramenta útil ao problema e eficientemente usada poderá determinar o modelo mais adequado a um certo conjunto de dados.

3 - O método do estimador de máxima probabilidade é o que tem melhores características para solucionar o problema de estimação dos parâmetros que usamos.

4 - A previsão ótima nem sempre é linear. Nos casos práticos a previsão mais adequada situa-se entre linear e parcialmente linear. A forma funcional da previsão mais adequada depende fundamentalmente das características estatísticas dos processos estocásticos envolvidos. Destas, a função densidade de probabilidade da variável observada é a mais importante, pois define exclusivamente o termo de correção no esquema recursivo de predição-correção que adotamos.

5 - Testes estatísticos sobre as distribuições de probabilidade poderão ser úteis na determinação das distribuições dos dados, mas somente o conhecimento prévio ou o estudo do comportamento físico dos sistemas poderão definir a distribuição de probabilidade de um determinado conjunto de dados oriundo do sistema. A repetição dos testes, auxiliado pelo feedback dos resultados obtidos, poderá levar ao estabelecimento de critérios de avaliação dos resultados dos testes.

6 - Esquemas adaptativos são facilmente estabelecidos e apresentam excelentes resultados, levando inclusive à obtenção de um con

junto de parâmetros ótimos. O esquema adaptativo deve ser utilizado principalmente nos sistemas dinâmicos, em constante mutação, como os sistemas econômicos.

Na realidade, existe uma grande dificuldade em obter-se o modelo a partir dos dados. Deve-se complementar estas informações com conhecimentos a priori sobre o sistema e sobre o comportamento físico do processo, porque os dados disponíveis são limitados. Mesmo se possuímos muitos dados, não poderemos dar-lhes total confiança porque muitas mudanças ocorrem e afetam indiretamente os sistemas.

Para obtermos os parâmetros A e B fizemos a aproximação de que não há erro nas observações Y, i.é., $Y = X$. Posteriormente, à determinação de A e B, consideramos observações com ruído para prever X.

Nossa aproximação é melhor, se conhecermos os parâmetros do modelo. Caso contrário, nosso método de determinação dos parâmetros e da previsão é sub-ótimo.

Assumimos modelos recursivos para previsões pela facilidade de computação. No entanto, pode ocorrer que, se a exatidão das previsões é um requisito importante, os modelos não recursivos sejam melhores.

Finalmente sugerimos um esquema de avanço cronológico sobre o problema prático de previsão, e que mostramos graficamente na figura 49.

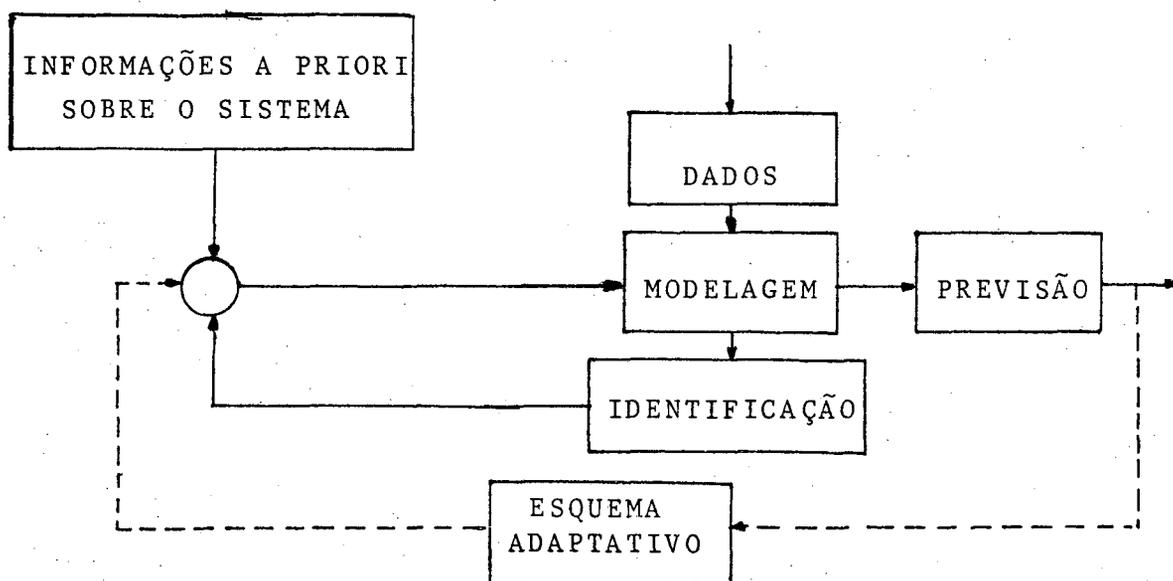


FIG. 49 - Esquema de avanço cronológico do problema

Observe-se o relacionamento indissolúvel entre a Modelagem e a Identificação, pois a Identificação faz uso do Modelo e fornece parâmetros para o mesmo. Informações a priori são indispensáveis para um problema físico real. Evidentemente que um esquema adaptativo só funcionará após o primeiro ciclo do problema, i.ê., após a obtenção da primeira previsão.

NOTAÇÃO E SIMBOLOGIA USADA

1. ESCALARES

a, b	:	coeficientes do modelo escalar
x, u, y	:	componentes dos processos aleatórios
t	:	tempo em sua forma contínua
n	:	tempo em sua forma discreta
w	:	um experimento
μ	:	média
σ^2	:	variância
β	:	coeficiente da transformação de estado
α	:	parâmetro de distribuições de probabilidade
R	:	risco esperado

2. VETORES

X, Y	:	processos estocásticos
W	:	processo de Ruído Branco Gaussiano
V	:	processo de Ruído Branco Genérico
θ	:	vetor dos parâmetros de um modelo
$\hat{\theta}$:	vetor dos estimadores do modelo

3. MATRIZES

A, B, C, D	:	matrizes coeficientes dos modelos vetoriais
Λ	:	matriz de covariância

4. FUNÇÕES ESCALARES

$f(.)$:	função densidade de probabilidade
$f(. .)$:	função densidade de probabilidade condicional
$f(.,.,.,.,.,.)$:	função densidade de probabilidade conjunta
$F(.)$:	função densidade de probabilidade acumulada
$F^{-1}(.)$:	função densidade de probabilidade acumulada inversa.
$\text{Exp}\{.\}$:	exponencial
δ_{JK}	:	função salto (Delta de Kronecker)

$\phi(\cdot)$:	função genérica de correção
$g(\cdot)$:	função distribuição de probabilidade genérica
$\psi(\cdot)$:	função genérica de previsão

5. FUNÇÃO DE CONJUNTO

$P\{\cdot\}$:	probabilidade
$P\{\cdot \cdot\}$:	probabilidade condicional

6. CONJUNTOS

I	:	conjunto indicial
Ω	:	espaço amostral
Ω_0	:	conjunto dos resultados da Hipótese 0
Ω_1	:	conjunto dos resultados possíveis na Hipótese 1

7. OPERADORES

$E[\cdot]$:	espectância
$E[\cdot \cdot]$:	espectância condicional
$(\cdot)'$:	transposto
$\frac{d}{dx}$:	derivada
\int	:	integral
$\det(\cdot)$:	determinante
$\langle \cdot \rangle$:	produto interno
Σ	:	somador
Π	:	produtório

8. NORMA

$$\| \phi_i \| = \sqrt{\sum_{i=1}^n \phi_i^2}$$

9. ESPAÇOS

R^n	:	espaço euclideano de dimensão n
\mathcal{H}	:	espaço de Hilbert
H	:	sub-espaço do espaço de Hilbert

10. SÍMBOLOS

ϵ	:	pertence a
\perp	:	perpendicular a
$ $:	dado que
\forall	:	qualquer
U	:	união
$\underline{\Delta}$:	igual por definição
$\{ \cdot \}$:	conjunto (processo)
$ \cdot $:	módulo
$\hat{\cdot}$:	valor estimado de

REFERÊNCIAS BIBLIOGRÁFICAS
 =====

- [1] AGUIAR, Márcia B. de - "Gerador de números aleatórios para o sistema IBM/1130" - Rio de Janeiro, 1972. (Tese para obtenção do grau de Mestre em Ciências em Informática).
- [2] AIGNER, Dennis J. - "Basic Econometrics" - New Jersey, Prentice Hall, 1971.
- [3] ASTROM, Karl J. - "Introduction to Stochastic Control Theory" - New York, Academic Press, 1970, p. 225-230.
- [4] CHOU, Y. A-lun - "Statistical Analysis with Business and Economic Applications" - New York, Holt, Rinehart & Winston, 1969, p. 449-452.
- [5] FRIDMAN, Milton A. - "Theory of Consumption Function"-Princeton University Press, 1957.
- [6] IBM - "IBM 1130/1800 Plotter Subroutines Manual"- C26-3755-0
- [7] KAILATH, T. - "An Innovations Approach to Least Square Estimation" - Part 1. IEEE Transactions on Automatic Control, V.AC 13, n.6, Dec.1968.
- [8] KALMAN, R.E. - "A New approach to linear filtering and Prediction Problems" - Transactions of ASME, V.82,p. 34-45, Mar.1960.
- [9] KNUTH, Donald E. - "The Art of Computer Programming" - Reading, Addison-Wesley, 1969, V.2 .
- [10] LUENBERGER, David G. - "Optimization by Vector Spaces Methods" - New York, J.Willey & Sons, 1969, p. 78-82 .
- [11] MOOD, Alexander M. & GRAYBILL, Franklin A. - "Introduction to the Theory of Statistics" - 2.ed.,Tokyo, McGraw-Hill - 1963.

- [12] OWEN, D.B. - "Handbook of Statistical Tables" - Reading , Addison-Wesley, 1962, p. 49-56 .
- [13] PACITTI, Tércio - "Fortran-Monitor Princípios" - 2.ed.-Rio de Janeiro, Livro Técnico, 1968.
- [14] PORTER, Willian A. - "Modern Foundations of Systems Engineering" - New York, MacMillan, 1969, p. 60-70 .
- [15] VAN TREES, Harry L. - "Detection, Estimation and Modulation Theory" - J.Willey & Sons, New York, 1968.
- [16] VAN TREES, Harry L. - "Applications of State Variable Techniques in Detection Theory" - May 1970, Proceedings of IEEE, V.58, n.5 .
- [17] WHEELWRIGHT, S.C. & MAKRIDAKIS, Spyros - "An Examination - of the use of adaptive filtering in forecasting"-March 1973, Operational Research Quarterly, V.24, n.1 .
- [18] ZELLNER, A. & GEISEL, M.S. - "Analysis of Distributed Lag Models with Applications to Consumption Function Estimation" - Nov.1970, Econometrica, V.38, n.6 .