Filtragem Adaptativa

Charles Casimiro Cavalcante

charles@gtel.ufc.br

Grupo de Pesquisa em Telecomunicações Sem Fio – GTEL
Programa de Pós-Graduação em Engenharia de Teleinformática
Universidade Federal do Ceará – UFC
http://www.gtel.ufc.br/∼charles

"Filtros adaptativos, os quais têm como meta transformar os sinais portadores de informação em versões 'limpas' ou 'melhoradas', ajustam suas características de acordo com os sinais encontrados. Eles formam os exemplos mais simples de algoritmos no campo de aprendizado de máquinas."

Philip A. Regalia, 2005

IEEE Control Systems Magazine, Agosto de 2005

Conteúdo do curso

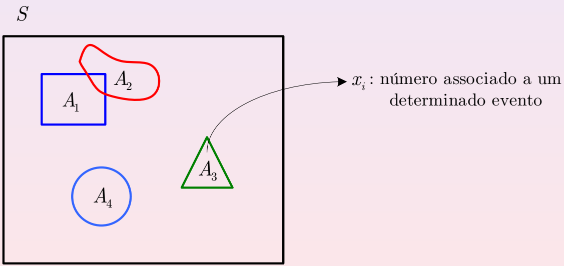
1. Introdução
2. Revisão de Processos Estocásticos
3. Filtragem Linear Ótima
4. Algoritmos Recursivos no Tempo
5. Método dos Mínimos Quadrados
6. Estruturas Alternativas de Filtragem Adaptativa
7. Tópicos Avançados

Parte III

Estatísticas de Segunda Ordem

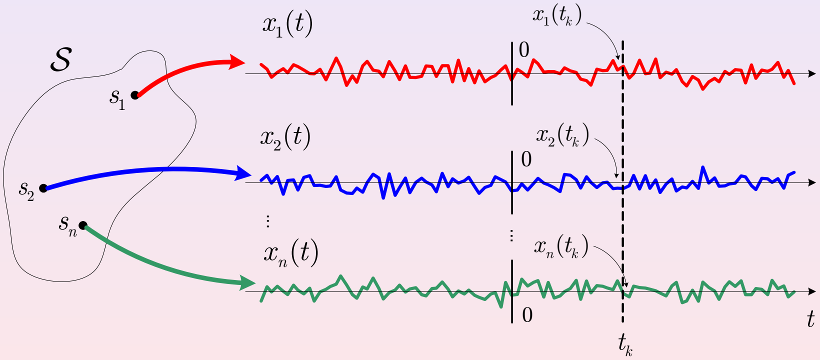
Processos estocásticos discretos

Sinal discreto $x(n)$: seqüencia de números indexada por um argumento $n \in \mathbb{Z}$. Constitui um sinal aleatório na medida em que, para um dado $n = n_0$ pré-estabelecido, existe alguma incerteza sobre o valor $x(n_0)$, ou seja, $x(n_0)$ é uma variável aleatória.



Processo estocástico

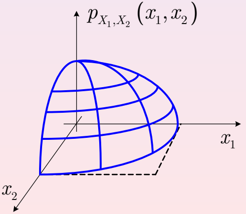
Conjunto das possíveis realizações de um dado sinal aleatório



Como caracterizar $x(n_0)$?

1. função de densidade de probabilidade (pdf): $p_{X_0}(x_0)$
2. função de distribuição (cumulativa) de probabilidade: $F_{X_0}(x_0)$

$$\left.\begin{array}{c}\underbrace{\left[\begin{array}{ccc} x_{-1} & \cdots & x_K \end{array}\right]}_{\substack{\text{intervalo de realização}\\ \text{do processo}}}\end{array}\right\} p_{X_{-1},\cdots,X_K}(x_{-1},\cdots,x_K)$$



Introdução geral - cont.

Estacionaridade

Propriedades estatísticas se preservam com translação no tempo

$$p_{X_L,\cdots,X_K}(x_L,\cdots,x_K) = p_{X_{L+n_0},\cdots,X_{K+n_0}}(x_{L+n_0},\cdots,x_{K+n_0}) \tag{22}$$

Ergodicidade

Comportamento estatístico é idêntico ao comportamento temporal

$$\underset{\forall i}{p_{X^{(i)}}\left(x^{(i)}\right)} = \underset{\forall j}{p_{X_j}(x_j)} \tag{23}$$

Correlação

Seja um vetor de v.a. $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \end{bmatrix}$, temos que a *correlação* entre dois elementos, i e j, como

$$\begin{aligned} r_{ij} = \mathbb{E}\{x_i x_j\} &= \int_{-\infty}^{\infty} x_i x_j^* p_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_i x_j^* p_{x_i,x_j}(x_i, x_j)\, dx_i\, dx_j \end{aligned} \tag{24}$$

De uma maneira mais intuitiva, a correlação pode ser definida como a medida de relação entre as duas variáveis aleatórias. Note ainda que a mesma pode assumir valores positivos ou negativos.

Matrix de correlação

Podemos unificar uma notação para avaliar todos os pares de variáveis na correlação. Desta forma, definimos a *matriz de correlação* definida como

$$\begin{aligned}\mathbf{R}_{\mathbf{x}} &= \mathbb{E}\left\{\mathbf{x}\mathbf{x}^H\right\} \\ &= \begin{bmatrix} \mathbb{E}\{|x_1|^2\} & \mathbb{E}\{x_1 x_2^*\} & \cdots & \mathbb{E}\{x_1 x_N^*\} \\ \mathbb{E}\{x_2 x_1^*\} & \mathbb{E}\{|x_2|^2\} & \cdots & \mathbb{E}\{x_2 x_N^* *\} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}\{x_N x_1^*\} & \mathbb{E}\{x_N x_2^*\} & \cdots & \mathbb{E}\{|x_N|^2\} \end{bmatrix}_{N\times N}\end{aligned} \tag{25}$$

Se escrevermos o vetor $\mathbf{x} = \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_i \end{bmatrix}$ explicitando suas partes real $(\mathbf{x}_r)$ e imaginária $(\mathbf{x}_i)$, temos

$$\boxed{\begin{aligned} \mathbb{E}\left\{\mathbf{x}\mathbf{x}^H\right\} &= \mathbb{E}\left\{\begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_i \end{bmatrix} \begin{bmatrix} \mathbf{x}_r^H & \mathbf{x}_i^H \end{bmatrix}\right\} \\ &= \begin{bmatrix} \mathbb{E}\{\mathbf{x}_r\mathbf{x}_r^H\} & \mathbb{E}\{\mathbf{x}_r\mathbf{x}_i^H\} \\ \mathbb{E}\{\mathbf{x}_i\mathbf{x}_r^H\} & \mathbb{E}\{\mathbf{x}_i\mathbf{x}_i^H\} \end{bmatrix} \end{aligned}} \tag{26}$$

Desta forma, denotamos

$$\begin{aligned} \mathbb{E}\{\mathbf{x}_r\mathbf{x}_r^H\} = \mathbb{E}\{\mathbf{x}_i\mathbf{x}_i^H\} = \mathbf{R}_{\mathbf{x}}^{\mathsf{E}} \\ \mathbb{E}\{\mathbf{x}_i\mathbf{x}_r^H\} = -\mathbb{E}\{\mathbf{x}_r\mathbf{x}_i^H\} = \mathbf{R}_{\mathbf{x}}^{\mathsf{O}} \end{aligned} \tag{27}$$

em que E e O denotam as partes par (even) e ímpar (odd).

⇒ **Propriedades de $\mathbf{R_x}$**

1. É uma matriz *simétrica*: $\mathbf{R_x} = \mathbf{R_x}^H$
2. É uma matriz *semi-definida positiva*

$$\mathbf{a}^H \mathbf{R_x} \mathbf{a} \geq 0 \tag{28}$$

para qualquer vetor N-dimensional $\mathbf{a} \neq \mathbf{0}$. Na prática, geralmente $\mathbf{R_x}$ é definida positiva para qualquer vetor N-dimensional $\mathbf{a} \neq \mathbf{0}$

3. Todos os autovalores de $\mathbf{R_x}$ são reais e *não-negativos* (positivos se $\mathbf{R_x}$ for definida positiva). Além disso, os autovetores de $\mathbf{R_x}$ são reais e podem sempre ser escolhidos tal que sejam *mutuamente ortogonais*.
4. $\mathbf{R_x} = 2\mathbf{R_x^{\mathsf{E}}} + j2\mathbf{R_x^{\mathsf{O}}}$

⇒ **Propriedades de $\mathbf{R_x}$** (cont.)

5. É uma matriz Toeplitz
6. A correlação indica quanto um sinal "tem haver" com outro
7. Quanto mais lento for o processo, maior a correlação
8. Uma correlação positiva indica que o crescimento (decrescimento) de uma v.a implica em crescimento (decrescimento) da outra e uma correlação negativa indica que o crescimento (decrescimento) de uma v.a. implica no decrescimento (crescimento) da outra

Seja $v(n)$ um ruído branco, ou seja, $\mathbf{R_v} = \sigma_v^2\mathbf{I}$. Seja então um sinal qualquer

$$s(n) = \underbrace{\underbrace{x(n)}_{\text{sinal de interesse}} + \underbrace{v(n)}_{\text{ruído branco}}}_{\textbf{independentes}}$$

Se passarmos a seqüência $s(n)$ por um filtro

$$\mathbf{w} = [\ w_0\ w_1\ \ldots\ w_{N-1}\ ],$$

teremos então a seguinte saída

$$y(n) = \mathbf{w}^H\mathbf{x}(n) + \mathbf{w}^H\mathbf{v}(n)$$

Temos então:

$$\begin{aligned} \text{SNR}_{\text{saída}} &= \frac{\text{Potência interesse}}{\text{Potência ruído}} \\ &= \frac{\mathbb{E}\{\mathbf{w}^H \mathbf{x}(n)\mathbf{x}^H \mathbf{w}(n)\}}{\mathbb{E}\{\mathbf{w}^H \mathbf{v}(n)\mathbf{v}^H \mathbf{w}(n)\}} \\ &= \frac{\mathbf{w}^H \mathbf{R}_\mathbf{x} \mathbf{w}}{\mathbf{w}^H \mathbf{R}_\mathbf{v} \mathbf{w}} = \frac{\mathbf{w}^H \mathbf{R}_\mathbf{x} \mathbf{w}}{\mathbf{w}^H \sigma_v^2 \mathbf{I} \mathbf{w}} \end{aligned} \tag{29}$$

Então, escreve-se

$$\underbrace{\boxed{\frac{\mathbf{w}^H \mathbf{R}_\mathbf{x} \mathbf{w}}{\mathbf{w}^H \mathbf{w}}}}_{\substack{\textbf{coeficiente} \\ \textbf{de Rayleigh}}} \cdot \frac{1}{\sigma_v^2} \tag{30}$$

- Relembrando: momentos centrados são definidos de maneira similar aos momentos usuais, apenas a média é envolvida no cálculo da esperança.

Definimos então a **matriz de covariância** de $\mathbf{x}$ como

$$\boxed{\mathbf{C}_{\mathbf{x}} = \mathbb{E}\left\{(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^H\right\}} \tag{31}$$

e os elementos

$$c_{ij} = \mathbb{E}\left\{(x_i - \mu_i)(x_j - \mu_j)^H\right\} \tag{32}$$

da matriz $\mathbf{C}_{\mathbf{x}}$ de dimensão $N \times N$, são chamados de *covariâncias* e eles são os momentos centrados correspondentes às correlações r_{ij}.

- A matriz de covariância $\mathbf{C_x}$ possui as mesmas propriedades de simetria que a matriz de correlação $\mathbf{R_x}$.
- Usando as propriedades do operador esperança, é fácil mostrar que

$$\mathbf{R_x} = \mathbf{C_x} - \boldsymbol{\mu}_\mathbf{x}\boldsymbol{\mu}_\mathbf{x}^H \tag{33}$$

- Se o vetor de médias for $\boldsymbol{\mu}_\mathbf{x} = \mathbf{0}$, as matrizes de correlação e covariância são as mesmas

Para esperanças conjuntas, ou seja, envolvendo mais de uma variável aleatória, podemos definir a **matriz de correlação cruzada**

$$\mathbf{R}_{\mathbf{xy}} = \mathbb{E}\left\{\mathbf{x}\mathbf{y}^H\right\} \tag{34}$$

e a **matriz de covariância cruzada**

$$\mathbf{C}_{\mathbf{xy}} = \mathbb{E}\left\{(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})^H\right\} \tag{35}$$

Note que as dimensões dos vetores $\mathbf{x}$ e $\mathbf{y}$ podem ser diferentes. Assim, as matriz de correlação e covariância cruzadas não são necessariamente quadradas e são, em geral, não-simétricas. Entretanto, de suas definições segue que:

$$\begin{aligned} \mathbf{R}_{\mathbf{xy}} &= \mathbf{R}_{\mathbf{yx}}^H \\ \mathbf{C}_{\mathbf{xy}} &= \mathbf{C}_{\mathbf{yx}}^H \end{aligned} \tag{36}$$

Quando temos uma soma de dois vetores $\mathbf{x}$ e $\mathbf{y}$, temos as seguintes relações:

1. Correlação

$$\mathbf{R_{x+y}} = \mathbf{R_x} + \mathbf{R_{xy}} + \mathbf{R_{yx}} + \mathbf{R_y} \tag{37}$$

2. Covariância

$$\mathbf{C_{x+y}} = \mathbf{C_x} + \mathbf{C_{xy}} + \mathbf{C_{yx}} + \mathbf{C_y} \tag{38}$$

Vale relembrar que:

- Variáveis ortogonais implica em correlação zero $(\mathbf{R_{xy}} = \mathbf{0})$
- Variáveis descorrelacionadas implica em covariância zero, $(\mathbf{C_{xy}} = \mathbf{0})$

Então, temos

1. $\mathbf{R_{x+y}} = \mathbf{R_x} + \mathbf{R_y}$ para $\mathbf{x}$ e $\mathbf{y}$ ortogonais
2. $\mathbf{C_{x+y}} = \mathbf{C_x} + \mathbf{C_y}$ para $\mathbf{x}$ e $\mathbf{y}$ descorrelacionados

- Transformação tempo-freqüência diz respeito à transformada de Fourier quando estudamos sinais determinísticos
- A transformação requer o conhecimento da função do sinal a ser avaliada na freqüência

$$\begin{aligned} S(\omega) &= \mathfrak{F}\{s(t)\} \\ &= \int\limits_{-\infty}^{\infty} s(t)\exp(-j\omega t)\ dt \end{aligned} \tag{39}$$

- **Pergunta**: como fazer no caso de sinais aleatórios? Não se conhece os valores do sinal com precisão e a integral pode não existir.
- **Resposta**: definir de outra forma a transformada de Fourier de um processo aleatório

Classificação de processos estocásticos

1. Estacionários no *sentido estrito* (SSS)
    - TODAS as estatísticas (momentos) são independentes do tempo, ou seja,

$$p_{X(t)}[x(t)] = p_{X(t)}(x)$$

2. Estacionários no *sentido amplo* (WSS)
    - Apenas as estatísticas de primeira e segunda ordem (média e correlação) são independentes do tempo, ou seja,

$$E\{X(t)\} = \mu$$
$$E\left\{(X(t)-\mu)^2\right\} = \sigma_x^2$$

Definição

Para a classe de processos estocásticos WSS definimos a **função de densidade espectral** como a transformada de Fourier da correlação na forma

$$S_x(\omega) = \mathfrak{F}\{r_x(\tau)\} = \int_{-\infty}^{\infty} r_x(\tau)\exp(-j\omega\tau)\ d\tau \tag{40}$$

A função de densidade espectral de potência é um indicador da distribuição da potência do sinal como uma função da freqüência.

Pode-se ainda calcular a correlação a partir da função de densidade espectral, uma vez que a transformada de Fourier é única, como

$$r_x(\tau) = \mathfrak{F}^{-1}\{S_x(\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_x(\omega)\exp(j\omega\tau)\ d\omega \qquad (41)$$

NOTA: Valem ressaltar que τ significa a diferença entre os tempos que os processos WSS requerem como parâmetro para caracterizar a correlação. Neste caso, seria o equivalente à diferença entre as amostras i e j que temos na correlação definida na Equação (24)

Propriedades

1. O valor médio quadrático de um processo WSS é dado por

$$E\{X^2(t)\} = r_x(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(\omega)\; d\omega \qquad (42)$$

2. A densidade espectral de potência de um processo WSS é sempre não-negativa

$$S_x(\omega) \geq 0 \quad \text{para todo } \omega \qquad (43)$$

Propriedades - cont.

3. A densidade espectral de potência de um processo WSS real é uma função par de ω

$$S_x(\omega) = S_x(-\omega) \tag{44}$$

4. O valor da densidade espectral de potência em $\omega = 0$ é

$$S_x(0) = \int_{-\infty}^{\infty} r_x(\tau)\, d\tau \tag{45}$$

Momentos de ordem superior

- São estatísticas de **ordem superior a 2**
- Descrevem o comportamento estatístico dos dados com alguma informação a mais
- São importante por portarem informação da fase dos dados
- O momento de ordem 3 recebe o nome de obliquidade (*skewness* em inglês) e mede a assimetria da distribuição
- O momento de ordem 4 recebe o nome de **kurtosis** e é denotada por $\mathcal{K}$
- Podemos também classificar as distribuições através da kurtosis
    1. Para $\mathcal{K} = 0$: distribuição *mesocúrtica* (gaussiana)
    2. Para $\mathcal{K} > 0$: distribuição *leptocúrtica* (super-gaussiana)
    3. Para $\mathcal{K} < 0$: distribuição *platicurtica* (sub-gaussiana)

Primeira função característica

$$C(\omega) \triangleq \int_{-\infty}^{\infty} p_X(x) \cdot \exp(j\omega x)\ dx \tag{46}$$

Importante

- $C(-\omega) \triangleq \int_{-\infty}^{\infty} p_X(x) \cdot \exp(-j\omega x)\ dx$
- $C(-\omega) = \mathfrak{F}\{p_X(x)\}$ (transformada de Fourier de $p_X(x)$)

Importante - cont.

- Notação

$$
\begin{aligned}
P_X(\omega) = \mathfrak{F}\{p_X(x)\} &= \int_{-\infty}^{\infty} p_X(x) \cdot \exp(-j\omega x)\ dx \\
&= \mathbb{E}\{\exp(-j\omega X)\}
\end{aligned}
$$

- Assim

$$
\boxed{p_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P_X(\omega) \cdot \exp(j\omega x)\ d\omega}
$$

Primeira função característica - Variáveis discretas

$$
\begin{aligned}
P_X(w) &= \sum_{i=-\infty}^{\infty} \Pr\{X = x_i\} \cdot \exp\left(-j\omega x_i\right) \\
&= \mathbb{E}\{\exp(-j\omega X)\}
\end{aligned}
$$

Propriedades $P_X(\omega)$

1. $|P_X(\omega)| \leq 1$
   Prova:

$$|P_X(\omega)| = \left| \int_{-\infty}^{\infty} p_X(x) \cdot \exp(-j\omega x)\ dx \right|$$

$$\leq \int_{-\infty}^{\infty} |p_X(x)| \cdot |\exp(-j\omega x)|\ dx = \int_{-\infty}^{\infty} p_X(x)\ dx$$

desigualdade de Schwartz

2. $P_X(0) = 1$

Função geradora de momentos

$$P_X(\omega) = \mathbb{E}\{\exp(-j\omega X)\}$$

$$\exp(-j\omega x) = \sum_{k=-\infty}^{\infty} \frac{(-j\omega X)^k}{k!} \quad \text{(Série de Taylor em torno de } X = 0\text{)}$$

$$\mathbb{E}\{\exp(-j\omega x)\} = \mathbb{E}\left\{\sum_{k=-\infty}^{\infty} \frac{(-j\omega)^k \cdot X^k}{k!}\right\}$$

$$\mathbb{E}\{\exp(-j\omega x)\} = \sum_{k=-\infty}^{\infty} \frac{(-j\omega)^k}{k!} \cdot \underbrace{\mathbb{E}\left\{X^k\right\}}_{\mu_k}$$

$$\Rightarrow \boxed{P_X(\omega) = \sum_{k=-\infty}^{\infty} \frac{(-j\omega)^k}{k!} \cdot \mu_k}$$

Função geradora de momentos - cont.

Ainda

$$P_X(\omega) = \sum_{k=0}^{\infty} \left.\frac{d^k P_X(\omega)}{d\omega^k}\right|_{\omega=0} \cdot \frac{1}{k!} \cdot \omega^k$$

Logo

$$\sum_{k=0}^{\infty} \left.\frac{d^k P_X(\omega)}{dw^k}\right|_{\omega=0} \cdot \frac{\omega^k}{k!} = \sum_{k=0}^{\infty} \mu_k \cdot (-j)^k \cdot \frac{\omega^k}{k!}$$

$$\boxed{\mu_k \cdot (-j)^k = \left.\frac{d^k P_X(\omega)}{dw^k}\right|_{\omega=0}}$$

Exemplo

Sejam X e Y v.a.s independentes com $p_X(x)$ e $p_Y(y)$ conhecidas. Se $Z = X + Y$, $p_Z(z) = ?$

Solução

$$P_Z(\omega) = \mathbb{E}\{\exp(-j\omega Z)\} = \mathbb{E}\{\exp[-j\omega(X+Y)]\}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-j\omega X) \cdot \exp(-j\omega Y) \cdot p_{X,Y}(x,y)\ dx\ dy$$

$$\underbrace{\int_{-\infty}^{\infty} \exp(-j\omega X) \cdot p_X(x)\ dx}_{\mathbb{E}\{\exp(-j\omega X)\}} \cdot \underbrace{\int_{-\infty}^{\infty} \exp(-j\omega Y) \cdot p_Y(y)\ dy}_{\mathbb{E}\{\exp(-j\omega Y)\}}$$

$$P_Z(\omega) = P_X(\omega) \cdot P_Y(\omega)$$

$$\boxed{p_Z(z) = p_X(x) \star p_Y(y)}$$

Segunda função característica

$$\Psi(\omega) = \ln[P_X(\omega)] \tag{47}$$

Importante

- A segunda função característica é também chamada de **função geradora de cumulantes**
- Os cumulantes são de extrema importância na caracterização estatística de uma v.a.

História

Os cumulantes foram inicialmente introduzidos pelo astrônomo, contador, matemático e estaticista dinamarquês Thorvald N. Thiele (1838-1910) que os denominou *semi-invariantes*.

O termo *cumulante* surgiu pela primeira vez em 1931 no artigo "The Derivation of the Pattern Formulæ of Two-Way Partitions from Those of Simpler Patterns", Proceedings of the London Mathematical Society, Series 2, vol. 33, pp. 195-208, publicado pelo geneticista e estaticista Sir Ronald Fisher e o estaticista John Wishart, epônimo da distribuição de Wishart.

O historiador Stephen Stigler comenta que o termo cumulante foi sugerido a Fisher numa carta de Harold Hotelling. Em um outro artigo publicado em 1929, Fisher chamou-os de funções de momentos cumulativos.

Definição

O cumulante de ordem k é definido como

$$\boxed{\kappa_k = \frac{\partial^k \Psi(\omega)}{\partial \omega^k}} \tag{48}$$

Propriedades dos cumulantes

1. **Invariância e equivariância**

$$\kappa_1(Y+\alpha) = \kappa_1(Y) + \alpha$$
$$\kappa_k(Y+\alpha) = \kappa_k(Y)$$

para α uma constante qualquer.

2. **Homogeneidade** (ou multilinearidade)

$$\kappa_k(\alpha Y) = \alpha^k \cdot \kappa_k(Y)$$

3. **Aditividade**

$$\kappa_k(X+Y) = \kappa_k(X) + \kappa_k(Y)$$

se X e Y são v.a.s independentes

Cumulantes e momentos

Os cumulantes são relacionados com os momentos através da seguinte recursão:

$$\boxed{\kappa_k = \mu_k - \sum_{i=1}^{k-1} \begin{pmatrix} k-1 \\ i-1 \end{pmatrix} \kappa_i \cdot \mu_{k-i}} \tag{49}$$

Cumulantes e momentos - cont.

Desta forma, o $k-$ésimo momento é um polinômio de grau k dos k primeiros cumulantes, dados, para o caso em que $k = 6$, na seguinte forma:

$$
\begin{aligned}
\mu_1 &= \kappa_1 \\
\mu_2 &= \kappa_2 + \kappa_1^2 \\
\mu_3 &= \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3 \\
\mu_4 &= \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4 \\
\mu_5 &= \kappa_5 + 5\kappa_4\kappa_1 + 10\kappa_3\kappa_2 + 10\kappa_3\kappa_1^2 + 15\kappa_2^2\kappa_1 + 10\kappa_2\kappa_1^3 \\
\mu_6 &= \kappa_6 + 6\kappa_5\kappa_1 + 15\kappa_4\kappa_2 + 15\kappa_4\kappa_1^2 + 10\kappa_3^2 + 60\kappa_3\kappa_2\kappa_1 + \\
&\quad 20\kappa_3\kappa_1^3 + 15\kappa_2^3 + 45\kappa_2^2\kappa_1^2 + 15\kappa_2\kappa_1^4 + \kappa_1^6.
\end{aligned}
$$

Em geral, podemos escrever a descrição de um modelo de entrada e saída de um modelo estocástico como

$$\begin{pmatrix} \text{valor atual} \\ \text{da saída} \\ \text{do modelo} \end{pmatrix} + \begin{pmatrix} \text{combinação linear} \\ \text{dos valores passados} \\ \text{da saída do modelo} \end{pmatrix} = \begin{pmatrix} \text{combinação linear} \\ \text{dos valores} \\ \text{presente e passados} \\ \text{da entrada do modelo} \end{pmatrix} \quad (50)$$

Os processos que obedecem o comportamento acima são ditos **processos lineares**.

A estrutura do filtro linear que processará os dados, é determinada pela maneira que as duas combinações lineares indicadas na Eq. (50) são formuladas. Podemos então identificar três tipos usuais de modelos estocásticos lineares:

1. **Modelo auto-regressivo (AR)** - no qual não são utilizadas valores passados da entrada.
2. **Modelo moving average (MA)** - no qual não são usados valores passados da saída. Também chamado de modelo de média móvel.
3. **Modelo ARMA** - junção dos modelos AR e MA.

Dizemos que uma série temporal $x(n), x(n-1), \ldots, x(n-M)$ representa uma realização de um processo AR de ordem M se ela satisfaz a equação diferença seguinte:

$$x(n) + a_1 x(n-1) + \cdots + a_M x(n-M) = v(n) \tag{51}$$

em que $a_1, \ldots, a_M$ são chamados parâmetros AR e $v(n)$ é um processo de ruído branco.

É mais simples enxergar o motivo de tal processo se chamar AR se escrevermos a Eq. (51) da seguinte forma:

$$x(n) = b_1 x(n-1) + b_2 x(n-2) + \cdots + b_M x(n-M) + v(n) \quad (52)$$

em que $b_k = -a_k$. Desta maneira vê-se facilmente que o instante atual do processo, ou seja $x(n)$ é igual a uma combinação dos valores passados do processo mais um termo de erro $v(n)$.

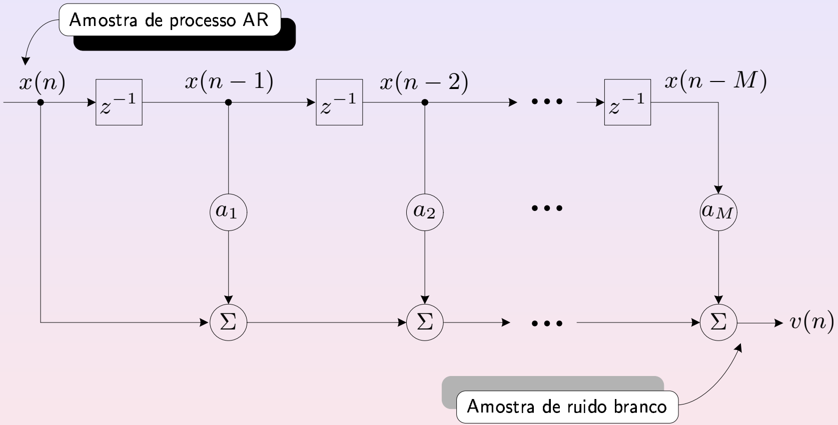


Figure: Analisador de processo AR

Modelo auto-regressivo (AR) - cont.

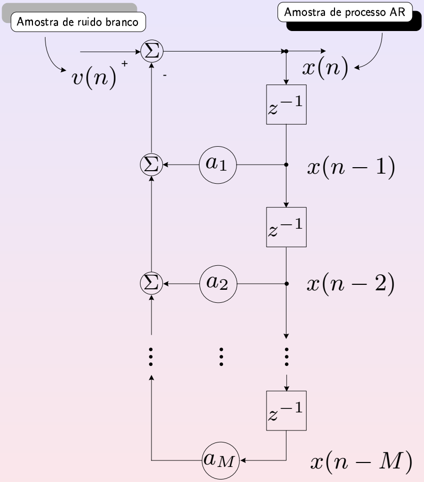


Figure: Gerador de processo AR

Em um modelo de média móvel (MA, *moving average*), o sistema é um filtro apenas com zeros e com ruído branco como entrada. O processo resultante $x(n)$ produzido é então dado pela seguinte equação diferença

$$x(n) = v(n) + b_1 v(n-1) + b_2 v(n-2) + \ldots + b_K v(n-K) \quad (53)$$

em que $b_1, \ldots, b_K$ são constantes chamadas de *parâmetros MA* e $v(n)$ é um processo de ruído branco de média zero e variância σ_v^2.

A Equação (53), representa uma versão escalar de um produto interno. Com isso, podemos representá-la como:

$$x(n) = \sum_{i=0}^{K} b_i v(n-i) = \mathbf{v}\mathbf{b}^T \tag{54}$$

em que $\mathbf{v} = [\ v(n)\ v(n-1)\ \ldots\ v(n-K)\ ]$ e $\mathbf{b} = [\ 1\ b_1\ b_2\ \ldots\ b_K\ ]$.

A ordem do processo MA é dada por K. O termo média móvel surge pois constrói-se uma estimativa do processo x a partir de uma média ponderada das amostras do processo v.

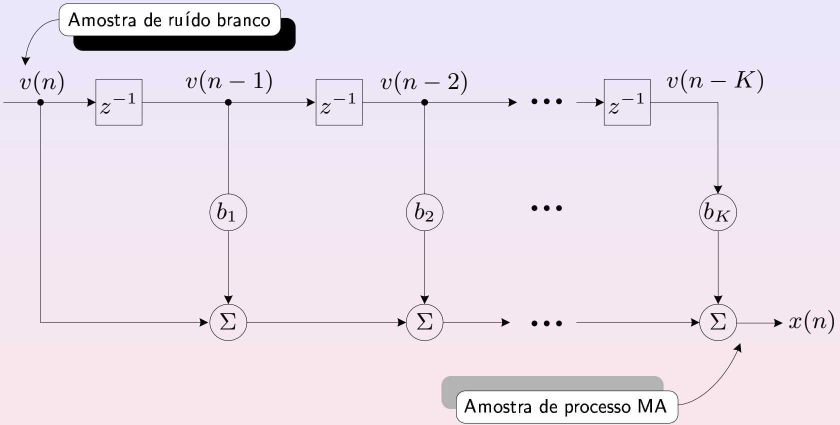


Figure: Modelo gerador de um processo de média móvel.

Para gerar um modelo **auto-regressivo-média móvel**, utilizando um processo de ruído branco como entrada, temos a seguinte equação diferença

$$x(n) + a_1 x(n-1) + \ldots + a_M x(n-M) = v(n) + b_1 v(n-1) \\ + b_2 v(n-2) + \ldots + b_K v(n-K) \tag{55}$$

em que $a_1, \ldots, a_M$ e $b_1, \ldots, b_K$ são os parâmetros ARMA.

A ordem do modelo ARMA é dada por (M, K).

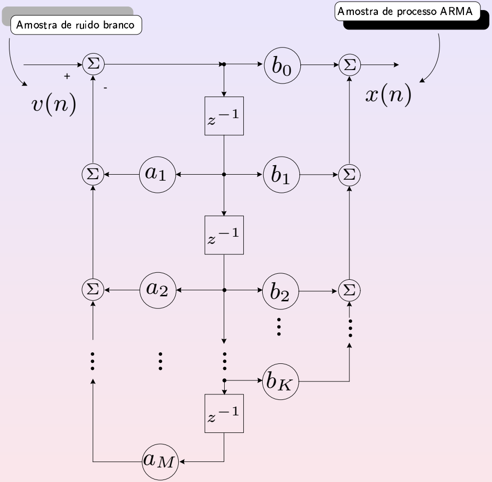
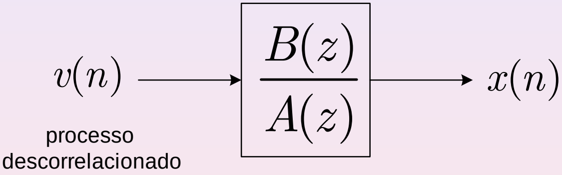


Figure: Modelo gerador de um processo ARMA de ordem (M, K), supondo $M > K$.

Modelos: ARMA, AR e MA

Filtragem linear de um processo do tipo ruído branco, também chamado de **processo inovação**



$$S_X(\omega) = \left.\frac{B(z)}{A(z)}\right|_{z=e^{j\omega}}$$

$$x(n) = \sum_{i=1}^{M} a_i x(n-i) + \sum_{j=0}^{K} b_j v(n-j) \tag{56}$$

ARMA

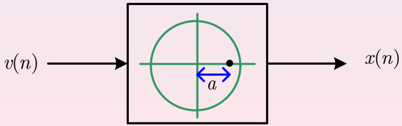
AR $\Rightarrow b_j = 0$ para $j > 0$ (autorregressivo)
MA $\Rightarrow a_i = 0$ para $i > 0$ (média móvel)

Modelo AR

$$x(n) = v(n) + \sum_{i=1}^{M} a_i x(n-i) \tag{57}$$

Os a_i ótimos são obtidos como uma solução de um sistema de equações lineares

Caso específico: AR(1): $x(n) = v(n) + ax(n-1)$



O modelo AR(1) é também um **processo markoviano** $\rightarrow$ sinal com memória 1

MAC de um processo AR(1)

$$\begin{aligned} r_X(0) &= \mathbb{E}\left\{x^2(n)\right\} = \mathbb{E}\left\{v^2(n) + 2\cdot a\cdot v(n)\cdot x(n-1) + a^2x^2(n-1)\right\} \\ &= \sigma_v^2 + 2a\overbrace{\mathbb{E}\left\{v(n)\cdot x(n-1)\right\}}^{0} + a^2\cdot r_X(0) \\ &= \sigma_v^2 + a^2\cdot r_X(0) \end{aligned} \tag{58}$$

Daí resulta

$$\boxed{r_X(0) = \frac{\sigma_v^2}{1-a^2}} \tag{59}$$

MAC de um processo AR(1) - cont.

$$
\begin{aligned}
r_X(1) &= \mathbb{E}\{x(n)x(n-1)\} \\
&= \mathbb{E}\{[v(n)+ax(n-1)][v(n-1)+ax(n-2)]\} \\
&= \mathbb{E}\left\{ax(n-1)v(n-1)+a^2x(n-1)x(n-2)\right\} \\
&= \mathbb{E}\left\{a[v(n-1)+ax(n-2)]\cdot v(n-1)+a^2x(n-1)x(n-2)\right\} \\
&= a\sigma_v^2 + a^2\cdot\mathbb{E}\{x(n-1)x(n-2)\} \\
&= a\sigma_v^2 + a^2 r_X(1)
\end{aligned}
\tag{60}
$$

Daí, temos

$$
\begin{aligned}
r_X(1)[1-a^2] &= a\sigma_v^2 \\
r_X(1) &= a\cdot r_X(0)
\end{aligned}
\tag{61}
$$

MAC de um processo AR(1) - cont.

Pode-se mostrar por recursividade que as seguintes relações são válida

$$\begin{aligned} r_X(i) &= a^i \cdot r_X(0) \\ r_X(i) &= a \cdot r_X(i-1) \end{aligned} \tag{62}$$

Assim, temos que a MAC (ordem 2) de um processo AR(1) é dada como

$$\mathbf{R_x} = \begin{bmatrix} \frac{\sigma_v^2}{(1-a^2)} & \frac{a \cdot \sigma_v^2}{(1-a^2)} \\ \frac{a \cdot \sigma_v^2}{(1-a^2)} & \frac{\sigma_v^2}{(1-a^2)} \end{bmatrix} \tag{63}$$

MAC de um processo AR(1) - cont.

Calcular λ_1 e λ_2 (autovalores)

$$
\begin{aligned}
\det(\mathbf{R_x} - \lambda \mathbf{I}) &= 0 \\
\left(\frac{\sigma_v^2}{1-a^2} - \lambda\right)^2 - \left(\frac{a\cdot\sigma_v^2}{1-a^2}\right)^2 &= \frac{\sigma_v^4}{(1-a^2)^2} - 2\frac{\lambda\sigma_v^2}{(1-a^2)} + \lambda^2 \\
-\frac{a^2\sigma_v^4}{(1-a^2)^2} &= \frac{\sigma_v^4}{(1-a^2)^2}\cdot(1-a^2) - 2\lambda\frac{\sigma_v^2}{(1-a^2)} + \lambda^2 \\
&= \lambda^2 - 2\lambda\cdot\frac{\sigma_v^2}{(1-a^2)} + \frac{\sigma_v^4}{(1-a^2)} = 0 \\
&= (1-a^2)\cdot\lambda^2 - 2\sigma_v^2\lambda + \sigma_v^4 = 0
\end{aligned} \tag{64}
$$

MAC de um processo AR(1) - cont.

Calcular λ_1 e λ_2 (autovalores) - cont.

$$\begin{aligned}\lambda_{1,2} &= \frac{2\sigma_v^2 \pm \sqrt{4\sigma_v^4 - 4\sigma_v^4(1-a^2)}}{2\cdot(1-a^2)} \\ &= \frac{2\sigma_v^2 \pm 2a\sigma_v^2}{2(1-a^2)}\end{aligned}$$

Então, teremos

$$\begin{aligned}\lambda_1 &= \frac{\sigma_v^2}{(1-a)} \\ \lambda_2 &= \frac{\sigma_v^2}{(1+a)}\end{aligned} \tag{65}$$

MAC de um processo AR(1) - cont.

Assim, a MAC de um processo AR(1) tem como autovalores

$$\lambda_1 = \frac{\sigma_v^2}{(1-a)} \qquad \text{e} \qquad \lambda_2 = \frac{\sigma_v^2}{(1+a)} \tag{66}$$

E então, podemos definir o **número de condicionamento** ou **espalhamento** da matriz, dado por

$$\boxed{\mathcal{C}\left(\mathbf{R_x}\right) = \frac{\lambda_{\max}}{\lambda_{\min}}} \tag{67}$$

No caso do processo AR(1): $\mathcal{C}\left(\mathbf{R_x}\right) = \frac{1+a}{1-a}$

MAC de um processo AR(1) - cont.

Informações fornecidas por $\mathcal{C}\left(\mathbf{R_x}\right)$

- O menor autovalor fica confinado entre σ_v^2 e $\frac{\sigma_v^2}{2}$
- O maior autovalor vai de σ_v^2 a infinito
- O número $\mathcal{C}\left(\mathbf{R_x}\right) = \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}}$ exprime quão plano é $S_X(\omega)$ ou quão descorrelacionado é $\mathbf{x}(n)$, ou ainda quão "próximo" $\mathbf{R_x}$ está de uma matriz diagonal
  1. $\mathcal{C}\left(\mathbf{R_x}\right) = 1 \Rightarrow S_X(\omega)$ plano
  2. $\mathcal{C}\left(\mathbf{R_x}\right) = 1 \Rightarrow x(n)$ é branco
  3. $\mathcal{C}\left(\mathbf{R_x}\right) \to \infty \Rightarrow S_X(\omega)$ é "colorido"
  4. $\mathcal{C}\left(\mathbf{R_x}\right) \to \infty \Rightarrow x(n)$ tem correlação