IE009 - Processamento adaptativo de sinais

Uma das principais finalidades das variadas estratégias de filtragem adaptativa é tentar lidar com as incertezas presentes nos sinais e sistemas práticos. Por isso, é necessário lançar mão da teoria da probabilidade para definir adequadamente a informação que está sendo processada.

Teoria da Probabilidade

- experimento aleatório
  - possível de ser repetido sob condições idênticas
  - o resultado do experimento em qualquer realização não pode ser pré-determinado antes de sua ocorrência.
  - deve haver uma regularidade estatística quando um grande número de realizações do experimento é executado, no sentido que um comportamento médio pode ser observado

Esta regularidade estatística dá origem à noção de freqüência relativa.

- Seja Ω o conjunto de todos os possíveis resultados $\omega_i, i=1,\ldots N_\omega$ de um experimento aleatório. Um evento A é $\forall$ subconjunto de Ω (elemento, subconjunto ou conjunto de subconjuntos). Se N é o número de realizações do experimento e N_A é o número de ocorrências do evento A, então a freqüência relativa é $f = \frac{N_A}{N}$

Neste caso, define-se a probabilidade como $P\{A\} = \lim_{N \to \infty} N_A/N$.

- Kolmogorov

"A teoria da probabilidade, sendo uma disciplina da matemática, pode e deve ser desenvolvida com base em axiomas, tal como a algebra e a geometria".

Definição axiomática de probabilidade:

– seja um campo de probabilidade formado pela tripla $\{F, \Omega, P\{A\}\}$ onde Ω é o conjunto de todos os resultados possíveis (espaço amostral), F é um campo que contém todos os eventos possíveis e $P\{A\}$ é a probabilidade do evento A. Esta medida, $P\{A\}$, é um número que satisfaz:

1) $P\{A\} \geq 0$
2) $P\{\Omega\} = 1$
3) Se $A \cap B = \emptyset$ (eventos disjuntos), então $P(A \cup B) = P(A) + P(B)$

Alguns resultados importantes:

- $P\{\bar{A}\} = 1 - P\{A\}$, onde $\bar{A}$ é o complemento do conjunto A.
- $P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\}$
- Independência: $P\{A \cap B\} = P\{A\} \cdot P\{B\}$
- Prob. condicional: $P\{A | B\} \triangleq \dfrac{P\{A \cap B\}}{P\{B\}}$

- Lei da probabilidade total: $P\{B\} = \sum_i P\{B \cap A_i\} = \sum_i P\{B | A_i\} \cdot P\{A_i\}$,

onde $\{A_i\}$ representa um conjunto contável de eventos mutuamente exclusivos ($A_i \cap A_j \neq \emptyset$ para $i \neq j$) e exaustivo ($\cup_i A_i = \Omega$).

* Variável Aleatória

É uma função $\Omega \to \mathbb{R}$ que mapeia cada possível resultado ω_i em um valor $X(\omega_i)$. Por exemplo, no caso do lançamento de uma moeda, $X(\text{CARA}) = +1$ e $X(\text{COROA}) = -1$.

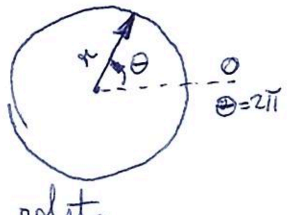
Duas classes: VA discreta – pode assumir um nº finito de valores (dado)

VA contínua – pode assumir um nº infinito (não contável) de valores (roleta)

Uma V.A. pode ser caracterizada por sua <u>função distribuição de probabilidade</u>, $F_X(x) = P\{X \leq x\}$ ou, equivalentemente, pela <u>função densidade de probabilidade</u> (p.d.f), $f_X(x) = \partial F_X(x) / \partial x$

<u>Exemplo:</u>

<u>VA contínua</u>



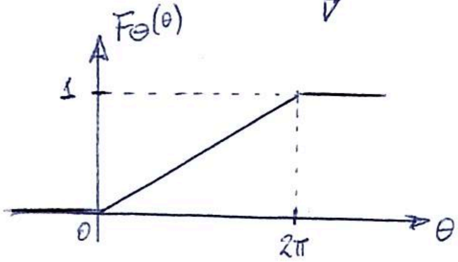
roleta

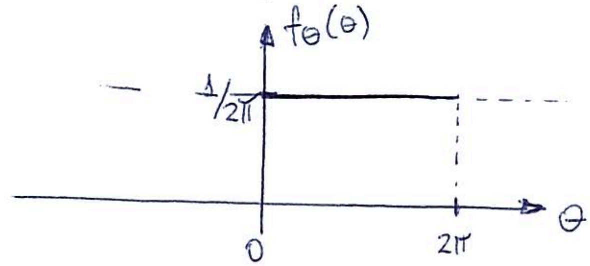
$F_\Theta(\theta) = P\{\Theta \leq \theta\}$

$= \dfrac{r \cdot \theta}{2\pi r} = \dfrac{\theta}{2\pi}$

Derivando em relação a θ:

$f_\Theta(\theta) = 1/2\pi$





<u>Propriedades:</u> * Monotonicamente crescente

* $F_X(-\infty) = 0$, $F_X(\infty) = 1$

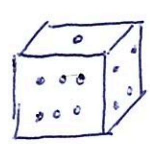
* Sempre não-negativa

* $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$

$P\{x_1 < X \leq x_2\} = P\{X \leq x_2\} - P\{X \leq x_1\} =$

$= F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\, dx$

<u>Exemplo:</u>

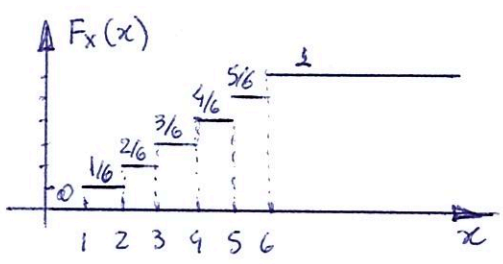


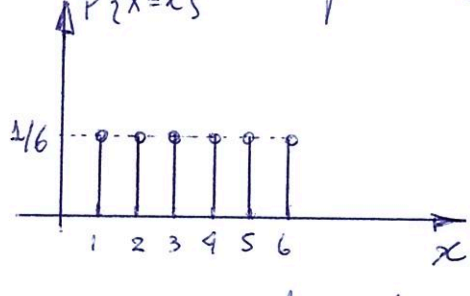
$x \in \{1, 2, 3, 4, 5, 6\}$

<u>VA discreta</u>

Dado

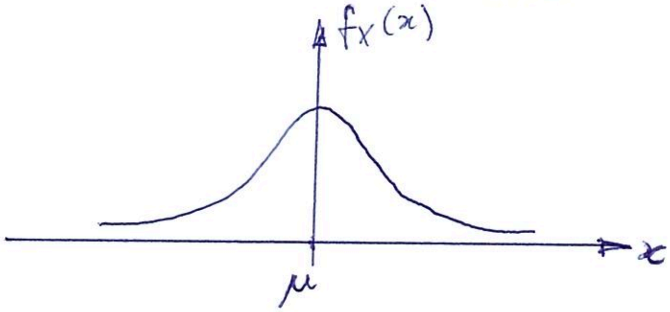
A PMF (probability mass function) define $P\{X = x\}$





Se pensássemos em pdf, teríamos impulsos nos valores de x em que

• V.A. gaussiana: $f_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$ → totalmente definida pela média (μ) e pela variância (σ^2)



Qualquer transformação linear feita sobre uma VA gaussiana dá origem a uma outra VA gaussiana também.

• Quando temos duas ou mais VAs, falamos em distribuição e densidade conjuntos: $F_{X_1,X_2}(x_1,x_2) = P\{X_1 \leq x_1, X_2 \leq x_2\}$

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

Se X_1 e X_2 forem independentes,

$$F_{X_1,X_2}(x_1,x_2) = F_{X_1}(x_1) \cdot F_{X_2}(x_2)$$

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$$

• Distribuição e densidade condicional:

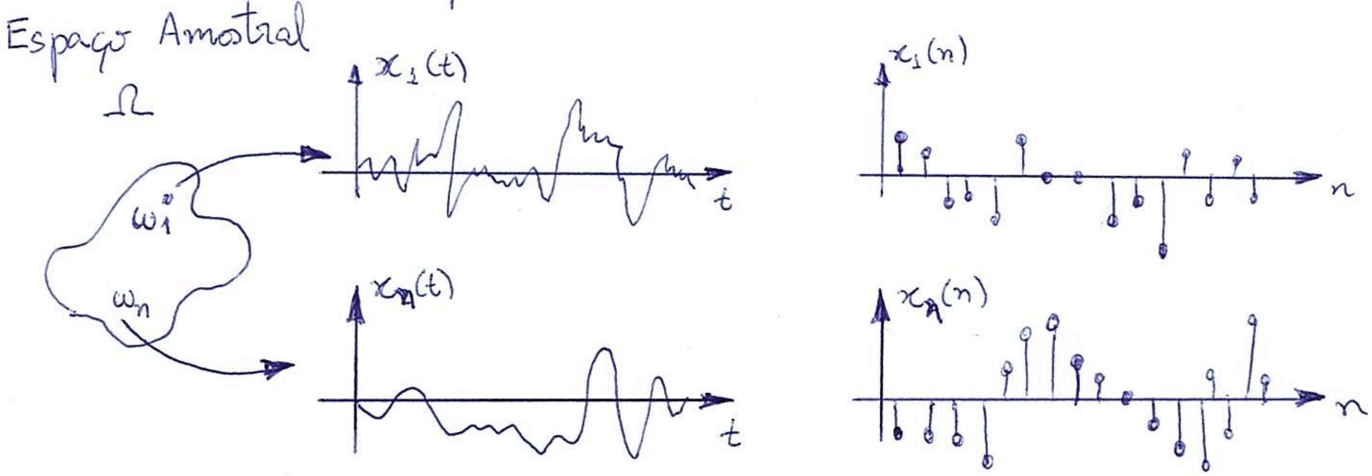
$$F_{X_1|X_2}(x_1|x_2) = \frac{F_{X_1,X_2}(x_1,x_2)}{F_{X_2}(x_2)} \xRightarrow{\text{Teorema de Bayes}} F_{X_1|X_2}(x_1|x_2) = \frac{F_{X_1}(x_1)}{F_{X_2}(x_2)} \cdot F_{X_2|X_1}(x_2|x_1)$$

• Esperança ou média: $E\{X\} = \int_{-\infty}^{\infty} x f_X(x)\, dx$

• Variância: $V\{X\} = E\{(X - E\{X\})^2\} = \int_{-\infty}^{\infty} (x - E\{X\})^2 f_X(x)\, dx$

* Processos Estocásticos

Um processo aleatório $X(t)$ (ou $X(t,\omega_i)$) é a coleção de funções geradas por uma regra que associa uma função $x_i(t)$ a cada possível resultado ω_i de um experimento aleatório, $\omega_i \in \mathbb{R}$.

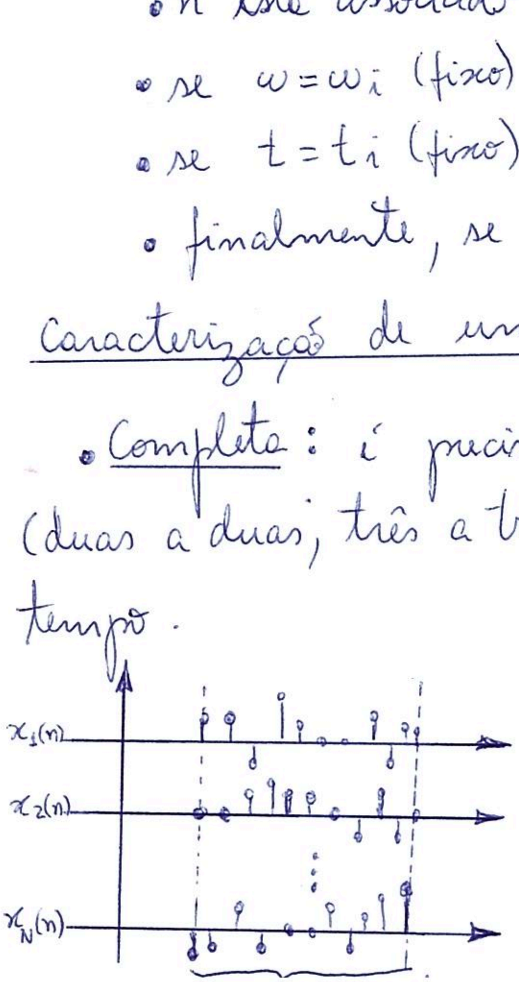


No caso discreto, usamos a notação $X(n)$ (ou $X(n,\omega_i)$)

- ω está associado ao comportamento estatístico do processo
- n está associado ao comportamento temporal do processo
- se $\omega = \omega_i$ (fixo) $\longrightarrow x_i(t)$ = função amostra do processo
- se $t = t_i$ (fixo) $\longrightarrow X(t_i,\omega)$ = variável aleatória
- finalmente, se $\omega = \omega_i$ e $t = t_i$, $X(t_i,\omega_i)$ é apenas um número.

Caracterizações de um processo aleatório

- Completa: é preciso conhecer as pdf.s marginais e conjuntas (duas a duas, três a três, etc) dos V.A.s em todos os instantes de tempo.



K variáveis aleatórias : $X_1, X_2, \ldots X_K$

$F_{X_1,X_2,\ldots,X_K}(x_1,x_2,x_3,\ldots,x_K)$ ou $F_{\underline{X}}(\underline{x})$, onde

$\underline{X} = [X_1, X_2, \ldots X_K]^T$ é um vetor aleatório

A distribuição (ou pdf) conjunta que caracteriza o processo é o máximo de informação que podemos ter a seu respeito.

Entretanto, é bastante difícil e/ou raro ter esse tipo de conhecimento a respeito do processo estocástico.

• <u>Parcial</u>: baseia-se em medidas estatísticas até 2ª ordem.

- <u>média</u>: $E\{\mathbf{X}(n)\} = \int_{-\infty}^{\infty} x(n)\, f_X(x)\, dx$ → valor médio ou nível DC

- <u>variância</u>: $E\{[\mathbf{X}(n) - E\{\mathbf{X}(n)\}]^2\}$ → espalhamento ou dispersão em torno da média

- <u>autocorrelação</u>: $R_X(n_1, n_2) = E\{X(n_1)X(n_2)\}$

⇒ <u>Estacionariedade</u>: dizemos que um processo estocástico $\mathbf{X}(n)$ é estritamente estacionário se o seu comportamento estatístico, representado por $f_{X(n_1), X(n_2), \ldots X(n_K)}(x_1, \ldots, x_K)$ não varia com o tempo (ou com a janela de observações):

$$f_{X(n_1), X(n_2), \ldots, X(n_K)}(x_1, x_2, \cdots, x_K) = f_{X(n_1+m), \ldots, X(n_K+m)}(x_1, \ldots, x_K)$$

Se a igualdade é verdadeira para $\forall m$, o processo é estacionário no sentido estrito.

- <u>$K=1$</u>: $f_{X(n)}(x) = f_{X(n+m)}(x) = f_X(x)$ → para todos n e m.

- <u>$K=2$</u>: $f_{X(n_1), X(n_2)}(x_1, x_2) = f_{X(0), X(n_2-n_1)}(x_1, x_2)$ → para todo n_1 e n_2

<u>e $m = -n_1$</u>

A pdf. conjunta de segunda ordem de um processo estacionário depende somente da diferença de tempo entre as amostras, e não dos instantes particulares de observação.

Consequências: — média constante: $E\{X(n)\} = m_X$

— autocorrelação depende apenas da diferença de tempo: $R_X(m) = E\{X(n)X(n-m)\}$

Opção menos restritiva: Estacionariedade no sentido amplo ou WSS (Wide-Sense Stationary)

• neste caso, as condições impostas são (1) média constante e (2) função de autocorrelação que não depende dos instantes das observações, mas apenas do intervalo de tempo entre elas. Essas duas características são observadas também em processos estacionários no sentido estrito, conforme vimos anteriormente. Ou seja, estacionariedade implica WSS mas o contrário não é necessariamente verdadeiro.

A caracterização de um processo aleatório, mesmo no contexto mais simples (WSS), requer o cálculo de esperanças estatísticas levando em consideração todos os possíveis resultados ω_i do espaço amostral subjacente. Na prática, porém, temos acesso a um nº limitado de funções amostra $x_i(t)$. Seria interessante poder, então, substituir médias estatísticas por médias temporais.

<u>Ergodicidade</u>: as médias temporais "equivalem" às médias estatísticas do processo.

- média temporal: $\underbrace{K_1(x,T)}_{\downarrow} = \frac{1}{2T}\int_{-T}^{T} x(t)\,dt$ → média estatística: $E\{X(t)\} = k_1(X)$

função de variável aleatória → é uma V.A.

Um processo $X(t)$ é <u>ergódico na média</u> se:

1) $\lim_{T\to\infty} k_1(x,T) = k_1(X)$

2) $\lim_{T\to\infty} \mathrm{Var}\{k_1(x,T)\} = 0$

- Autocorrelação: $R_x(\tau,T) = \frac{1}{2T}\int_{-T}^{T} x(t+\tau)\,x(t)\,dt$ → estimativa temporal

$X(t)$ é <u>ergódigo em relação à função de autocorrelação</u> se:

1) $\lim_{T\to\infty} R_x(\tau,T) = R_X(\tau)$

2) $\lim_{T\to\infty} \mathrm{Var}\{R_x(\tau,T)\} = 0$

<u>Cicloestacionariedade</u>: as propriedades estatísticas do processo variam com o tempo, mas são periódicas com período T.

$$f_{X(n_1),X(n_2),\ldots,X(n_k)}(x_1,x_2,\ldots x_k) = f_{X(n_1+T),X(n_2+T),\ldots,X(n_k+T)}(x_1,x_2,\ldots,x_k)$$

⇒ Autocorrelação e Densidade Espectral de Potência

• a função de ~~au~~autocorrelação é uma estatística de 2ª ordem que exprime uma similaridade entre duas variáveis aleatórias, $X(n_1)$ e $X(n_2)$, i. e., duas amostras temporais de um processo separadas por um intervalo de tempo m

$R_X(m) = E\{X(n) X^*(n-m)\}$, onde $X(n)$ é um processo estacionário e discreto no tempo.

Para um processo estacionário, $R_X(m)$ é uma função determinística. O tipo de informação presente nela pode ser avaliado também no domínio da freqüência por meio da transformada de Fourier de $R_X(m)$.

A densidade espectral de potência de $X(n)$ é definida como a transformada de Fourier de $R_X(m)$:

$$S_X(e^{j\omega}) = \mathcal{F}\{R_X(m)\} = \sum_{m=-\infty}^{\infty} R_X(m) e^{-j\omega m} \quad , \quad \omega \in [-\pi, \pi]$$

Analogamente:

$$R_X(m) = 1/2\pi \int_{-\pi}^{\pi} S_X(e^{j\omega}) e^{j\omega m} d\omega \quad \text{para } m \in \mathbb{Z}$$

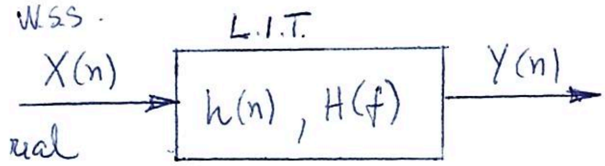
<u>Autocovariância</u>: $C_X(m) = E\{[X(n) - m_x][X(n-m) - m_x]\}$, $m_x = E\{X(n)\}$

• quando $R_X(m) = 0$, temos a condição de <u>ortogonalidade</u>; quando $C_X(m) = 0$, falamos de <u>descorrelação</u>.

• σ_X^2 = variância de $X(n) = C_X(0)$.

• $R_X(0) = E\{X^2(n)\}$ = valor quadrático médio de $X(n)$.

* Filtragem de um processo estocástico WSS



A saída $Y(n)$ do sistema linear e invariante com o tempo também será um processo estocástico. É possível relacionar as propriedades estatísticas de $Y(n)$ com as de $X(n)$.

- <u>MÉDIA</u>: $E\{Y(n)\} = E\left\{\sum_{k=-\infty}^{\infty} h(k)\, x(n-k)\right\}$

$$= \sum_{k=-\infty}^{\infty} h(k)\, E\{X(n-k)\}$$

$$m_Y = \sum_{k=-\infty}^{\infty} h(k) \,.\, m_X = \underline{H(0).m_X}\,,$$

onde m_X é a média do processo de entrada $X(n)$ e

$H(f) = \sum_{n=-\infty}^{\infty} h(n)\, e^{-j2\pi f n}$ e $H(0) = H(f)\Big|_{f=0}$.

- <u>AUTOCORRELAÇÃO</u>:

$$E\{Y(n)Y(n+k)\} = E\left\{\sum_{i=-\infty}^{\infty} h(i)x(n-i) \sum_{j=-\infty}^{\infty} h(j)\, x(n+k-j)\right\}$$

$$= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} h(i)h(j).\underbrace{E\{x(n-i)\,x(n+k-j)\}}_{R_X(k-j+i)}$$

Logo, podemos escrever:

$$E\{Y(n)\,Y(n+k)\} = \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} h(i)h(j)\, R_X([k+i]-j)$$

$$= \sum_{i=-\infty}^{\infty} h(i) \sum_{j=-\infty}^{\infty} h(j)\, R_X([k+i]-j)$$

Seja $g(n) = h(n) * R_X(n) = \sum_{m=-\infty}^{\infty} h(m) R_X(n-m)$. Então:

$$E\{Y(n)Y(n+k)\} = \sum_{i=-\infty}^{\infty} h(i) \sum_{j=-\infty}^{\infty} h(j) R_X([k+i]-j)$$

$$R_Y(-k) = \sum_{i=-\infty}^{\infty} h(i)\, g(k+i)$$

Trocando o índice do somatório por $l = -i$:

$$R_Y(-k) = \sum_{l=-\infty}^{\infty} h(-l)\, g(k-l)$$

$$R_Y(-k) = h(-k) * g(k)$$

Como $g(k) = h(k) * R_X(k)$, chegamos a $\boxed{R_Y(-k) = h(-k) * h(k) * R_X(k)}$

$Y(n)$ real $\rightarrow R_Y(k) = R_Y(-k)$

- DENSIDADE ESPECTRAL DE POTÊNCIA:

$$S_Y(f) = \mathcal{F}\{R_Y(k)\} = \mathcal{F}\{h(-k) * h(k) * R_X(k)\}$$

Como $\mathcal{F}\{h(k)\} = H(f)$, $\mathcal{F}\{h(-k)\} = H^*(f)$ e $\mathcal{F}\{R_X(k)\} = S_X(f)$, temos:

$$S_Y(f) = H(f)H^*(f)S_X(f) \Rightarrow \boxed{S_Y(f) = |H(f)|^2 S_X(f)}$$

OBSERVAÇÃO: Se $X(n)$ for um processo aleatório gaussiano, $Y(n)$ também será gaussiano.

⇒ MATRIZ DE AUTOCORRELAÇÃO

Considerando M amostras de um processo $X(n)$, podemos construir o vetor $\underline{X}(n) = [X(n)\ X(n-1)\ \ldots\ X(n-M+1)]^T \rightarrow$ M variáveis aleatórias

$$R_X = E\{\underline{X}(n)\underline{X}^H(n)\}$$

$$= E\left\{ \begin{bmatrix} X(n) \\ \vdots \\ X(n-M+1) \end{bmatrix} \cdot \begin{bmatrix} X^*(n) & \ldots & X^*(n-M+1) \end{bmatrix} \right\}$$

$$R_X = \begin{bmatrix} E\{x(n)x^*(n)\} = R_X(0) & E\{x(n)x^*(n-1)\} = R_X(+1) & \cdots & R_X(M-1) \\ E\{x(n-1)x^*(n)\} = R_X(-1) & E\{x(n-1)x^*(n-1)\} = R_X(0) & \cdots & \vdots \\ \vdots & \vdots & & \\ R_X(-(M-1)) & \cdots & & R_X(0) \end{bmatrix}$$

Se $X(n)$ é complexo, R_X é hermitiana pois $R_X = R_X^H$.
Se $X(n)$ é real, R_X é simétrica pois $R_X(i) = R_X(-i)$.

Ruído Branco:

É um processo aleatório WSS composto por uma sequência de variáveis aleatórias descorrelacionadas, todas com a mesma variância.

- Ruído branco gaussiano (WGN)

Sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com p.d.f. marginal gaussiana

$X(n) \sim N(0, \sigma^2)$ para $-\infty < n < \infty$

Por ser i.i.d., o processo é estacionário

$$f_{X(n_1),X(n_2),\dots,X(n_N)}(x_1,x_2,\dots,x_N) \overset{\text{indep.}}{=} f_{X(n_1)}(x_1)\, f_{X(n_2)}(x_2) \cdots f_{X(n_N)}(x_N)$$

$$= \prod_{i=1}^{N} f_{X(n_i)}(x_i)$$

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

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \cdot \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N} x_i^2\right)$$

A média do WGN é nula em todos os instantes de observação.

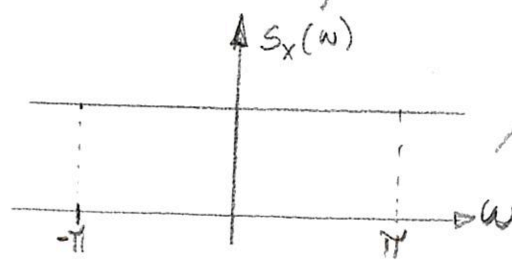
Autocorrelação:

$$R_X(k) = E\{X(n)\,X(n-k)\} \overset{IND}{=} E\{X(n)\}\cdot E\{X(n-k)\} = \begin{cases} E\{X^2(n)\} = \sigma^2, & k=0 \\ 0, & k\neq 0 \end{cases}$$

$$\therefore \boxed{R_X(k) = \sigma^2 \cdot \delta(k)}$$

PSD:

$$\boxed{S_X(\omega) = \mathcal{F}\{R_X(k)\} = \sigma^2}, \quad -\pi < \omega < \pi$$



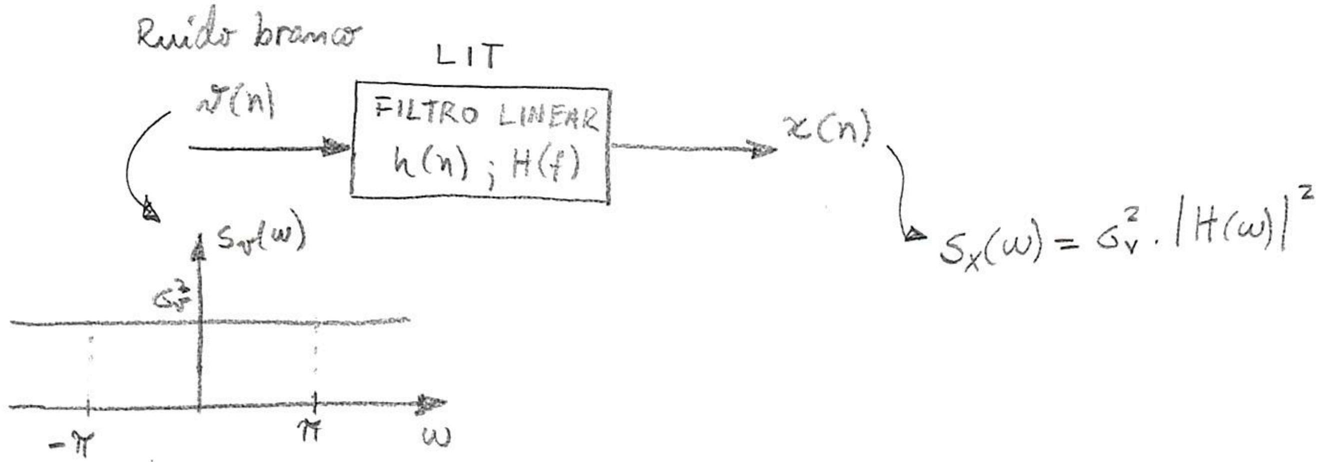
espectro plano

Matriz de autocorrelação:

$$\underline{X}(n) = [\,X(n) \quad X(n-1) \quad \ldots \quad X(n-M)\,]^T$$

$$R_X = E\{\underline{X}(n)\underline{X}^H(n)\} = \sigma^2 \cdot \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \sigma^2 \cdot I_M$$

Processos obtidos a partir da filtragem de um ruído branco



a) Processo de média móvel (MA)

$$x(n) = a_0 v(n) + a_1 v(n-1) + \dots + a_{M-1} v(n-M+1) = \sum_{i=0}^{M-1} a_i v(n-i)$$

$$X(z) = H(z)\,V(z) \Rightarrow H(z) = a_0 + a_1 z^{-1} + \dots + a_{M-1} z^{-(M-1)}$$

↳ filtro FIR (somente zeros)

b) Processo autoregressivo (AR)

$$x(n) = \underbrace{b_1 x(n-1) + b_2 x(n-2) + \dots + b_Q x(n-Q)}_{} + v(n)$$

autoregressão de $x(n)$

$$H(z) = \frac{1}{(1 - b_1 z^{-1} - b_2 z^{-2} - \dots - b_Q z^{-Q}} \Rightarrow \text{Filtro IIR (somente pólos)}$$

c) Processo autorregressivo com média móvel (ARMA)

$$x(n) = a_0 v(n) + \dots + a_M v(n-M) + b_1 x(n-1) + \dots + b_Q x(n-Q)$$

$$H(z) = \frac{a_0 + a_1 z^{-1} + \dots + a_M z^{-M}}{1 - b_1 z^{-1} - \dots - b_Q z^{-Q}} \Rightarrow \text{pólos e zeros}$$

Processo AR

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

$$x(n) - \sum_{i=1}^{K} b_i x(n-i) = v(n)$$

Multiplicando por $x^*(n-\ell)$ e calculando a esperança dos dois lados da equação, temos:

$$\underbrace{E\{x(n)x^*(n-\ell)\}}_{R_x(\ell)} - \sum_{i=1}^{K} b_i \underbrace{E\{x(n-i)x^*(n-\ell)\}}_{R_x(\ell-i)} = E\{v(n)x^*(n-\ell)\}$$

Note que $E\{v(n)x^*(n-\ell)\} = 0$ quando $\ell > 0$ pois, nesse caso, $v(n)$ e $x^*(n-\ell)$ são independentes e, como $E\{v(n)\}=0$, $E\{v(n)\}E\{x^*(n-\ell)\} = 0$. Assim,

$$R_x(\ell) = \sum_{i=1}^{K} b_i R_x(\ell - i) \quad , \ \ell > 0$$

A partir desse resultado, podemos construir o seguinte sistema de equações:

$$\underbrace{\begin{bmatrix} R_x(0) & R_x(1) & \dots & R_x(K-1) \\ R_x(1) & R_x(0) & \dots & R_x(K-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_x(K-1) & R_x(K-2) & & R_x(0) \end{bmatrix}}_{\substack{\text{Matriz de autocorrelação de } x(n) \\ R_x}} \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix}}_{\underline{b}} = \underbrace{\begin{bmatrix} R_x(1) \\ R_x(2) \\ \vdots \\ R_x(M) \end{bmatrix}}_{\underline{r}_x} \left.\begin{matrix} \to \ell=1 \\ \to \ell=2 \\ \vdots \\ \to \ell=M \end{matrix}\right\} \text{Equações de Yule-Walker}$$

$R_x \underline{a} = \underline{r}_x \Rightarrow \boxed{\underline{a} = R_x^{-1} \underline{r}_x} \rightarrow$ coeficientes do filtro que gera um processo com autocorrelação $R_x(n)$ a partir de um ruído branco.