Processos Estocásticos

Prof. Charles Casimiro Cavalcante

charles@gtel.ufc.br

Departamento de Engenharia de Teleinformática
Universidade Federal do Ceará – UFC
http://www.gtel.ufc.br/charles

UNIVERSIDADE FEDERAL DO CEARÁ

"*Tudo que existe no universo é fruto do acaso.*"

Demócrito (pensador grego, 460 a.C.-370 a.C.)

Conteúdo do curso

1. Caracterização de Processos Estocásticos
2. Processos Estacionários e Ergódicos
3. Cadeias de Markov
4. Análise Espectral de Processos Estocásticos

Caracterização de Processos Aleatórios

O que são sinais e sistemas?

- **Sinal:** um ente matemático que usamos para representar algum tipo de informação que queremos tornar disponível, ou seja, é uma representação matemática de algum processo físico;
- **Sistema:** modelo matemático para um determinado ambiente ou conjunto de operações que são aplicadas à informação.

Desta forma, sinais e sistemas são representações de informações com as quais necessitamos trabalhar.

Classificação de sinais

Podemos dividir as classes de sinais de acordo com várias de suas características.

Em relação à natureza do tipo de informação a classificação usual é:

- Sinais determinísticos;
- Sinais **aleatórios**.

Sinais determinísticos

No caso determinístico, um sinal é completamente definido a partir de uma **função geradora e da entrada** da mesma.

Com isso, pode-se concluir que sinais determinísticos são *previsíveis*.

Se tomarmos em relação à representação temporal podemos ter:

- **Contínuos**: definido para todos os instantes temporais - $y(t)$;
- **Discretos**: definido para alguns instantes de tempo - $y(n)$.

Sinal determinístico contínuo

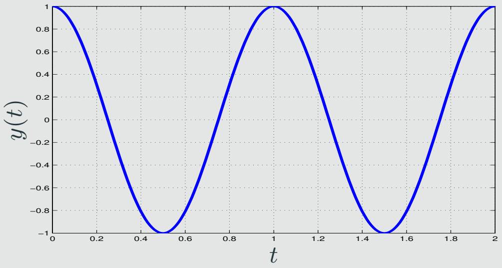


Figura 1: Sinal determinístico contínuo $y(t) = \cos(2\pi t)$.

Teste: Qual o valor de $y(t)$ para $t = 1.25$?

Sinal determinístico discreto

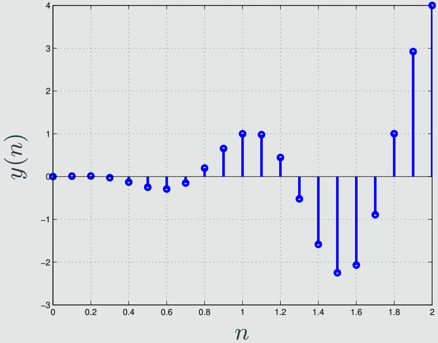


Figura 2: Sinal determinístico discreto $y(n) = \cos(2\pi n) \cdot (n)^2$.

Determinismo é suficiente?

Nosso interesse? Trabalhar com **sinais aleatórios**, os quais não possuem uma natureza tão simplificada

Porque? Modelos matemáticos que usualmente trabalhamos não são capazes “traduzir” o comportamento dos sistemas e sinais reais que encontramos na prática.

Como? Uso de modelos probabilísticos para incluir nos modelos algumas componentes que sejam um reflexo (espera-se que bastante fiel!) do que ocorre no mundo real.

Mas como modelar a aleatoriedade?

Sinais aleatórios - tipos de incerteza

Sinais e sistemas aleatórios tratam de *incertezas*. Entretanto, a incerteza pode ser classificada de duas maneiras:

1. **Probabilística:** "incertezas objetivas" podem ser mensuradas;
2. **Lógica nebulosa (fuzzy):** "incertezas subjetivas" mais difíceis de quantizar.

No nosso curso trataremos apenas com as incertezas mensuráveis, ou probabilísticas.

Sinais aleatórios - exemplo

Experimento: medir a tensão elétrica de corrente alternada na rede elétrica doméstica.

? Qual o valor que você espera ter?

- Instrumentos de medição apresentam variação nas características
- *Confiabilidade* da medida diretamente relacionada ao aparelho
- Processo de transmissão de energia: "flutuação" do valor da tensão

Provável: não ser precisamente uma senóide com 220 V.

Sinais aleatórios - exemplo (cont.)

Uma representação possível do sinal de energia transmitido pela rede seria

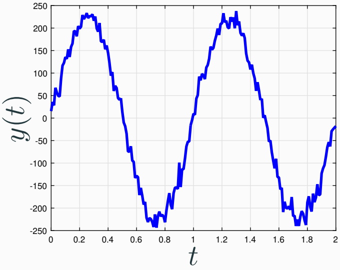
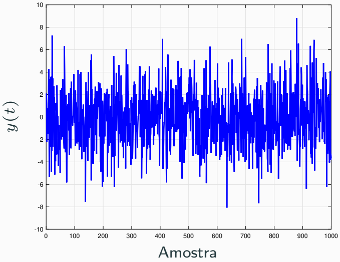


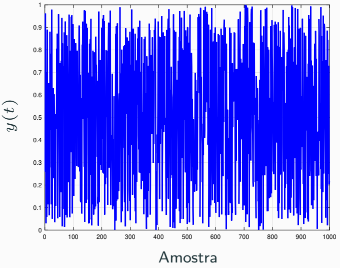
Figura 3: Representação da energia na rede elétrica.

Pergunta: Que sinal é esse?

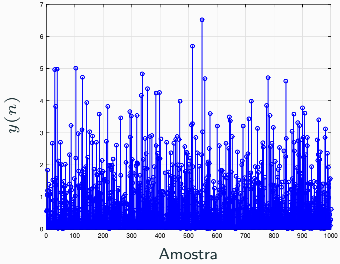
Resposta: Uma senóide + **"incerteza"**. Incerteza gerada a partir de diversas fontes!



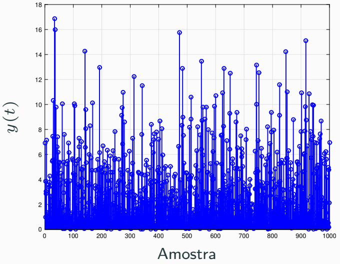
$y(t)$
Amostra
10
8
6
4
2
0
-2
-4
-6
-8
-10
0
100
200
300
400
500
600
700
800
900
1000



$y(t)$
Amostra
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
100
200
300
400
500
600
700
800
900
1000



$y(n)$
Amostra
7
6
5
4
3
2
1
0
0
100
200
300
400
500
600
700
800
900
1000



$y(t)$
Amostra
18
16
14
12
10
8
6
4
2
0
0
100
200
300
400
500
600
700
800
900
1000

Exemplos de sinais aleatórios

- Sinal de voz/fala;
- Sinal de eletrocardiograma (ECG);
- Sinal de eletroencefalograma (EEG);
- Sinal de comunicação digital;
- Música produzida por uma fita cassete ou outro dispositivo de reprodução;
- Velocidade de rotação de um motor de indução;
- Onda de tensão produzido por um gerador elétrico;
- Onda de tensão nos TAPs do secundário de um transformador;
- Sinal gerado por um instrumento musical (e.g. flauta, violino).

Caracterização das incertezas

- Probabilidade
- Variáveis aleatórias
- Distribuições de probabilidade
- Distribuição cumulativa
- ...

Estatística para Engenharia

Processos estocásticos

Definição

Um **processo estocástico** é uma coleção de variáveis aleatórias que são indexadas pelo *tempo*. Neste caso podemos ter a representação em tempo contínuo ou discreto. Neste caso precisamos de uma expressão que contenha a incerteza do processo, por exemplo,

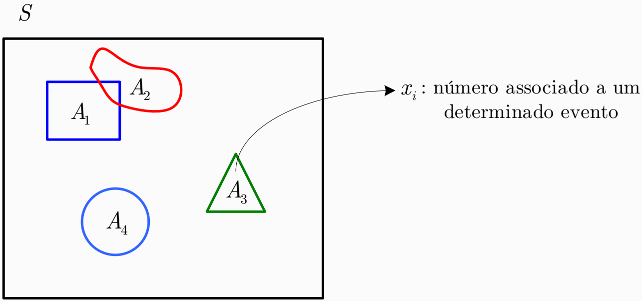
$$X(t) = A \cdot \cos(2\pi f_0 t), \quad A \sim \mathcal{N}(\mu, \sigma^2) \tag{1}$$

Por vezes, podemos nomear de forma mais genérica como *processos aleatórios*.

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

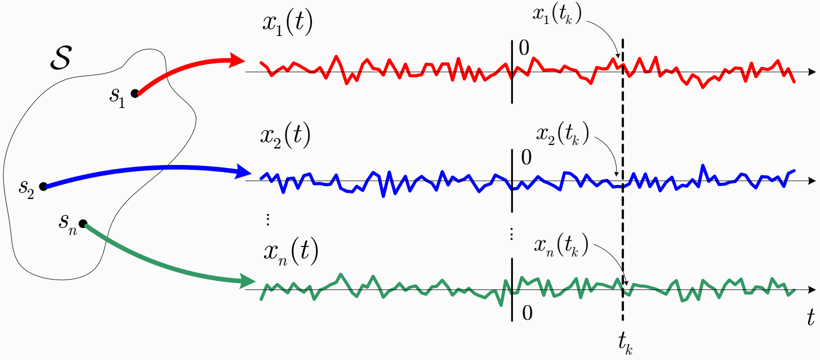


Figura 4: Representação de um processo estocástico.

Processo estocástico - notação

- Uma realização de um processo aleatório é apenas uma das curvas da Figura 4.
- O conjunto de várias realizações é geralmente denominado de ensemble (conjunto) do processo estocástico.
- Tipicamente, assumimos que são possíveis *infinitas* realizações de um processo estocástico.

Processo estocástico - aleatoriedade

Uma decorrência do processo de construção de um processo estocástico é que, em um *dado instante de tempo*, se torna uma **variável aleatória**.

Desta forma, podemos usar os conceitos de probabilidade para caracterizar a incerteza do processo, para cada instante.

Caracterização de processos aleatórios

Como caracterizar $x(n_0)$?

① função de densidade de probabilidade (pdf): $p_{X_0}(x_0)$

② função de distribuição (cumulativa) de probabilidade (cdf): $F_{X_0}(x_0)$

Classificação de processos aleatórios

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

$$\begin{pmatrix} \text{valor atual} \\ \text{da saída} \\ \text{do modelo} \end{pmatrix} + \begin{pmatrix} \text{combinação linear} \\ \text{dos valores passados} \\ \text{da saída do modelo} \end{pmatrix} = \begin{pmatrix} \text{combinação linear} \\ \text{dos valores} \\ \text{presente e passados} \\ \text{da entrada do modelo} \end{pmatrix} \quad (2)$$

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

Classificação de processos aleatórios

A estrutura do filtro linear que processará os dados, é determinada pela maneira que as duas combinações lineares indicadas na Eq. (2) 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.

Modelo auto-regressivo (AR)

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:

$$x(n) + a_1 x(n-1) + \cdots + a_M x(n-M) = v(n) \qquad (3)$$

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. (3) da seguinte forma:

$$x(n) = b_1 x(n-1) + b_2 x(n-2) + \cdots + b_M x(n-M) + v(n) \quad (4)$$

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 $v(n)$.

Modelo auto-regressivo (AR) - cont.

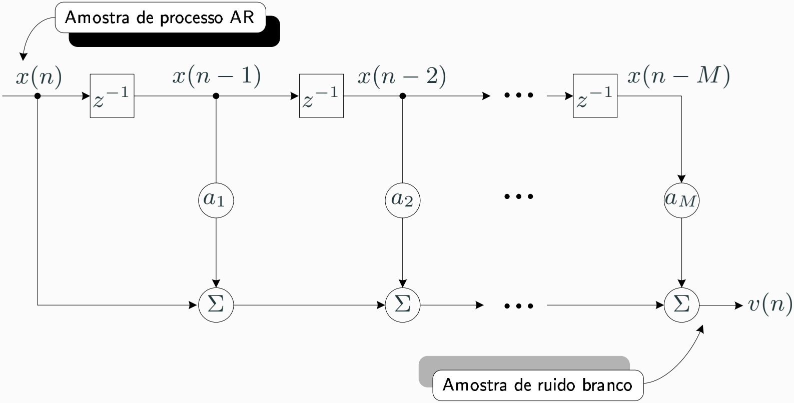


Figura 5: Analisador de processo AR

Modelo auto-regressivo (AR) - cont.

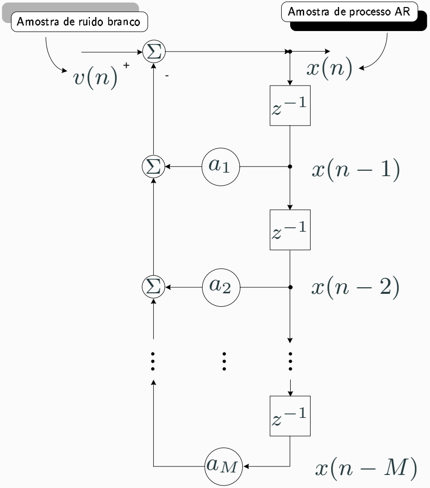


Figura 6: Gerador de processo AR

Modelo média móvel (MA)

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

$$x(n) = v(n) + b_1 v(n-1) + b_2 v(n-2) + \ldots + b_K v(n-K) \quad (5)$$

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.

Modelo média móvel (MA) - cont.

A Equação (5), representa uma versão escalar de um produto interno. Com isso, podemos representá-la como:

$$x(n) = \sum_{i=0}^{K} b_i v(n-i) = \mathbf{v}\mathbf{b}^T \tag{6}$$

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.

Modelo média móvel (MA) - cont.

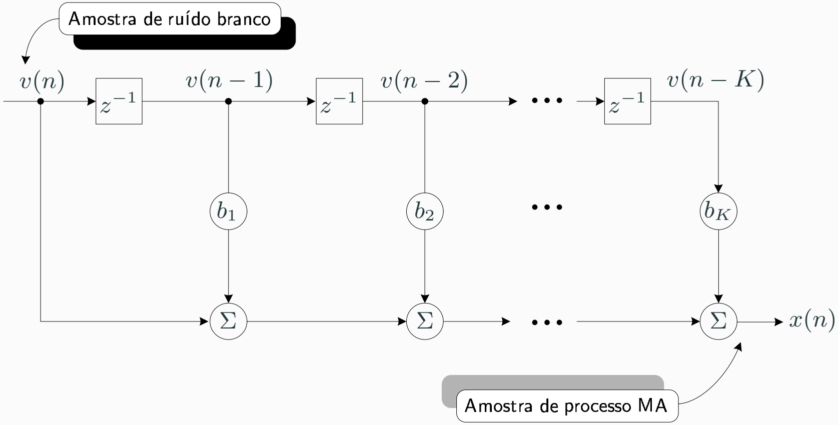


Figura 7: Modelo gerador de um processo de média móvel.

Modelo auto-regressivo-média móvel (ARMA)

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

$$\begin{aligned} x(n) + a_1 x(n-1) + \ldots + a_M x(n-M) = v(n) + b_1 v(n-1) \\ + b_2 v(n-2) + \ldots + b_K v(n-K) \end{aligned} \tag{7}$$

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).

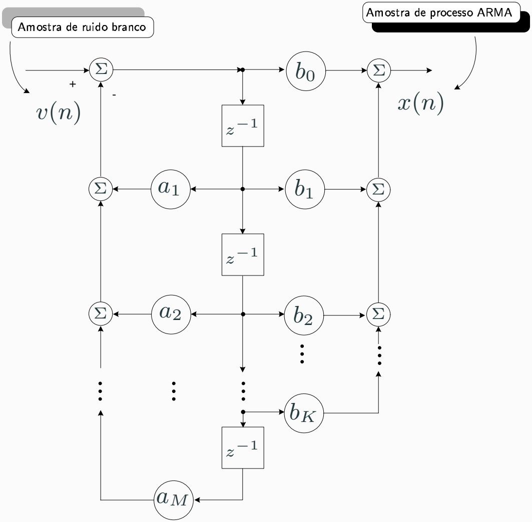


Figura 8: Modelo gerador de um processo ARMA de ordem (M, K), supondo $M > K$.

✔ **Mencionado anteriormente:** *em cada instante de tempo, temos uma variável aleatória*;

✔ **Consideração:** no instante t_n estamos observando a v.a. X_n;

➥ **Expressões de representação da incerteza**

$$\boxed{F_{X_n}(x_n; t_n) = \Pr\{X_n \leq x_n; t_n\}} \tag{8}$$

e

$$\boxed{p_{X_n}(x_n; t_n) = \frac{\partial F_{X_n}(x_n; t_n)}{\partial x_n}} \tag{9}$$

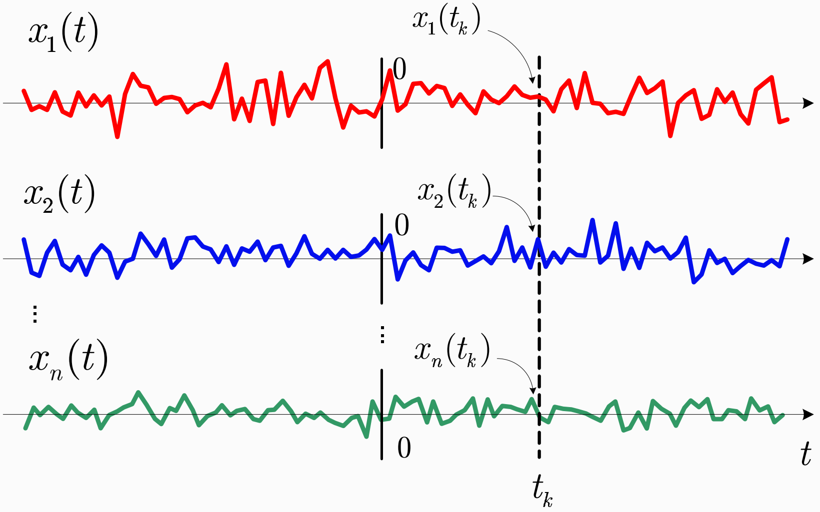
Estatísticas de processos estocásticos - cont.

As expressões em (14) e (11) são chamadas, respectivamente,

- **cdf de primeira ordem**, ou cdf unidimensional, de $X(t)$
- **pdf de primeira ordem**, ou pdf unidimensional de $X(t)$

Ambas caracterizam o comportamento da v.a. em cada instante de tempo.

Descrição de processo estocástico em um ponto



Estatísticas de segunda ordem

Podemos ainda observar um processo em dois pontos distintos no tempo. Teremos assim duas v.a.'s para observar e neste caso precisamos de métricas **conjuntas** para tal caracterização.

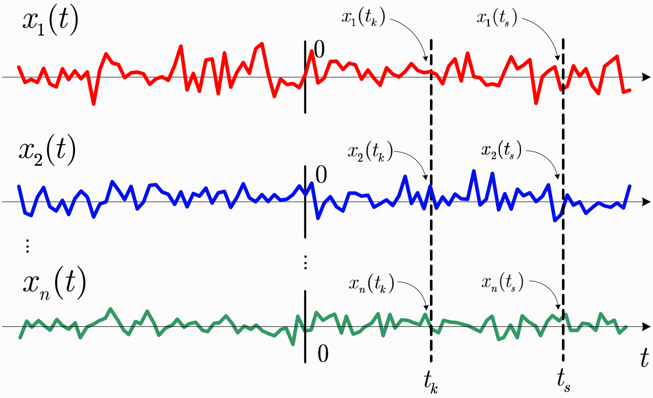
Temos então:

$$\boxed{F_{X_k X_s}(x_k, x_s; t_k, t_s) = \Pr\{X_k \leq x_k, X_s \leq x_s; t_k, t_s\}} \tag{10}$$

e

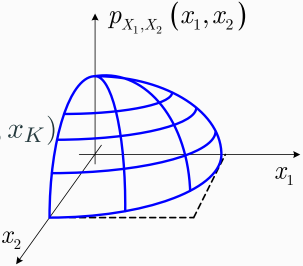
$$\boxed{p_{X_k X_s}(x_k, x_s; t_k, t_s) = \frac{\partial F_{X_k X_s}(x_k, x_s; t_k, t_s)}{\partial x_k \partial x_s}} \tag{11}$$

Descrição de processo estocástico em dois pontos



Representação de processo em dois pontos

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



Generalizando...

$$\boxed{F_{X_1 \ldots X_n}(x_1, \ldots, x_n; t_1 \ldots, t_n) = \Pr\{X_1 \leq x_1 \ldots, X_n \leq x_n; t_1 \ldots, t_n\}} \tag{12}$$

e

$$\boxed{p_{X_1 \ldots X_n}(x_1, \ldots, x_n; t_1 \ldots, t_n) = \frac{\partial F_{X_1 \ldots X_n}(x_1 \ldots, x_n; t_1 \ldots, t_n)}{\partial x_1 \ldots \partial x_n}} \tag{13}$$

Distribuições e densidades marginais

$$\boxed{F_{X_k}(x_k; t_k) = F_{X_k X_s}(x_k, \infty; t_k, t_s)} \tag{14}$$

$$\boxed{p_{X_k}(x_k; t_k) = \int_{-\infty}^{\infty} p_{X_k X_s}(x_k, x_s; t_k, t_s)\ dx_s} \tag{15}$$

Para simplificar/encurtar a notação podemos usar vetores.

Pode-se tomar o vetor

$$\mathbf{x} = \left[\begin{array}{cccc} x_1 & x_2 & \dots & x_n \end{array} \right], \tag{16}$$

em que cada $x_i \in \mathbb{R}$ é uma v.a.

Neste caso, poderíamos ter as seguintes correspondências:

$$x_i = X(t_i) \qquad \text{ou} \qquad x_i = X(n_i) \tag{17}$$

Notação compacta - vetorial (cont.)

Pode-se ainda definir um outro vetor

$$\mathbf{x}_0 = \begin{bmatrix} x_{0,1} & x_{0,2} & \ldots & x_{0,n} \end{bmatrix}, \tag{18}$$

cujas entradas $x_{0,i}$ são valores **fixos**.

Com isso, podemos ter uma representação mais compacta para a cdf e pdf quando representando processos estocástico em múltiplos pontos.

Notação compacta - vetorial (cont.)

$$\boxed{F_{\mathbf{x}}(\mathbf{x};\mathbf{t}) = \Pr\left\{\mathbf{x} \leq \mathbf{x}_0;\mathbf{t}\right\}} \tag{19}$$

e

$$\boxed{p_{\mathbf{x}}(\mathbf{x};\mathbf{t}) = \frac{\partial F_{\mathbf{x}}(\mathbf{x};\mathbf{t})}{\partial \mathbf{x}}} \tag{20}$$

em que $\mathbf{t} = \begin{bmatrix} t_1 & t_2 & \dots & t_n \end{bmatrix}$, cujas entradas podem ser referentes a amostras de tempo contínuo ou discreto.

Notação compacta - vetorial (cont.)

A notação vetorial é bastante útil quando queremos usar muitas variáveis.

Também é interessante pois permite a mesma notação para o vetor de variáveis em relação ao tempo contínuo ou discreto.

Próximos passos

Na sequência iremos discutir algumas medidas estatísticas importantes para variáveis aleatórias e processos estocásticos.

Processos Estacionários e Ergódicos

Valor médio

Uma medida estatística importante também para processos estocásticos é a noção de **valor médio**.

O valor médio, ou esperança, é definida para uma **variável aleatória**. Assim, para um processo estocástico, a média seria definida para cada instante de tempo, contínuo ou discreto.

Ou seja, teríamos a média, para um dado instante t_i, considerando $x_i = X(t_i)$

$$\mu_X(t_i) = \mathbb{E}\left\{X(t_i)\right\} = \int_{-\infty}^{\infty} x_i p_X(x_i; t_i) dx_i \tag{21}$$

Valor médio

Assim, podemos definir para um instante genérico de tempo a seguinte expressão para o valor médio de um processo $X(t)$

$$\boxed{\mu_X(t) = \mathbb{E}\left\{X(t)\right\} = \int_{-\infty}^{\infty} x(t) p_X(x;t) dx} \tag{22}$$

Pode-se definir o valor médio para processos discretos, substituindo a integral pelo somatório.

Valor médio - cont.

Observações sobre a média de um sinal aleatório

- A média de um sinal estocástico é função de t_i.
- A média pode ter valor diferente para diferentes valores de t_i.
- Para resolver a integral na Eq. (22) devemos conhecer todos os *infinitos* possíveis valores que x_i pode assumir em t_i.
- Assim, para saber o valor médio da amplitude do sinal no instante t_i necessitamos de **infinitas** realizações do sinal estocástico $X(t)$.

Valor médio - exemplo

$\mathbb{E}\left\{X(t)\right\} =$?

Seja $X(t) = \cos(\omega_0 t + \theta)$ em que ω_0 é uma constante e θ uma v.a. de acordo com

$$p_\theta\left(\theta\right) = \begin{cases} \frac{1}{2\pi}, & \text{para } -\pi \leq \theta \leq \pi \\ 0, & \text{caso contrário} \end{cases}$$

Calcular $\mathbb{E}\left\{X(t)\right\}$.

Resposta:

$$\mu(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(\omega t + \theta) d\theta = 0$$

Outras características?

Embora o valor médio forneça uma importante característica dos processos estocásticos ele não é a única maneira de ter acesso à informações.

Uma vez que um processo estocástico tem uma evolução temporal uma importante medida é avaliar a relação entre as amostras do processo em dois instantes de tempo distintos.

Esta medida é o que chamamos de **função de autocorrelação** e será o objeto de estudo na sequência.

Definição

A **Função de Autocorrelação** (FAC) de $X(t)$ para dois instantes de tempo $t_1, t_2 \in \mathbb{R}$ e $t_2 > t_1$, é definida como:

$$\begin{aligned} R_X(t_1, t_2) &= \mathbb{E}\left\{X(t_1)X(t_2)\right\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 p_{X_1X_2}(x_1, x_2) dx_1 dx_2 \end{aligned} \tag{23}$$

em que $x_1 = X(t_1)$, $x_2 = X(t_2)$ e $p_{X_1X_2}(x_1, x_2)$ é a pdf conjunta de $x_1 = X(t_1)$ e $x_2 = X(t_2)$.

Pergunta

Mas o que mede a função autocorrelação?

❶ A FAC avalia o quanto a amplitude do sinal em um instante de tempo t_1 está **estatisticamente associada** à amplitude do sinal em outro instante de tempo t_2.

❷ A FAC fornece valores na faixa $(-\infty, +\infty)$.

Função de autocorrelação - cont.

Importante 1

- Valores positivos indicam que quando a amplitude no instante t_1 tende a crescer, a amplitude no instante t_2 tende a crescer também.
- Valores negativos indicam que quando a amplitude no instante t_1 tende a crescer, a amplitude no instante t_2 tende a decrescer.
- Valores próximos de zero indicam que tanto faz se a amplitude no instante t_1 tende a crescer (ou descrescer), pois não se verifica nenhuma tendência de crescimento (ou decrescimento) da amplitude no instante t_2.

Importante 2

- Valores elevados da FAC para dois instantes quaisquer t_1 e t_2 indicam a presença de componentes de baixa freqüência no sinal.
  - Quando o sinal é predominantemente de baixa freqüência as amplitudes consecutivas $X(t_1)$ e $X(t_2)$ possuem valores muito parecidos.
- Valores baixos da FAC para dois instantes quaisquer t_1 e t_2 indicam a presença de componentes de alta freqüência no sinal.
  - Quando o sinal é predominantemente de alta freqüência as amplitudes consecutivas $X(t_1)$ e $X(t_2)$ possuem valores bem diferentes.

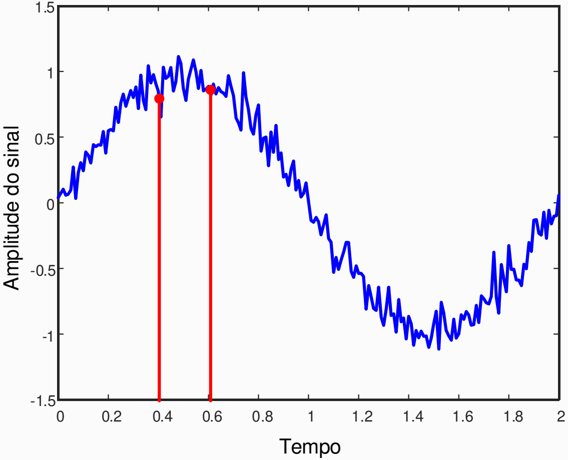


Figura 9: Sinal aleatório com conteúdo harmônico considerado de baixa frequência.

Função de autocorrelação - cont.

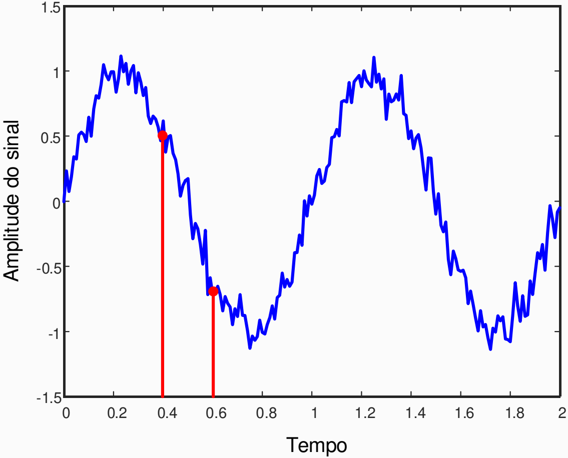


Figura 10: Sinal aleatório com conteúdo harmônico de alta frequência (em relação ao anterior).

Função de autocorrelação - cont.

- ✔ A amplitude do sinal mostrado na segunda figura varia bastante entre instantes consecutivos t_1 e t_2, $t_2 > t_1$, enquanto a amplitude do sinal da primeira figura varia menos.
- ✔ Quanto mais distintos forem os valores consecutivos das amplitudes $X(t_1)$ e $X(t_2)$, menor será a correlação entre eles.
- ✔ Quanto mais parecidos forem os valores consecutivos das amplitudes $X(t_1)$ e $X(t_2)$, maior será a correlação entre eles.

Função de autocorrelação - cont.

Comentário Importante

A interpretação da FAC em termos do conteúdo harmônico de um sinal será muito útil mais adiante quando formos definir o conceito de função densidade espectral de potência.

Função de autocorrelação - cont.

Matriz de autocorrelação (MAC)

Podemos unificar uma notação para avaliar todos os pares de variáveis na autocorrelação dada na Eq. (23). Desta forma, definimos a *matriz de autocorrelação* (MAC) como

$$\mathbf{R}_{\mathbf{x}} = \mathbb{E}\left\{\mathbf{x}\mathbf{x}^T\right\} = \begin{bmatrix} \mathbb{E}\{|x_1|^2\} & \mathbb{E}\{x_1x_2\} & \cdots & \mathbb{E}\{x_1x_N\} \\ \mathbb{E}\{x_2x_1\} & \mathbb{E}\{|x_2|^2\} & \cdots & \mathbb{E}\{x_2x_N\} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}\{x_Nx_1\} & \mathbb{E}\{x_Nx_2\} & \cdots & \mathbb{E}\{|x_N|^2\} \end{bmatrix} \tag{24}$$

Função de autocorrelação - cont.

Propriedades de $\mathbf{R_x}$

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

$$\mathbf{a}^H \mathbf{R_x} \mathbf{a} \geq 0 \tag{25}$$

   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. É uma matriz Toeplitz.

Função de autocorrelação - cont.

- Note que a definição da FAC dada na Eq. (23) assume, implicitamente, que as médias $\mu_X(t_i)$ do sinal em t_1 e t_2 são nulas.
- Esta suposição nem sempre é válida! O que nos levará à definição de **Função de autocovariância**.
- Relembrando a expressão da covariância para v.a.'s

$$
\begin{aligned}
C(X,Y) &= \mathbb{E}\left\{(X-\mu_X)(Y-\mu_Y)\right\} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \cdot p_{XY}(x,y)\; dxdy
\end{aligned}
\tag{26}
$$

Definição

Seja $p_{X_1X_2}(x_1,x_2)$ a PDF conjunta do sinal $X(t)$ nos instantes t_1 e t_2. Então a **Função de Autocovariância** (FACV) de $X(t)$ nestes dois instantes de tempo é definida como:

$$\boxed{C_X(t_1,t_2) = \mathbb{E}\left\{[X(t_1)-\mu_X(t_1)]\,[X(t_2)-\mu_X(t_2)]\right\}} \tag{27}$$

em que $\mu_X(t_i) = \mathbb{E}\left\{X(t_i)\right\}$ é a média do sinal no instante t_i, $i = 1, 2$.

- Após alguma manipulação algébrica, chegamos a uma expressão bastante útil que relaciona a FAC com a FACV:

$$\boxed{C_X(t_1, t_2) = R_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2)} \tag{28}$$

Para obter a FACV basta calcular a FAC e subtrair o produto das médias correspondentes.

- Para sinais de **médias nulas** tem-se, obviamente, que

$$\boxed{C_X(t_1, t_2) = R_X(t_1, t_2)} \tag{29}$$

Função de autocovariância - cont.

Matriz de autocovariância

De forma análoga à autocorrelação, podemos ainda definimos uma matriz que tenha todas as autocovariâncias de pares de v.a's como

$$\boxed{\mathbf{C}_{\mathbf{x}} = \mathbb{E}\left\{(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T\right\}} \tag{30}$$

e os elementos

$$c_{ij} = \mathbb{E}\left\{(x_i - \mu_i)(x_j - \mu_j)\right\} \tag{31}$$

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

Função de autocovariância - cont.

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

$$\boxed{\mathbf{C_x} = \mathbf{R_x} - \boldsymbol{\mu}_\mathbf{x}\boldsymbol{\mu}_\mathbf{x}^T} \tag{32}$$

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

Função de autocovariância - cont.

Algumas propriedades da FACV

1. Se $X(t_i)$ e $X(t_j)$ forem independentes

$$\mathbb{E}\{X(t_i)X(t_j)\} = \mathbb{E}\{X(t_i)\} \cdot \mathbb{E}\{X(t_j)\} = \mu_X(t_i)\mu_X(t_j)$$

$$\boxed{\Rightarrow \text{cov}(X(t_i), X(t_j)) = 0}$$

2. Dizemos que $X(t_i)$ e $X(t_j)$ são **descorrelacionados** se e somente se

$$\text{cov}(X(t_i), X(t_j)) = 0$$

3. Se $\text{cov}(X, Y) = 0$ **não implica** que X e Y são independentes

Exemplo

$\theta \sim$ v.a. uniforme entre $[0, 2\pi]$

$$\begin{cases} X = \sin(\theta) \\ Y = \cos(\theta) \end{cases}$$

$$\text{cov}(X, Y) = \mathbb{E}\{XY\} - \mathbb{E}\{X\}\mathbb{E}\{Y\}$$

$$\mathbb{E}\{X\} = 0 \quad \text{e} \quad \mathbb{E}\{Y\} = 0$$

$$\mathbb{E}\{XY\} = \mathbb{E}\{\sin(\theta) \cdot \cos(\theta)\} = \int_0^{2\pi} \frac{1}{2\pi} \sin(\theta) \cdot \cos(\theta)\ d\theta = 0$$

$\text{cov}(X, Y) = 0$ não implica X e Y independentes!

Função de autocovariância - cont.

Uma classe importante de processos é quando temos processos cujas amostras são **independentes**. Neste caso, devemos ter a seguinte relação entre as densidades conjunta e as marginais:

$$\boxed{p_{\mathbf{x}}(\mathbf{x};\mathbf{t}) = \prod_{i=1}^{N} p_X(x_i;t_i)} \tag{33}$$

Isso implica então que

$$\boxed{\mathbb{E}\left\{X_1 \cdot X_2 \cdots X_N\right\} = \mathbb{E}\left\{X_1\right\} \cdot \mathbb{E}\left\{X_2\right\} \cdots \mathbb{E}\left\{X_N\right\}} \tag{34}$$

Função de autocovariância - cont.

Pode-se notar que a independência implica em descorrelação. Mas, em geral, o contrário não é verdade (como visto no exemplo anterior!).

Um único caso que descorrelação implica em independência ocorre para variáveis com distribuição gaussiana (normal).

Uma definição importante que deriva da definição de FACV é a da **variância** de um sinal aleatório.

Definição

Seja $p_{X_i}(x_i)$ a PDF do sinal $X(t)$ no instante t_i. Então a variância de $X(t)$ neste instante é definida como:

$$\boxed{\begin{aligned}\sigma_X^2(t_i) &= \mathbb{E}\left\{[x(t_i) - m_X(t_i)]^2\right\} \\ &= \int_{-\infty}^{\infty} [x_i - \mu_X(t_i)]^2 \, p_{X_i}(x_i; t_i) dx_i\end{aligned}} \quad (35)$$

em que $\mu_X(t_i) = \mathbb{E}\{X(t_i)\}$ é a média do sinal em t_i.

Função de autocovariância - cont.

Por sua vez, da definição de variância deriva-se o conceito de **valor quadrático médio** de um sinal aleatório.

Definição

Considere que o sinal $X(t)$ tem média nula. Então, o valor quadrático médio de $X(t)$ é definido como:

$$\boxed{\begin{aligned}\mathbb{E}\left\{x^2(t_i)\right\} &= R_X(t_i, t_i) \\ &= \int_{-\infty}^{\infty} x_i^2 p_{X_i}(x_i; t_i) dx_i\end{aligned}} \tag{36}$$

- Valores da correlação ou covariância pertencem ao intervalo $(-\infty, \infty)$
- Versão $normalizada$ da FACV é conhecida como **coeficiente de autocorrelação** (CAC).

Definição

Sejam $C_X(t_1, t_2)$, $\sigma_X(t_1)$ e $\sigma_X(t_1)$, respectivamente, a FACV e as variâncias do sinal $X(t)$ para os instantes t_1 e t_2. Então, o coeficiente de autocorrelação (CAC) de $X(t)$ é dado por

$$\boxed{\rho_X(t_1, t_2) = \frac{C_X(t_1, t_2)}{\sigma_X(t_1)\sigma_X(t_2)}} \tag{37}$$

Função de autocovariância - cont.

Pode-se mostrar (exercício proposto!) que

$$\boxed{-1 \leq \rho_X(t_1, t_2) \leq 1} \tag{38}$$

Com isso, temos que:

1. Valores de ρ próximos de 1: fortemente correlacionados positivamente
2. Valores de ρ próximos de -1: fortemente correlacionados negativamente
3. Valores de ρ muito próximos de 0: sem correlação

Estacionariedade

- Um conceito importante em processamento de sinais aleatórios é a **estacionariedade**.
- Este conceito está associado com a *invariância* (ou constância) das propriedades estatísticas de um processo estocástico.
- Considere a seguinte realização de um processo estocástico:

$$\{X(t_1), X(t_2), ..., X(t_k)\}$$

- Considere agora **uma nova realização** obtida τ instantes de tempo depois, ou seja:

$$\{X(t_1 + \tau), X(t_2 + \tau), ..., X(t_k + \tau)\}$$

- **As estatísticas serão modificadas por conta da diferença temporal?**

Estacionariedade - cont.

Definição

Um processo estocástico é *fortemente estacionário* ou *estacionário no sentido estrito* se, e somente se, **todas** as pdfs de ordem $k \geq 1$ forem invariantes sob uma translação do tempo $\tau \in \mathbb{R}$.

Ou seja,

$$\boxed{\begin{aligned} p_{X_1}(x_1) &= p_{X_1}(x_1; t+\tau) \\ p_{X_1X_2}(x_1, x_2) &= p_{X_1X_2}(x_1, x_2; t_1+\tau, t_2+\tau) \\ \vdots \quad &\vdots \qquad \vdots \\ p_{\mathbf{x}}(\mathbf{x}; \mathbf{t}) &= p_{\mathbf{x}}(\mathbf{x}; \mathbf{t}+\tau) \end{aligned}} \tag{39}$$

Estacionariedade - cont.

Observações importantes

1. A condição de estacionariedade forte é um requisito bastante severo, pois exige que **todas** as pdfs associadas sejam invariantes no tempo;
2. Esta condição é bastante difícil de ser verificada ou testada na prática. Uma definição "menos" exigente de estacionariedade é freqüentemente usada (ou assumida).

Observações importantes - cont

3. Quando a condição só é válida para $1 \leq r \leq k$, diz-se que o processo estocástico $X(t)$ é *estacionário de ordem r*.
4. Um caso importante e de bastante interesse prático diz respeito a um processo estocástico estacionário de ordem $r = 2$.
5. Este processo estacionário é chamado **fracamente estacionário** ou **estacionário no sentido amplo** (WSS).

Estacionariedade - cont.

Processo WSS - Condições

Para um processo estacionário no sentido amplo as seguintes condições devem ser válidas:

Condição ❶ A média do processo $\mu_X(t_i)$ é constante.

Condição ❷ A variância do processo $\sigma_X^2(t_i)$ é constante.

Condição ❸ A FAC do processo $R_X(t_1, t_2) = R_X(\tau)$, em que $\tau = t_2 - t_1$ $(t_1 < t_2)$.

As condições acima são mais fáceis de se verificar na prática, permitindo classificar um sinal estocástico em estacionário ou não-estacionário.

Condição 1

Seja $X(t)$ um processo estocástico contínuo. Se o processo é estacionário, então podemos escrever que:

$$p_{X_t}(x;t) = p_{X_t}(x;t+\tau), \quad \forall \tau \in \mathbb{R} \tag{40}$$

Assim, a média de $X(t+\tau)$ é dada por:

$$\begin{aligned}\mu_X(t+\tau) = \mathbb{E}[X(t+\tau)] &= \int_{-\infty}^{\infty} x(t+\tau) p_{X_t}(x;t+\tau)dx \\ &= \int_{-\infty}^{\infty} x(t+\tau) p_{X_t}(x;t)dx = \mathbb{E}[X(t)] = \mu_X(t)!\end{aligned}$$

Condição 3

Seja $X(t)$ um processo estocástico contínuo. Assim, a FAC de $X(t)$ é dada por:

$$
\begin{aligned}
\mathbb{E}[X(t_1)X(t_2)] &= R_X(t_1, t_2) = R_X(t_1 + \tau, t_2 + \tau) \\
&= R_X(t_1 - t_2, 0) = R_X(\tau), \quad \forall \tau \in \mathbb{R}
\end{aligned} \tag{41}
$$

em que assumimos que

$$
p_{X_1 X_2}(x_1, x_2; t_1, t_2) = p_{X_1 X_2}(x_1, x_2; t_1 + \tau, t_2 + \tau) \tag{42}
$$

Note que a FAC depende apenas da diferença entre t_1 e t_2. Por isso, em vez de escrever $R_X(t_1 - t_2, 0)$, costumamos abreviar a notação e escrever $R_X(t_1 - t_2, 0) = R_X(\tau)$.

Estacionariedade - cont.

A figura abaixo mostra um sinal aleatório estacionário (pelo menos para o período de observação do mesmo!).

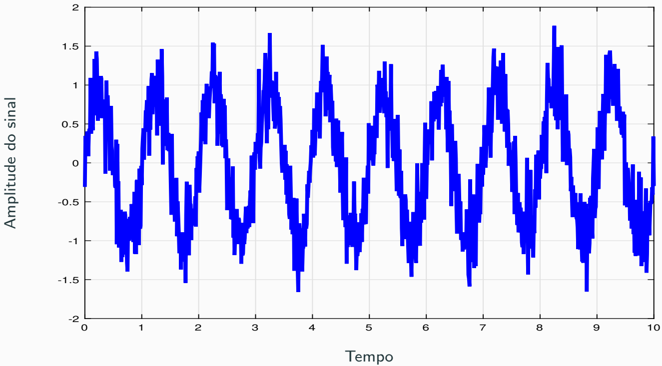


Figura 11: Sinal aleatório estacionário.

Estacionariedade - cont.

- A figura anterior mostra uma realização do seguinte processo estocástico, simulado no Matlab/Octave:

$$X(t) = \sin(\omega t) + \varepsilon(t) \tag{43}$$

  em que $\varepsilon(t)$ é uma variável aleatória gaussiana de média zero e variância $\sigma_\varepsilon^2 = 0.1$, $\omega = 2\pi/T$ é a freqüência angular constante e $T = 10$ segundos.
- O sinal gerado obedece todas as três condições de estacionariedade fraca.
- O tipo de sinal estocástico mostrado na Eq. (43) simula um fenômeno comumente encontrado na prática: um sinal determinístico contaminado com ruído.

Estacionariedade - cont.

- A parte senoidal do sinal poderia corresponder a uma onda de tensão/corrente gerada por determinado equipamento elétrico, enquanto a v.a. $\varepsilon(t)$ corresponderia ao ruído que interfere na medição desta onda de tensão/corrente.
- Na prática, **mede-se apenas** o sinal ruidoso $X(t)$.
- Muitas vezes o nível de ruído é tão elevado que não temos certeza absoluta que o sinal medido possui uma componente senoidal.
- Mais adiante estudaremos uma técnica baseada na FAC para detectar sinais periódicos imersos em ruído.

A figura abaixo mostra um sinal aleatório não-estacionário na média (por quê?).

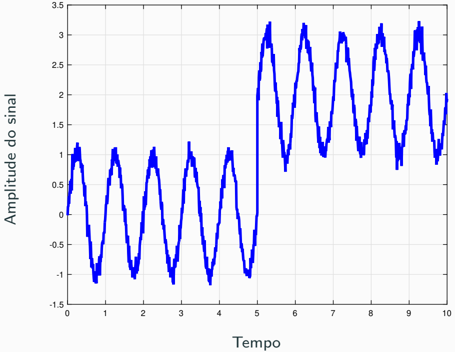


Figura 12: Processo não-estacionário na média.

Neste caso temos mostra um sinal aleatório não-estacionário na variância (por quê?).

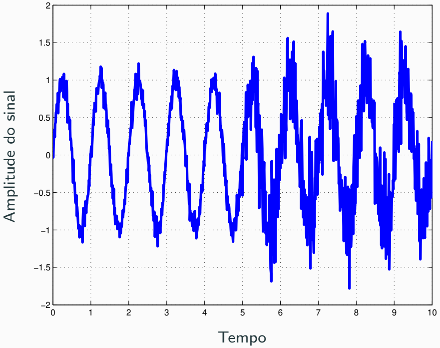


Figura 13: Processo não-estacionário na variância.

E agora um sinal aleatório não-estacionário na FAC (por quê?).

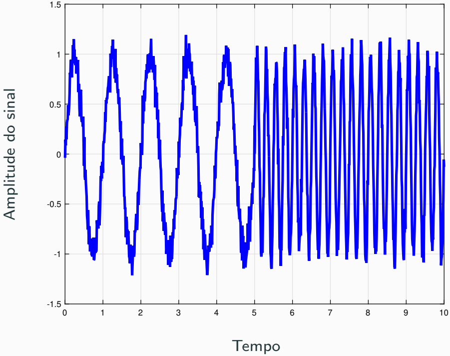


Figura 14: Processo não-estacionário na FAC.

Estacionariedade - cont.

A consideração de processos WSS permite simplificar as expressões da FAC, FACV e CAC, escrevendo-as em função de $\tau = t_2 - t_1$.

- Por exemplo, a FAC passa ser escrita como:

$$R_X(t_1, t_2) = R_X(\tau) = \mathbb{E}\left\{X(t)X(t+\tau)\right\} \tag{44}$$

- A FACV por sua vez, passa ser escrita como:

$$C_X(t_1, t_2) = C_X(\tau) = \mathbb{E}\left\{(X(t) - \mu_X)(X(t+\tau) - \mu_X)\right\} \tag{45}$$

- Por fim, o CAC passa ser escrito como:

$$\rho_X(t_1, t_2) = \rho_X(\tau) = \frac{C_X(\tau)}{\sigma_X^2} \tag{46}$$

Como a FAC de um processo WSS depende apenas de τ ($\tau > 0$), pode-se escrever a FAC de duas maneiras equivalentes:

$$\boxed{R_X(\tau) = \mathbb{E}\left\{X(t)X(t+\tau)\right\}} \tag{47}$$

ou

$$\boxed{R_X(\tau) = \mathbb{E}\left\{X(t-\tau)X(t)\right\}} \tag{48}$$

Note que em ambas definições temos $t_2 - t_1 = \tau$.

A mesma observação vale para a definição da FACV e do CAC.

Propriedades gerais da FAC

1. $R_X(0)$ é o valor quadrático médio do processo $X(t)$:

$$\mathbb{E}\left\{x^2(t_i)\right\} = R_X(t_i, t_i) = R_X(t_i - t_i) = R_X(0) \quad (49)$$

2. $R_X(0) = \sigma_X^2$ se o processo tiver média nula.
3. $R_X(\tau)$ é uma função par de seu argumento: $R_X(\tau) = R_X(-\tau)$. Para fins de análise consideramos apenas os valores correspondentes a $\tau > 0$.
4. $|R_X(\tau)| \leq R_X(0)$, para todo τ.
5. Se $X(t)$ contém uma componente periódica, $R_X(t)$ também conterá uma componente de mesmo período.

Propriedades gerais da FAC (cont.)

❻ Se a componente periódica é do tipo senoidal, então $R_X(t)$ perderá informação sobre a fase da componente senoidal.

❼ Se $X(t)$ não contém componentes periódicas, $R_X(\tau)$ tende a zero, à medida que $\tau \to \infty$.

❽ Se para um determinado processo estocástico $X(t)$ obtivermos $R_X(\infty) = 0$ implica dizer que o processo tem média zero.

Estacionariedade - cont.

Modelo AR

$$x(n) = v(n) + \sum_{i=1}^{M} a_i x(n-i) \tag{50}$$

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.

Podemos então ver como as correlações de tal processo WSS estão organizadas.

Estacionariedade - cont.

Modelo AR - cont.

$$
\begin{aligned}
r_X(0) &= \mathbb{E}\left\{x^2(n)\right\} \\
&= \mathbb{E}\left\{v^2(n) + 2 \cdot a \cdot v(n) \cdot x(n-1) + a^2 x^2(n-1)\right\} \\
&= \sigma_v^2 + 2a\mathbb{E}\left\{v(n) \cdot x(n-1)\right\} + a^2 \cdot r_X(0) \\
&= \sigma_v^2 + a^2 \cdot r_X(0)
\end{aligned}
\tag{51}
$$

Daí resulta

$$
\boxed{r_X(0) = \frac{\sigma_v^2}{1-a^2}} \tag{52}
$$

Modelo AR - cont.

$$\begin{aligned} r_X(1) &= \mathbb{E}\left\{x(n)x(n-1)\right\} \\ &= \mathbb{E}\left\{[v(n)+ax(n-1)][v(n-1)+ax(n-2)]\right\} \\ &= \mathbb{E}\left\{ax(n-1)v(n-1)+a^2x(n-1)x(n-2)\right\} \\ &= \mathbb{E}\left\{a[v(n-1)+ax(n-2)]\cdot v(n-1)\right. \\ &\qquad\qquad \left. +a^2x(n-1)x(n-2)\right\} \\ &= a\sigma_v^2 + a^2 \cdot \mathbb{E}\left\{x(n-1)x(n-2)\right\} \\ &= a\sigma_v^2 + a^2 r_X(1) \end{aligned} \tag{53}$$

Daí, temos

$$\begin{aligned} r_X(1)[1-a^2] &= a\sigma_v^2 \\ r_X(1) &= a\cdot r_X(0) \end{aligned} \tag{54}$$

Modelo AR - cont.

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

$$\begin{aligned} r_X(i) &= a^i \cdot r_X(0) \\ r_X(i) &= a \cdot r_X(i-1) \end{aligned} \tag{55}$$

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

$$\mathbf{R_x} = \begin{bmatrix} \dfrac{\sigma_v^2}{(1-a^2)} & \dfrac{a \cdot \sigma_v^2}{(1-a^2)} \\ \dfrac{a \cdot \sigma_v^2}{(1-a^2)} & \dfrac{\sigma_v^2}{(1-a^2)} \end{bmatrix} \tag{56}$$

Modelo AR - cont.

Calcular λ_1 e λ_2 (autovalores)

$$\begin{aligned}
\det(\mathbf{R_x} - \lambda\mathbf{I}) &= 0 \\
\left(\frac{\sigma_v^2}{1-a^2} - \lambda\right)^2 - \left(\frac{a \cdot \sigma_v^2}{1-a^2}\right)^2 &= \frac{\sigma_v^4}{(1-a^2)^2} - 2\frac{\lambda\sigma_v^2}{(1-a^2)} + \lambda^2 \\
-\frac{a^2\sigma_v^4}{(1-a^2)^2} &= \frac{\sigma_v^4}{(1-a^2)^2} \cdot (1-a^2) - 2\lambda\frac{\sigma_v^2}{(1-a^2)} + \lambda^2 \\
&= \lambda^2 - 2\lambda \cdot \frac{\sigma_v^2}{(1-a^2)} + \frac{\sigma_v^4}{(1-a^2)} = 0 \\
&= (1-a^2) \cdot \lambda^2 - 2\sigma_v^2\lambda + \sigma_v^4 = 0
\end{aligned} \tag{57}$$

Modelo AR - 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

$$\begin{aligned}\lambda_1 &= \frac{\sigma_v^2}{(1-a)} \\ \lambda_2 &= \frac{\sigma_v^2}{(1+a)}\end{aligned} \tag{58}$$

Modelo AR - cont.

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

$$\lambda_1 = \frac{\sigma_v^2}{(1-a)} \quad \text{e} \quad \lambda_2 = \frac{\sigma_v^2}{(1+a)} \tag{59}$$

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

$$\boxed{\mathcal{C}\left(\mathbf{R_x}\right) = \frac{\lambda_{\max}}{\lambda_{\min}}} \tag{60}$$

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

Modelo AR - cont.

Informações fornecidas por $\mathcal{C}\left(\mathbf{R}_\mathbf{x}\right)$

- 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 $\mathcal{C}\left(\mathbf{R}_\mathbf{x}\right) = \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}}$ exprime quão descorrelacionado é $\mathbf{x}(n)$, ou ainda quão "próximo" $\mathbf{R}_\mathbf{x}$ está de uma matriz diagonal
  1. $\mathcal{C}\left(\mathbf{R}_\mathbf{x}\right) = 1 \Rightarrow x(n)$ é branco
  2. $\mathcal{C}\left(\mathbf{R}_\mathbf{x}\right) \to \infty \Rightarrow x(n)$ tem correlação

Ergodicidade

Definição

Um processo estocástico é **ergódico**, se seus momentos estatísticos, calculados ao longo de várias realizações do processo, forem equivalentes aos momentos temporais, calculados a partir de uma única realização do processo.

Definição

Um processo é **ergódico na média**, se e somente se:

$$
\begin{aligned}
\mathbb{E}\left\{X(t_i)\right\} &= \int_{-\infty}^{\infty} x_i p_{X_i}(x_i;t)dx_i \\
&\equiv \begin{cases} \frac{1}{k}\int_0^k x(t)dt, \quad k \to \infty & \text{(contínuo)} \\ \\ \frac{1}{k}\sum_{i=1}^{k} x(t_i), \quad k \to \infty & \text{(discreto)} \end{cases}
\end{aligned} \tag{61}
$$

Ergodicidade - cont.

Definição

Um processo é **ergódico na variância**, se e somente se:

$$
\begin{aligned}
\text{var}[X(t_i)] &= \int\limits_{-\infty}^{\infty} (x_i - \mu_X)^2 p_{X_i}(x_i; t) dx_i \\
&\equiv \begin{cases} \frac{1}{k} \int\limits_{0}^{k} (x(t) - \mu_X)^2 dt, \quad k \to \infty & \text{(contínuo)} \\ \\ \frac{1}{k} \sum\limits_{i=1}^{k} (x(t_i) - \mu_X)^2, \quad k \to \infty & \text{(discreto)} \end{cases}
\end{aligned} \tag{62}
$$

Definição

Um processo é **ergódico na correlação**, se e somente se:

$$
\begin{aligned}
R_X(\tau) &= \mathbb{E}\{X(t)X(t+\tau)\} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_t x_{t+\tau} p_{X_t X_{t+\tau}}(x_t, x_{t+\tau}) dx_t dx_{t+\tau} \\
&\equiv \begin{cases} \frac{1}{k}\int_0^k x(t)x(t+\tau)dt, \quad k \to \infty & \text{(contínuo)} \\ \\ \frac{1}{k-\tau}\sum_{i=1}^{k-\tau} x(t_i)x(t_i+\tau), \quad k \to \infty & \text{(discreto)} \end{cases}
\end{aligned} \tag{63}
$$

Ergodicidade - cont.

- O conceito de ergodicidade é de **fundamental** importância prática para a teoria de processos estocásticos.
- Dizer que um processo é ergódico implica em assumir que podemos inferir todas as informações estatísticas sobre o processo estocástico a partir de uma **única** realização do mesmo.
- Na prática, é muito difícil verificar a estacionariedade no sentido estrito e a ergodicidade de um processo estocástico.
- Assim, para simplificação, assumimos que o processo é estacionário no sentido amplo e ergódico.

Avaliando estacionariedade

Dica prática para verificar estacionariedade!

(1) Dividir o sinal original em um número finito K de sub-sinais de mesmo comprimento N.

(2) Para um certo t_i fixo, estimar a média $\mu_x(t_i)$ e variância $\sigma_x(t_i)$ ao longo das K pseudo-realizações:

$$\widehat{\mu}_X(t_i) \approx \frac{1}{K}\sum_{k=1}^{K} x^{(k)}(t_i), \tag{64}$$

$$\widehat{\sigma}_X^2(t_i) \approx \frac{1}{K}\sum_{k=1}^{K} \left[x^{(k)}(t_i) - \widehat{\mu}_X(t_i)\right]^2, \tag{65}$$

para $i = 1, \ldots, N$.

Dica prática para verificar estacionariedade! - cont.

(3) Calcular a média das médias e das variâncias do item (2):

$$\overline{M} = \frac{\sum_{i=1}^{N} \widehat{\mu}_X(t_i)}{N} \quad \text{e} \quad \overline{V} = \frac{\sum_{i=1}^{N} \widehat{\sigma}_X^2(t_i)}{N} \tag{66}$$

(4) Calcular a média e a variância das amplitudes da realização original completa:

$$\overline{X} \approx \frac{1}{N \cdot K} \sum_{i=1}^{N \cdot K} x(t_i), \tag{67}$$

$$s_X^2 \approx \frac{1}{N \cdot K} \sum_{i=1}^{N \cdot K} \left[x(t_i) - \overline{X} \right]^2. \tag{68}$$

Avaliando estacionariedade - cont.

Dica prática para verificar estacionariedade! - cont.

(5) Analisar a plausibilidade das seguintes hipóteses:

- A média é constante, ou seja, $\overline{M} \approx \overline{X}$.
- A variância é constante, ou seja, $\overline{V} \approx s_X^2$.

(6) Se as hipóteses forem verdadeiras, então pode-se assumir com relativa segurança que o processo é ergódico e estacionário.

Exercício resolvido

- Considere o seguinte processo estocástico:

$$X(t) = A \sin \omega t \tag{69}$$

em que A é uma variável aleatória gaussiana de média zero e variância σ_A^2, ω é uma constante de valor conhecido e t corresponde ao tempo contínuo.

- Suponha que obtemos um valor amostral para A, denotado por A_1. Assim, a realização correspondente de $X(t)$ para esse valor de A_1 é dada por:

$$X_A(t) = A_1 \sin \omega t \tag{70}$$

Exercício resolvido - cont

- Assim, usando a Eq. (69) para um sinal contínuo, obtemos:

$$\begin{aligned}
R_{X_A}(\tau) &= \lim_{k\to\infty} \frac{1}{k}\int_0^k x(t)x(t+\tau)dt \\
&= \lim_{k\to\infty} \frac{1}{k}\int_0^k A_1 \sin \omega t \cdot A_1 \sin \omega(t+\tau)dt \\
&= \frac{A_1^2}{2}\cos \omega\tau
\end{aligned}$$

- Para resolver a integral acima usar a seguinte identidade trigonométrica:

$$\boxed{\sin(a)\cdot\sin(b) = \frac{1}{2}\cos(a-b) - \frac{1}{2}\cos(a+b)}$$

Exercício resolvido - cont

- Por outro lado, usando a definição da FAC, Eq. (23), chegamos ao seguinte resultado:

$$\begin{aligned} R_X(t_1, t_2) &= \mathbb{E}\{X(t_1)X(t_2)\} = \mathbb{E}\{A\sin\omega t_1 \cdot A\sin\omega t_2\} \\ &= \mathbb{E}\left\{A^2\right\}\sin\omega t_1 \cdot \sin\omega t_2 \\ &= \sigma_A^2 \sin\omega t_1 \cdot \sin\omega t_2 \end{aligned} \tag{71}$$

- Como os resultados são **diferentes**, concluímos que este processo não é ergódico.
- Além disso, como a expressão da FAC dada na Eq. (71) não pode ser reduzida a uma simples função de $\tau = t_2 - t_1$, o processo também não é estacionário.

Definição 1

Sejam $X(t)$ e $Y(t)$ dois sinais aleatórios estacionários de tempo contínuo. A Função de Correlação Cruzada (FCC) entre $X(t)$ e $Y(t)$ pode ser definida como:

$$\boxed{\begin{aligned} R_{XY}(\tau) &= \mathbb{E}\left\{X(t)Y(t+\tau)\right\}, \quad \forall \tau > 0 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_t y_{t+\tau} p_{X_t Y_{t+\tau}}(x_t, y_{t+\tau}) dx_t dy_{t+\tau} \end{aligned}} \qquad (72)$$

em que $p_{X_t Y_{t+\tau}}(x_t, y_{t+\tau})$ é a FDP conjunta de $X(t)$ e $Y(t+\tau)$.

Função de correlação cruzada - cont.

Definição 2

A FCC entre $X(t)$ e $Y(t)$ também pode ser definida como:

$$\boxed{\begin{aligned} R_{YX}(\tau) = \mathbb{E}\left\{Y(t)X(t+\tau)\right\}, \quad \forall\tau > 0 \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_t x_{t+\tau} p_{Y_t X_{t+\tau}}(y_t, x_{t+\tau}) dy_t dx_{t+\tau} \end{aligned}} \tag{73}$$

em que $p_{Y_t X_{t+\tau}}(y_t, x_{t+\tau})$ é a FDP conjunta de $Y(t)$ e $X(t+\tau)$.

Propriedades

- A definição de $R_{YX}(\tau)$ dada na Eq. (73) é **invariante** a uma translação de $-\tau$.
- Desta forma, $R_{YX}(\tau)$ é também dada por:

$$R_{YX}(\tau) = \mathbb{E}\left\{Y(t-\tau)X(t)\right\} \tag{74}$$

- Comparando a Eq.(74) com a definição de $R_{XY}(\tau)$ dada na Eq.(72), percebemos que:

$$R_{XY}(\tau) = R_{YX}(-\tau) \tag{75}$$

- Logo, trocar a ordem dos índices da FCC tem o efeito de mudar o sinal do seu argumento.

Propriedades - cont.

- Os valores máximos de $R_{XY}(\tau)$ e $R_{YX}(\tau)$ geralmente não ocorrem para $\tau = 0$.
- Porém, pode-se mostrar que

$$\boxed{|R_{XY}(\tau)| \leq \sqrt{R_X(0)R_Y(0)}} \tag{76}$$

- Para dois processos estocásticos independentes temos que:

$$\boxed{R_{XY}(\tau) = R_{YX}(\tau)} \tag{77}$$

Função de correlação cruzada - cont.

Definição (sinais discretos)

Sejam $X(t) = \{..., x(-2), x(-1), x(0), x(1), x(2), ...\}$ e $Y(t) = \{..., y(-2), y(-1), y(0), y(1), y(2), ...\}$ dois sinais aleatórios estacionários e de tempo discreto. Neste caso, as FCCs são definidas como:

$$\boxed{\begin{aligned} R_{XY}(\tau) &= \sum_{t=-\infty}^{\infty} x(t)y(t+\tau) \\ R_{YX}(\tau) &= \sum_{t=-\infty}^{\infty} y(t)x(t+\tau) \end{aligned}} \tag{78}$$

para $\tau = 0, 1, 2, ...$

Estimando FCC (sinais discretos)

- Sejam $X(t) = \{x(1), x(2), ..., x(N)\}$ e $Y(t) = \{y(1), y(2), ..., y(N)\}$ duas seqüências aleatórias finitas (e.g. sinais amostrados).
- Neste caso, as FCCs podem ser estimadas por meio das seguintes expressões:

$$\boxed{\begin{aligned} r_{XY}(\tau) &= \frac{\sum_{t=1}^{N-\tau} x(t)y(t+\tau)}{N-\tau} \\ r_{YX}(\tau) &= \frac{\sum_{t=1}^{N-\tau} y(t)x(t+\tau)}{N-\tau}, \end{aligned}} \tag{79}$$

para $\tau = 0, 1, 2, ...$

Função de correlação cruzada - cont.

Exemplo

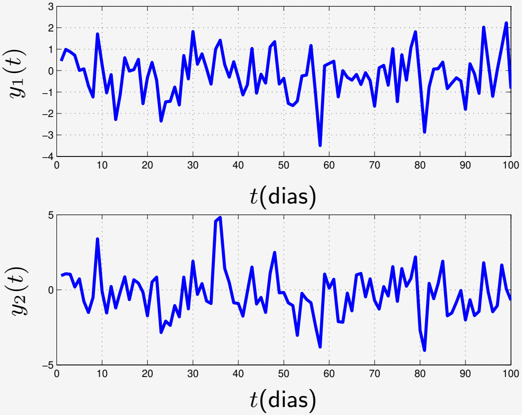


Figura 15: Sinais aleatórios distintos.

Exemplo - cont.

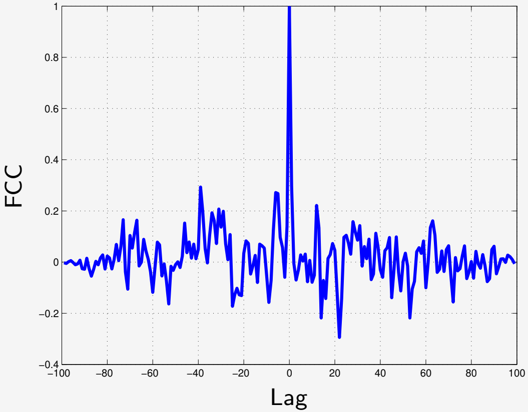


Figura 16: FCC normalizada dos sinais aleatórios distintos.

Para as esperanças conjuntas, ou seja, envolvendo mais de um processo aleatório, podemos definir outras matrizes:

Matriz de correlação cruzada

$$\boxed{\mathbf{R}_{\mathbf{xy}} = \mathbb{E}\left\{\mathbf{x}\mathbf{y}^T\right\}} \tag{80}$$

e

Matriz de covariância cruzada

$$\boxed{\mathbf{C}_{\mathbf{xy}} = \mathbb{E}\left\{(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})^T\right\}} \tag{81}$$

Função de correlação cruzada - cont.

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:

$$\begin{aligned} \mathbf{R_{xy}} &= \mathbf{R}_{\mathbf{yx}}^T \\ \mathbf{C_{xy}} &= \mathbf{C}_{\mathbf{yx}}^T \end{aligned} \tag{82}$$

Função de correlação cruzada - cont.

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

1. Correlação

$$\mathbf{R_{x+y}} = \mathbf{R_x} + \mathbf{R_{xy}} + \mathbf{R_{yx}} + \mathbf{R_y} \tag{83}$$

2. Covariância

$$\mathbf{C_{x+y}} = \mathbf{C_x} + \mathbf{C_{xy}} + \mathbf{C_{yx}} + \mathbf{C_y} \tag{84}$$

Função de correlação cruzada - cont.

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

- Na prática, é comum observar combinações aditivas de sinais estocásticos.
- Por exemplo, seja um processo $Z(t)$ dado pela **soma de dois outros processos estacionários** $X(t)$ e $Y(t)$:

$$Z(t) = X(t) + Y(t) \tag{85}$$

- Pode-se mostrar (exercício!) que a FAC do processo $Z(t)$ é dada por:

$$\begin{aligned} R_Z(\tau) &= \mathbb{E}\left\{Z(t)Z(t+\tau)\right\} \\ &= R_X(\tau) + R_{YX}(\tau) + R_{XY}(\tau) + R_Y(\tau) \end{aligned} \tag{86}$$

FAC na detecção de sinais - cont.

- Se $X(t)$ e $Y(t)$ são processos não-correlacionados e ambos têm média zero, os termos $R_{YX}(\tau) = R_{XY}(\tau)$ são nulos, resultando em

$$\boxed{R_Z(\tau) = R_X(\tau) + R_Y(\tau)} \tag{87}$$

- Assim, concluímos a FAC de um sinal formado pela soma de dois sinais não-correlacionados é equivalente à soma das FACs dos sinais individuais.
- Este resultado pode ser **estendido para a soma de mais de dois** processos estocásticos não-correlacionados.

FAC na detecção de sinais - cont.

- A Eq. (87) provê um procedimento muito interessante para ser utilizado na prática; por exemplo, na detecção de sinais determinísticos em um meio ruidoso, tal como um canal de comunicação.
- Nesta aplicação supõe-se que um determinado sinal estocástico é formado por componentes determinísticas (informação transmitida pelo meio) distorcidas por ruído e que **a informação desejada e o ruído (indesejado) são não-correlacionados**.

Seja o sinal senoidal de amplitude unitária e período $T = 40$ unidades de tempo, mostrado na figura abaixo.

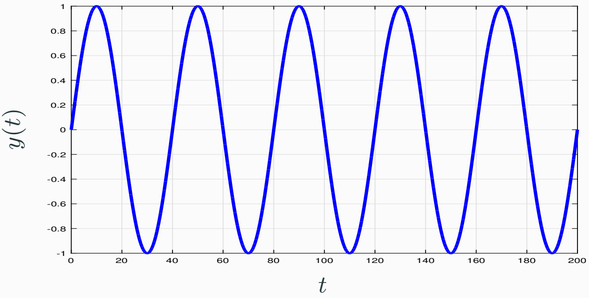


Figura 17: Sinal determinístico.

FAC na detecção de sinais - cont.

De acordo com a teoria, a FAC de um sinal periódico é também periódica com o mesmo período do sinal.

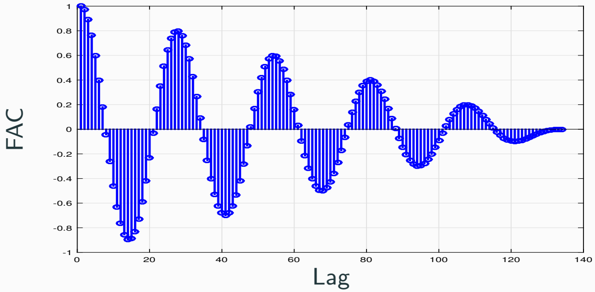


Figura 18: FAC normalizada do sinal determinístico.

Consideremos agora um sinal aleatório, gaussiano, de média 0 e variância 2.

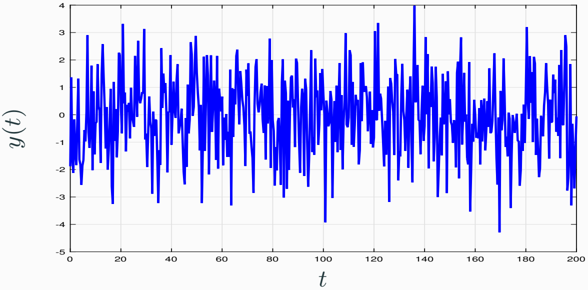


Figura 19: Sinal aleatório com distribuição $\mathcal{N}(0, 2)$.

A teoria nos diz que a FAC do ruído branco é nula, exceto para $\tau = 0$. A FAC estimada a partir do sinal mostrado na figura anterior é mostrado abaixo.

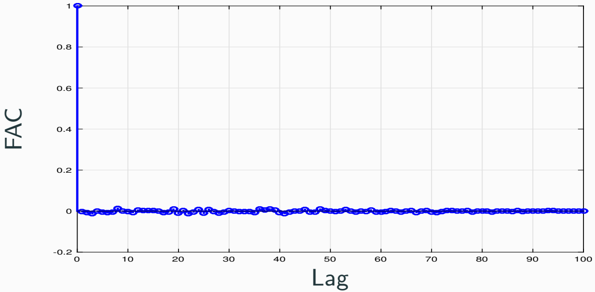


Figura 20: FAC normalizada do sinal aleatório com distribuição $\mathcal{N}(0, 2)$.

Considere agora um sinal formado pela soma da senóide (determinístico) com o sinal aleatório. Este sinal é mostrado abaixo.

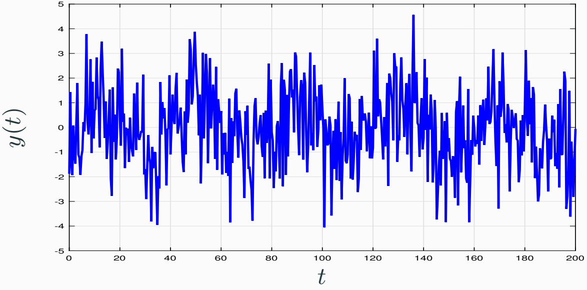


Figura 21: Sinal com senóide mais ruído.

FAC na detecção de sinais - cont.

Pergunta desafio

Você poderia confirmar só no "olhômetro", se no sinal resultando há apenas ruído ou se há alguma informação (sinal determinístico)?

- De acordo com a teoria, a FAC de dois sinais não-correlacionados é a soma das FACs dos sinais individuais.
- Será que isto se confirma?

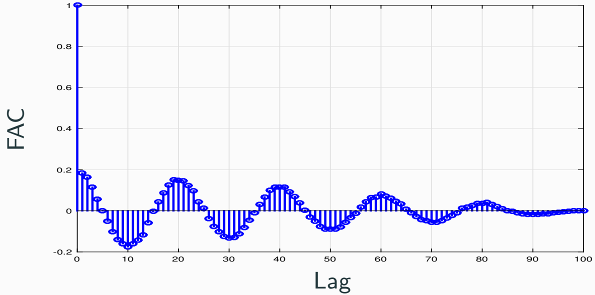


Figura 22: FAC normalizada do sinal senoidal adicionado de ruído.

Assim, concluímos que a FAC pode ser usada para detectar estruturas determinísticas periódicas imersas em sinais estocásticos, podendo inclusive estimar o período do sinal determinístico.

Gráfico de dispersão (lagplot)

Definição

Técnica qualitativa (gráfica) que avalia a correlação serial das amplitudes do sinal através do gráfico de dispersão.

Implementação

- Considere o seguinte sinal de tempo discreto:

$$X = \{x(1), x(2), x(3), ..., x(N)\}$$

- O *lagplot* consiste no gráfico de um conjunto de pares ordenados:

$$[x(n), x(n-\tau)], \quad n = \tau + 1, \ldots, N \tag{88}$$

em que a constante τ ($\tau > 1$) é chamada de *lag* e define o distanciamento temporal entre as amplitudes do sinal.

Gráfico de dispersão (lagplot) - cont.

Comentários

- Em geral, é necessário plotar *lagplots* para vários *lags* para se ter uma idéia da persistência da memória.
- Em outras palavras, vários *lagplots* são necessários para avaliar quanto tempo dura a influência dos valores passados sobre o valor atual do sinal.
- A influência dos valores passados sobre o valor atual pode ser positiva, negativa ou nula.

Exemplo 1

Considere um sinal de tempo discreto $X(n)$, de comprimento $N = 1000$.

Lagplot ($\tau = 1$)

Os pares ordenados que formam o lagplot para $\tau = 1$ são definidos como:

$$[x(n), x(n-1)], \quad n = 2, ..., 1000. \tag{89}$$

Ou seja,

$$[x(2), x(1)], \; [x(3), x(2)], \; \cdots \; , [x(1000), x(999)]. \tag{90}$$

Exemplo 2

Considere novamente um sinal de tempo discreto $X(n)$, de comprimento $N = 1000$.

Lagplot ($\tau = 2$)

Os pares ordenados que formam o lagplot para $\tau = 2$ são definidos como:

$$[x(n), x(n-2)], \quad n = 3, ..., 1000. \tag{91}$$

Ou seja,

$$[x(3), x(1)], \;\; [x(4), x(2)], \;\; \cdots \;\; , [x(1000), x(998)]. \tag{92}$$

E assim sucessivamente para outros *lagplots* do mesmo sinal (i.e.$\tau = 3, ..., 1000$).

Gráfico de dispersão (lagplot) - cont.

Sinal cujas as amplitudes distanciadas de $\tau = 1$ e estão positivamente correlacionadas.

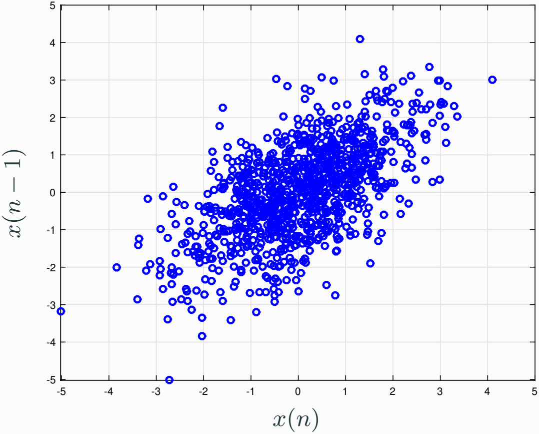


Figura 23: *Lagplot* com $\tau = 1$ e correlação positiva.

Sinal cujas as amplitudes distanciadas de $\tau = 1$ e estão negativamente correlacionadas.

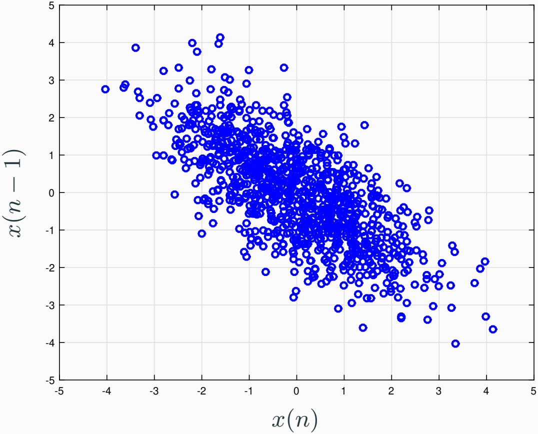


Figura 24: *Lagplot* com $\tau = 1$ e correlação negativa.

Sinal cujas as amplitudes distanciadas de $\tau = 1$ e estão descorrelacionadas.

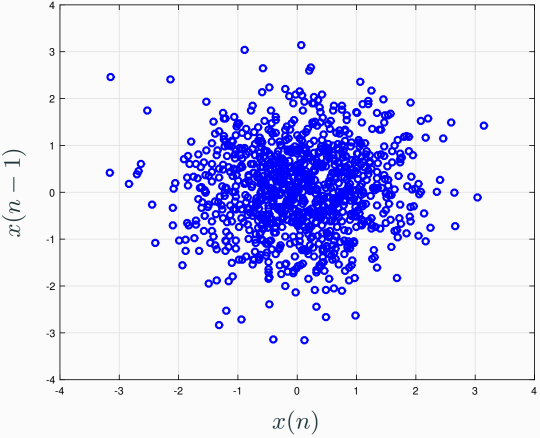


Figura 25: *Lagplot* com $\tau = 1$ e correlação nula.

Processo AR(1): $\boxed{x(n) = 0.6x(n-1) + v(n)}$

Os lagplots para $\tau = 1$ e $\tau = 2$ de um processo AR(1) são mostrados abaixo. O ruído é gaussiano com variância 1.

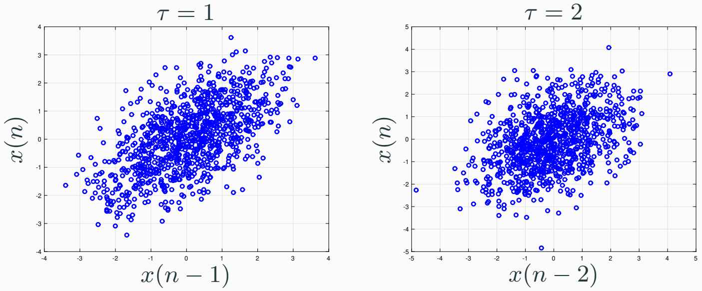


Figura 26: Lagplots para $\tau = 1$ e $\tau = 2$ do processo AR(1).

Processo AR(1): $\boxed{x(n) = 0.6x(n-1) + v(n)}$

Os lagplots para $\tau = 3$ e $\tau = 4$ de um processo AR(1) são mostrados abaixo. O ruído é gaussiano com variância 1.

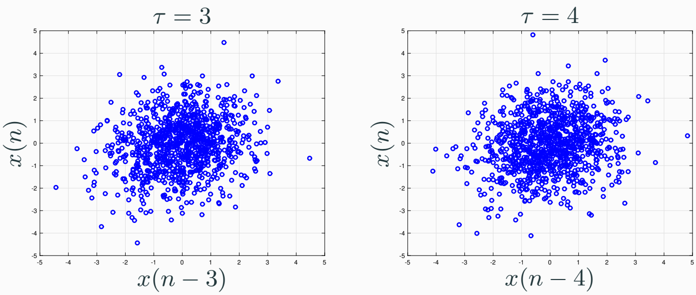


Figura 27: Lagplots para $\tau = 3$ e $\tau = 4$ do processo AR(1).

Processo de ruído branco

Um processo denominado como sendo **ruído branco** é um processo que possui as seguintes características:

- Amostras de diferentes instantes de tempo são independentes e identicamente distribuídas
- Amostras do mesmo instante de tempo possuem distribuição gaussiana de média zero e variância σ_X^2

Com isso, a função de autocorrelação do sinal é dada como

$$\boxed{r_X(\tau) = \begin{cases} \sigma_X^2, & \text{para } \tau = 0, \\ 0, & \text{para } \tau \neq 0 \end{cases}} \tag{93}$$

Processo de ruído branco - cont.

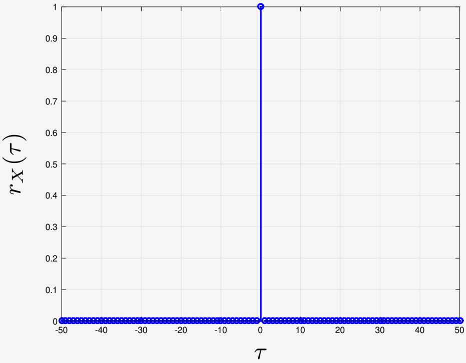


Figura 28: Autocorrelação normalizada teórica do processo de ruído branco.

Processo de ruído branco - cont.

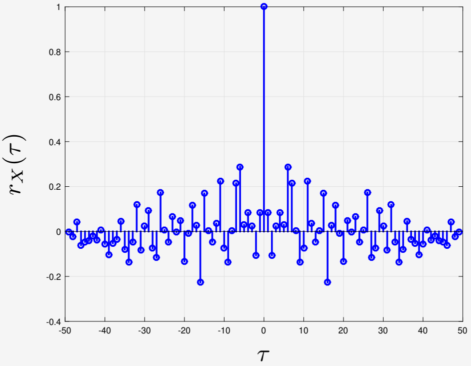


Figura 29: Autocorrelação normalizada estimada (amostral) do processo de ruído branco.

Alguns processos estocásticos usuais - cont.

Sequência binária aleatória

É um processo discreto que é típico na caracterização em sistemas de comunicações. Utilizamos uma sequência de bits ('0' e '1') onde cada os bits possuem probabilidade de ocorrência p e $1 - p$, respectivamente.

Cada sequência então possui as mesmas características estatísticas mas são potencialmente diferentes.

Sequência binária aleatória - cont.

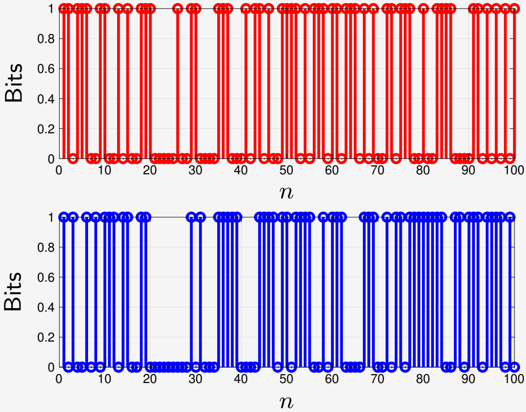


Figura 30: Sequências binárias aleatórias $\left(p = \frac{1}{2}\right)$.

Sequência binária aleatória - cont.

Uma sequência binária aleatória é também chamada de **processo aleatório de Bernoulli**. Assim, temos as seguintes características para este processo:

- Média: $\mu(n) = p$
- Variância: $\sigma^2(n) = p(1-p)$
- Autocovariância: $c_X(\tau) = \sigma^2\delta(\tau)$

Processo de Poisson

Outro processo importante, e que serve de base para outros tipos, é o **processo de Poisson**.

Imagine uma situação onde eventos ocorrem em instantes de tempo aleatórios, com uma taxa média de λ eventos por segundo. Um modelo de chegada de pacotes em um servidor de internet.

Seja $N(t)$ o número de ocorrências do evento no intervalo $[0, t]$. Este é então um processo *não-decrescente, de valores inteiros e contínuo no tempo.*

Processo de Poisson - cont.

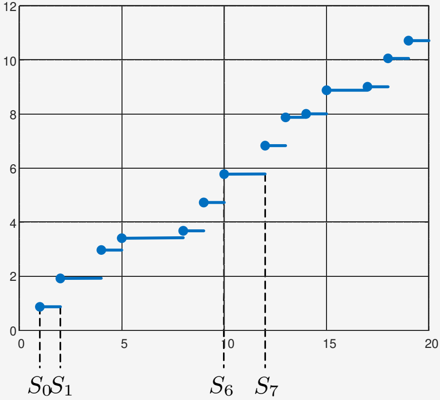


Figura 31: Realização de um processo de Poisson.

Processo de Poisson - cont.

Denotando os tempos de ocorrência dos eventos por $S_1, S_2, \ldots$ temos que o j-ésimo intervalo entre eventos é definido por $X_j = S_j - S_{j-1}$.

Suponha que o intervalo $[0, t]$ é dividido em n subintervalos pequenos de tal forma que as seguintes condições ocorrem:

① A probabilidade da ocorrência de um ou mais eventos em um subintervalo é desprezível comparado com a probabilidade de observar um ou nenhum evento;

② A ocorrência ou não de um evento em um subintervalo é independente do resultado em outros subintervalos.

Alguns processos estocásticos usuais - cont.

Processo de Poisson - cont.

A primeira hipótese implica que o resultado em cada subintervalo pode ser visto como uma amostra de Bernoulli. E a segunda hipótese implica que estas amostras de Bernoulli são independentes. As duas juntas implicam que o processo $N(t)$ pode ser aproximado como um *processo de contagem binomial*.

Processo de Poisson - cont.

Se a probabilidade de ocorrência de um evento em cada subintervalo é p então o número esperado (médio) no intervalo $[0, t]$ é np

Como os eventos ocorrem a uma taxa de λ eventos por segundo, o número médio de eventos no intervalo $[0, t]$ é λt.

Assim, nós devemos ter

$$\lambda t = np \tag{94}$$

Processo de Poisson - cont.

Se fizermos $n \to \infty$ e $p \to 0$ enquanto $\lambda t = np$ então pode-se mostrar que a distribuição binomial aproxima uma distribuição de Poisson com parâmetro λt.

Assim, pode-se concluir que o número de ocorrências $N(t)$ no intervalo $[0, t]$ tem uma distribuição de Poisson com média λt:

$$\boxed{\Pr[N(t) = k] = \frac{(\lambda t)^k}{k!} \exp(-\lambda t), \quad k = 0, 1, \ldots} \tag{95}$$

Processo de Poisson - cont.

Por esta razão $N(t)$ é chamado de **Processo de Poisson**.

O processo tem as seguintes características:

1. Média: $\mu(t) = \lambda t$
2. Variância: $\sigma^2(t) = \lambda t$
3. Autocovariância: $c_X(t_1, t_2) = \lambda t_1$

Processo de Poisson - cont.

Considere o tempo T entre ocorrências de eventos em um processo de Poisson. Suponha novamente que o intervalo de tempo $[0, t]$ é dividido em n subintervalos.

A probabilidade de que o tempo entre eventos T excede t segundos é equivalente à não ocorrência de eventos em t segundos, ou seja:

$$\begin{aligned}\Pr[T > t] &= \Pr[\text{sem eventos em } t \text{ segundos}] \\ &= (1-p)^n \\ &= \left(1 - \frac{\lambda t}{n}\right)^n \\ &\to \exp(-\lambda t)\end{aligned} \tag{96}$$

Alguns processos estocásticos usuais - cont.

Processo de Poisson - cont.

A equação (96) implica que a variável aleatória T segue uma distribuição exponencial com parâmetro λ.

Uma vez que os tempos entre chegadas são v.a. independentes, isto implica que a sequência de tempos entre eventos em um processo de Poisson é composta de v.a. independentes!

Assim, conclui-se que os tempos entre eventos em um processo de Poisson formam uma sequência iid de v.a. exponenciais com média $1/\lambda$.

Processo de Poisson - cont.

Uma outra quantidade de interesse é o tempo S_n no qual o n-ésimo evento ocorre em um processo de Poisson.

Seja T_j denotar os intervalos de tempo exponenciais iid. Com isso temos:

$$S_n = T_1 + T_2 + \cdots + T_n. \tag{97}$$

E S_n é uma v.a. com distribuição de Erlang, dada como

$$p_{S_n}(y) = \frac{(\lambda y)^{n-1}}{(n-1)!} \lambda \cdot \exp(-\lambda y), \quad y \geq 0. \tag{98}$$

Sinal telegráfico

Considere um processo aleatório $X(t)$ que assume os valores ± 1. Suponha que $X(0) = +1$ ou -1 com probabilidade $1/2$ e suponha também que $X(t)$ muda a polarização com cada ocorrência de um evento num processo de Poisson com taxa α.

Sinal telegráfico - cont.

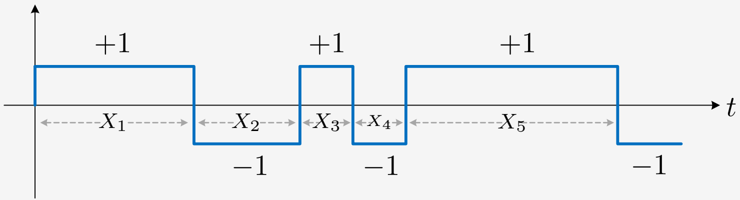


Figura 32: Realização de um sinal telegráfico. Tempos entre transições X_j são v.a. exponenciais iid.

Sinal telegráfico - cont.

- O sinal telegráfico aleatório é igualmente provável de ser ± 1 em qualquer tempo $t > 0$;
- A média e a variância do processo são dados por

$$\begin{aligned} \mu(t) &= 0 \\ \sigma^2(t) &= 1 \end{aligned} \tag{99}$$

- A autocovariância de $X(t)$ é dada por

$$\boxed{c_X(t_1, t_2) = \exp\left(-2\alpha|t_2 - t_1|\right)} \tag{100}$$

Alguns processos estocásticos usuais - cont.

Sinal telegráfico - cont.

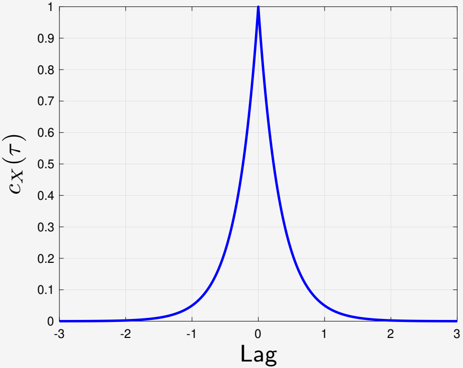
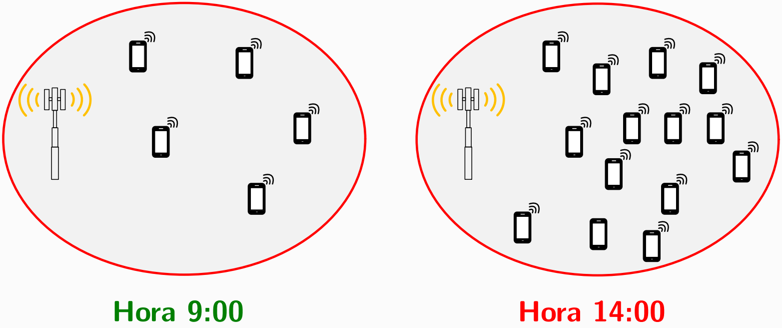


Figura 33: Autocovariância sinal telegráfico aleatório $(\lambda = 1.5)$.

Cadeias de Markov

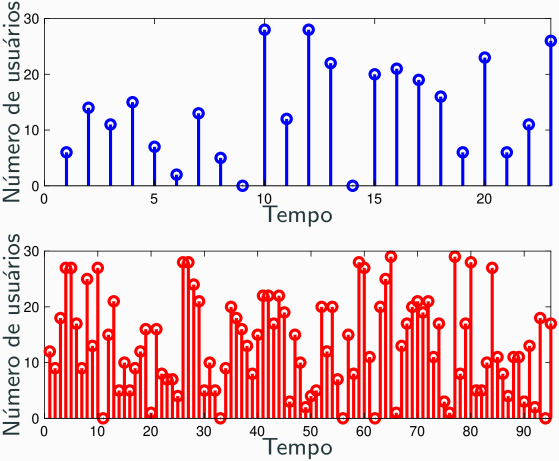
Motivação

Rede celular: número de usuários ativos em dado momento.



Motivação - cont.

Rede celular: número de usuários ativos em dado momento.



Definição

Um processo estocástico (aleatório) $X(t)$ é um *processo de Markov* se o futuro do processo **dado** o presente é independente do passado. Pode ser visto como uma generalização dos experimentos *independentes*.

? Mas o que significa isso?

- Qual a relação estatística que podemos extrair desta definição?

Para tempos arbitrários $t_0 < t_1 < \cdots < t_k < t_{k+1}$

Processo discreto

$$
\begin{aligned}
&\Pr[X(t_{k+1}) = x_{k+1} | X(t_k) = x_k, \ldots, X(t_0) = x_0] \\
&= \Pr[X(t_{k+1}) = x_{k+1} | X(t_k) = x_k],
\end{aligned} \tag{101}
$$

Processo contínuo

$$
\begin{aligned}
&\Pr[a < X(t_{k+1}) \leq b | X(t_k) = x_k, \ldots, X(t_0) = x_0] \\
&= \Pr[a < X(t_{k+1}) \leq b | X(t_k) = x_k],
\end{aligned} \tag{102}
$$

Processos de Markov - cont.

Os processos de Markov podem ser classificados não só pelo seu parâmetro, contínuo ou discreto, mas também pelo espaço dos estados (conjunto dos possíveis valores do processo estocástico), que também pode ser contínuo ou discreto. Os processos de Markov cujo espaço de estados é discreto, dizem-se **Cadeias de Markov**.

O caso que nos vai interessar será o das Cadeias de Markov de parâmetro discreto.

Propriedade markoviana

As Equações (101) e (102) caracterizam a **propriedade markoviana**.

Assim, podemos definir uma **cadeia de Markov** como sendo um processo estocástico que satisfaz as seguintes condições:

1. o parâmetro k é discreto
2. o espaço de estados E é discreto (finito ou infinito enumerável)
3. o estado inicial do processo ou do espaço de estados é conhecido
4. vale a **propriedade markoviana**
5. vale a *propriedade da estacionariedade*

Propriedade markoviana - cont.

A propriedade markoviana nos diz que as probabilidade (cumulativas e densidades também) de todos os estados futuros X_j $(j > k)$ dependem somente do estado atual X_k, ou seja, não dependem dos estados anteriores $X_0, X_1, \ldots, X_{k-1}$.

> *O estado "futuro" do sistema depende do "presente" mas não do "passado".*

Relembrando a propriedade markoviana (caso discreto), temos

$$\begin{aligned}\Pr[X(t_{k+1}) = x_{k+1} | X(t_k) = x_k, \ldots, X(t_0) = x_0] \\ = \Pr[X(t_{k+1}) = x_{k+1} | X(t_k) = x_k],\end{aligned} \tag{103}$$

Vamos então denotar os valores para o processo estocástico nos instantes determinados como estados.

Assim, vamos associar os valores $x_0, x_1, \ldots, x_k$ a um conjunto de estados denotados por $s_0, s_1, \ldots, s_k$.

Probabilidades de transição - cont.

Assim, de forma geral, a função da probabilidade de transição pode ser escrita como:

Função de probabilidade de transição

$$p_{ij}(m,n) = \Pr(X_n = s_j | X_m = s_i) \tag{104}$$

ou seja, a probabilidade do sistema estar no estado s_k no momento n dado que no momento m esteve no estado s_j.

Probabilidades de transição - cont.

Caso de interesse: Cadeia de Markov homogênea

- Cadeia de Markov é dita ser homogênea ou estacionária se $p_{j,k}(m,n)$ depende apenas da diferença $n-m$.
- Chamamos então função de **probabilidade de transição em n passos** da Cadeia de Markov homogênea a função

$$\begin{aligned} p_{ij}(n) &= \Pr(X_{n+t} = s_j | X_t = s_i) \\ &= \Pr(X_n = s_j | X_0 = s_i), \quad \forall t \geq 0, t \in \mathbb{N}. \end{aligned} \tag{105}$$

Probabilidades de transição - cont.

Uma transição de transição particularmente importante em uma cadeia de Markov homogênea é a transição de um passo, dada por

Probabilidade de transição de um passo

$$\begin{aligned} p_{ij}(1) &= \Pr(X_{n+1} = s_j | X_n = s_i) \\ &= \Pr(X_1 = s_j | X_0 = s_i), \quad \forall t \geq 0, t \in \mathbb{N}. \end{aligned} \tag{106}$$

Matriz de probabilidade de transição

Nota-se que as Equações (105) e (106) dependem apenas de duas informações: do estado inicial (i) e estado final (j).

Assim, uma representação natural para tais probabilidades de forma mais compacta é também uma representação de apenas duas coordenadas, o que nos leva a uma representação matricial.

Isso nos leva à definição das matrizes de probabilidade de transição, na sequência.

Matriz de probabilidade de transição - cont.

Matrizes de probabilidade de transição

$$\boxed{\mathbf{P}(n) = [p_{ij}(n)]} \tag{107}$$

para transição de n passos, e

$$\boxed{\mathbf{P}(1) = [p_{ij}(1)]} \tag{108}$$

para transição de um passo. Usualmente, a matriz de transição de um passo é também denotada apenas como $\mathbf{P}$.

Matriz de probabilidade de transição - cont.

As matrizes de probabilidade de transição não são matrizes quaisquer e possuem algumas condições:

✎ **Elementos são probabilidades**

$$p_{ij}(n) \geq 0, \quad \forall i, j \tag{109}$$

✏ **Soma dos eventos possíveis**

$$\sum_j p_{ij}(n) = 1, \quad \forall i, j \tag{110}$$

Matriz de probabilidade de transição - cont.

A matriz de transição (de um passo) tem então a seguinte estrutura geral:

$$\mathbf{P} = \begin{array}{c c} & \begin{array}{cccc} \color{red}{s_1} & \color{red}{s_2} & \color{red}{\cdots} & \color{red}{s_M} \end{array} \\ \begin{array}{c} \color{blue}{s_1} \\ \color{blue}{s_2} \\ \color{blue}{\vdots} \\ \color{blue}{s_M} \end{array} & \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & a_{24} \\ \vdots & \cdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{pmatrix} \end{array} \tag{111}$$

Escolha: linhas/colunas correspondem ao estado inicial/final. Ao longo dos slides iremos usar as linhas como o estado de origem e as colunas como estados destino.

? Mas o que a matriz de transição revela?

São as possibilidades do sistema (cadeia de Markov) *transitar* entre diferentes estados. E isso é muito útil para entendermos e calcularmos quais são os estados mais prováveis e possíveis.

Outras probabilidades em cadeias de Markov

- Outras probabilidades importantes em cadeias de Markov
- Probabilidade (não condicional) de no instante n o sistema estar no estado s_j, ou seja

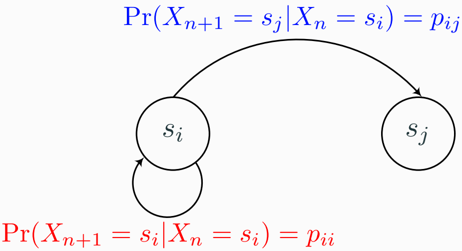
$$\boxed{p_j(n) = \Pr(X_n = s_j)} \tag{112}$$

- Vetor de probabilidades: agrupa as probabilidades para os diversos estados de uma cadeia de Markov, em um dado instante n,

$$\boxed{\mathbf{p}(n) = \begin{bmatrix} p_0(n) & p_1(n) & \cdots & p_M(n) \end{bmatrix}} \tag{113}$$

Diagramas de cadeias de Markov

- Uma ferramenta útil em cadeias de Markov discretas e homogêneas (ou estacionárias) é a ilustração da mesma como um grafo orientado.
- Como para as cadeias de Markov estacionárias, as probabilidades só dependem da diferença entre os estados nas quais ela se encontrava, tais probabilidades são constantes.
- Com isso, temos a seguinte possibilidade de representação



Diagramas de cadeias de Markov - cont.

Exemplo

Suponha que em *Nárnia* (ou *Terra Média*), nunca há dois dias ensolarados consecutivos.

- Se em um dia qualquer faz sol, no dia seguinte pode tanto nevar quanto chover.
- Se chover, metade das vezes continua chovendo no dia seguinte e nas outras ocasiões pode tanto fazer sol ou nevar.
- Se nevar, apenas metade das vezes o dia seguinte também neva.

Queremos:

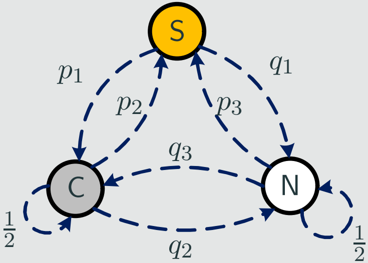
(a) Representar graficamente a Cadeia de Markov;

(b) Construir sua matriz de probabilidades de transição;

(c) Determinar a probabilidade de nevar daqui a dois dias?

Diagramas de cadeias de Markov - cont.

Exemplo - cont.



$$\mathbf{P} = \begin{pmatrix} 0 & p_1 & q_1 \\ p_2 & 1/2 & q_2 \\ p_3 & q_3 & 1/2 \end{pmatrix}$$

? Mas.... e qual a probabilidade de nevar daqui a dois dias?

Equações de Chapman-Kolmogorov

Como descrito, elemento $p_{ij}(n)$ da matriz $\mathbf{P}(n)$ representa a probabilidade de que o processo, iniciado no estado s_i, estará no estado s_j, depois de n passos.

Uma questão importante então é a consideração de sair, por exemplo, do estado s_i no instante $t = 0$, passando pelo estado s_k no instante $t = m$ e atingindo o estado s_j no instante $t = n + m$.

A resposta a esta questão leva às **equações de Chapman–Kolmogorov**.

Calculando a probabilidade de transição de n-passos...

Sabemos que a propriedade markoviana diz que:

$$\begin{aligned}\Pr[X(t_{k+1}) =& s_{k+1}|X(t_k) = s_k, \ldots, X(t_1) = s_1] = \\ &= \Pr[X(t_{k+1}) = s_{k+1}|X(t_k) = s_k]\end{aligned} \tag{114}$$

Da definição de probabilidade condicional que, para três eventos, temos:

$$\begin{aligned}&\Pr[X(t_3) = s_3|X(t_2) = s_2, X(t_1) = s_1] = \\ &\qquad = \frac{\Pr[X(t_3) = s_3, X(t_2) = s_2, X(t_1) = s_1]}{\Pr[X(t_2) = s_2, X(t_1) = s_1]}\end{aligned} \tag{115}$$

Equações de Chapman-Kolmogorov - cont.

Podemos ainda escrever, para dois eventos, que

$$\Pr[X(t_2) = s_2 | X(t_1) = s_1] = \frac{\Pr[X(t_2) = s_2, X(t_1) = s_1]}{\Pr[X(t_1) = s_1]} \quad (116)$$

Estamos interessados em calcular a probabilidade de transição entre os estados i e j de uma Cadeia de Markov em n passos, ou seja, calcular:

$$p_{ij}(n) = \Pr[X_{n+k} = s_j | X_k = s_i] = \Pr[X_n = s_j | X_0 = s_i]. \quad (117)$$

Equações de Chapman-Kolmogorov - cont.

Vamos iniciar com dois passos. Temos então o seguinte:

$$\begin{aligned} p_{ij}(2) &= \Pr[X_2 = s_j, X_1 = s_k, X_0 = s_i] \\ &= \frac{\Pr[X_2 = s_j, X_1 = s_k | X_0 = s_i]}{\Pr[X_0 = s_i]} \end{aligned} \tag{118}$$

E utilizando (115) na expressão em (118) temos que:

$$\begin{aligned} p_{ij}(2) &= \frac{\Pr[X_2 = s_j | X_1 = s_k] \cdot \Pr[X_1 = s_k | X_0 = s_i] \cdot \Pr[X_0 = s_i]}{\Pr[X_0 = s_i]} \\ &= \Pr[X_2 = s_j | X_1 = s_k] \cdot \Pr[X_1 = s_k | X_0 = s_i] \\ &= p_{kj}(1) \cdot p_{ik}(1) \end{aligned} \tag{119}$$

Equações de Chapman-Kolmogorov - cont.

Para todas as variações dos possíveis k estados que podem servir de caminho parcial entre i e j teremos a soma das probabilidades de tais caminhos, assim:

$$p_{ij}(2) = \sum_{k} p_{ik}(1) \cdot p_{kj}(1) \tag{120}$$

E generalizando para um número de passos $m + n$, feito através do uso de um maior número de expressões de eventos condicionados e do uso da propriedade markoviana dada em (114), temos o seguinte resultado:

$$\boxed{p_{ij}(m+n) = \sum_{k} p_{ik}(m) \cdot p_{kj}(n)} \tag{121}$$

Equações de Chapman-Kolmogorov - cont.

A Equação (121) (descrita novamente abaixo) é chamada de equação de **Chapman-Kolmogorov** e é um importante resultado de probabilidade para o estudo de Cadeias de Markov e cálculo de probabilidades de transição.

$$\boxed{p_{ij}(m+n) = \sum_{k} p_{ik}(m) \cdot p_{kj}(n)}$$

No caso de termos uma cadeia de Markov estacionária temos que as probabilidades de transição entre quaisquer dois estados, para dois instantes consecutivos, são constantes.

De acordo com a Equação (121) podemos escrever então a seguinte relação para a matriz de probabilidade de transição:

$$\mathbf{P}(n+m) = \{p_{ij}(n+m)\} \tag{122}$$

Assim, a matriz de probabilidade de transição para n passos, representada por $\mathbf{P}(n)$, tem a seguinte propriedade:

$$\boxed{\begin{aligned} \mathbf{P}(n+m) &= \mathbf{P}(n)\cdot\mathbf{P}(m) \\ \mathbf{P}(n) &= \mathbf{P}^n = \prod_{i=1}^{n}\mathbf{P} \end{aligned}} \tag{123}$$

Vetor de probabilidades iniciais

Em muitas situações se desconhece o estado da cadeia de Markov no instante inicial.

Para $i = 1, 2, \ldots, M$, consideramos p_i como a probabilidade de que o processo esteja no estado s_i no instante inicial.

Nestes casos, se associa probabilidades também para os estados da cadeia, no instante inicial. Assim, define-se o **vetor de probabilidades iniciais** $\mathbf{p}(0) = \begin{bmatrix} p_1(0) & \cdots & p_M(0) \end{bmatrix}$, cujos elementos do vetor são definidos segundo:

$$\begin{aligned} p_i(0) &\geq 0 \\ \sum_{i}^{M} p_i(0) &= 1 \end{aligned} \tag{124}$$

Probabilidade de estados

Agora vamos considerar as probabilidades dos estados em um instante de tempo n. Seja $\mathbf{p}(n)$ o vetor de probabilidades no instante n.

A probabilidade $p_j(n)$ é relacionada ao vetor $\mathbf{p}(n-1)$ pela seguinte equação:

$$\begin{aligned} p_j(n) &= \sum_i \Pr[X_n = s_j | X_{n-1} = s_i] \cdot \Pr[X_{n-1} = s_i] \\ &= \sum_i p_{ij} \cdot p_i(n-1) \end{aligned} \tag{125}$$

Probabilidade de estados - cont.

A Equação (125) diz que o vetor de estados no instante n $\mathbf{p}(n)$ é obtido multiplicando o vetor $\mathbf{p}(n-1)$ pela matriz $\mathbf{P}$, ou seja:

$$\boxed{\mathbf{p}(n) = \mathbf{p}(n-1) \cdot \mathbf{P}} \tag{126}$$

Probabilidade de estados - cont.

De forma similar, podemos escrever a dependência de $p_j(n)$ com o vetor de estados iniciais $\mathbf{p}(0)$ como

$$\begin{aligned} p_j(n) &= \sum_i \Pr[X_n = s_j | X_0 = s_i] \cdot \Pr[X_0 = s_i] \\ &= \sum_i p_{ij}(n) p_i(0) \end{aligned} \tag{127}$$

E em notação matricial temos

$$\boxed{\mathbf{p}(n) = \mathbf{p}(0)\mathbf{P}(n) = \mathbf{p}(0) \cdot \mathbf{P}^n} \tag{128}$$

Probabilidade estacionária

A partir da Equação (128), notamos que o produtório da matriz de probabilidades de transição tem um papel essencial na determinação dos estados atingíveis da cadeia de Markov.

? Surge então a questão: os valores de $\mathbf{p}(n)$ convergem para algum valor limite quando $n \to \infty$

A estas probabilidades de *equilíbrio* são chamadas probabilidades de estado estacionárias.

Matematicamente, o que temos é que, quando $n \to \infty$, a matriz de probabilidade de se aproxima de uma matriz na qual as linhas tem a mesma distribuição, ou seja

$$p_{ij}(n) \to \pi_j, \quad \forall i \tag{129}$$

Para escrever a Equação (129) em notação matricial podemos ter:

$$\mathbf{P}^n \to \mathbf{1}\boldsymbol{\pi}, \tag{130}$$

em que $\mathbf{1}$ é um vetor coluna com todas as posições igual a 1, ou seja, $\mathbf{1}^T = \begin{bmatrix} 1 & 1 & \dots \end{bmatrix}$ e $\boldsymbol{\pi} = \begin{bmatrix} \pi_0 & \pi_1 & \dots \end{bmatrix}$.

Probabilidade estacionária - cont.

Da Equação (127), a convergência de de $\mathbf{P}^n$ implica a convergência dos seguintes estados:

$$\boxed{p_j(n) = \sum_i p_{ij}(n) \cdot p_i(0) \rightarrow \sum_i \pi_i \cdot p_i(0) = \pi_j} \qquad (131)$$

Quando a Equação (131) é válida dizemos que o sistema (cadeia de Markov) atinge o "equilíbrio" ou "estado permanente".

Probabilidade estacionária - cont.

Podemos encontrar a distribuição de $\boldsymbol{\pi}$ uma vez que notamos que para $n \to \infty$, temos $p_j(n) \to \pi_j$ e também $p_j(n-1) \to \pi_j$.

Com isso, podemos escrever a Equação (125) como

$$\pi_j = \sum_i p_{ij}\pi_i, \tag{132}$$

o que em notação matricial é escrito como

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}. \tag{133}$$

Probabilidade estacionária - cont.

Também sabe-se que

$$\sum_i \pi_i = 1, \tag{134}$$

já que os elementos do vetor são probabilidades.

Chamamos o vetor $\boldsymbol{\pi}$ a **distribuição de probabilidade de estado estacionário** da cadeia de Markov.

Se iniciarmos a cadeia de Markov com o vetor de estados iniciais $\mathbf{p}(0) = \boldsymbol{\pi}$ então, pelas Equações (128) e (133), temos que

$$\mathbf{p}(n) = \boldsymbol{\pi}\mathbf{P}^n = \boldsymbol{\pi} \tag{135}$$

Isso significa que

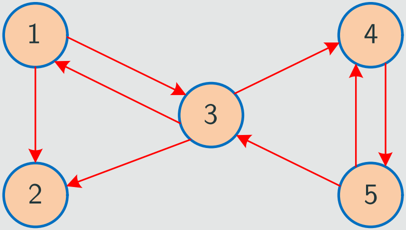
1. Probabilidades são independentes da escolha do instante inicial, e;
2. O processo é estacionário.

Exemplo

Escolha de páginas visitadas

Um navegador da internet procura páginas em um universo de 5 páginas, conforme a figura abaixo. O navegador seleciona a próxima página a ser vista selecionando com igual probabilidade as páginas apontadas pela página atual. Se uma página não tem link de saída (por exemplo, a página 2), o navegador então seleciona qualquer das páginas do universo com igual probabilidade.

Encontrar a probabilidade do navegador mostrar a página i.

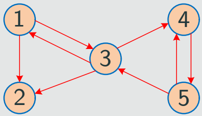


Exemplo - cont.

Escolha de páginas visitadas - cont.

Podemos modelar o comportamento descrito como uma cadeia de Markov onde o estado representa a página atualmente visualizada. Se a página atual aponta para k páginas, então a próxima página é selecionada daquele grupo com probabilidade $\frac{1}{k}$. Se a página atual não aponta para nenhuma outra, então a próxima página pode ser qualquer uma das 5 páginas com probabilidade $\frac{1}{5}$. A matriz de probabilidades de transição é escrita então como

$$\mathbf{P} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{bmatrix}$$



Exemplo - cont.

Escolha de páginas visitadas - cont.

Podemos então usar a Eq. (128) para encontrar os estados para cada instante.

Para tentar ver quando o sistema se estabiliza, podemos usar algum software de simulação para calcular $\mathbf{P}^n$ para $n = 50$, por exemplo. Neste caso, estamos avaliando a estabilização do processo e as probabilidades de estado estacionário. Para este caso específico, teríamos então

$$\mathbf{p}(50) = \begin{bmatrix} 0.1229 & 0.18447 & 0.25825 & 0.12298 & 0.31974 \end{bmatrix}$$

Mas podemos usar a Eq. (135) para calcular as probabilidades de estado estacionário exatas.

Exemplo - cont.

Escolha de páginas visitadas - cont.

O modelo de visualização aleatória por meio de uma cadeia de Markov forma a base do algoritmo do **PageRank** que foi introduzido pelo Google e que classifica a importância de uma página na Web.

Classes de estados

Os estados de uma cadeia de Markov podem ser divididas em uma ou mais classes separadas e estas classes podem ser de vários tipos.

Estado acessível - dizemos que o estado j é acessível a partir do estado i se, para algum $n \geq 0$, $p_{ij}(n) > 0$. Ou seja, que ha uma sequência de transições de i para j que possui probabilidade não-nula.

Estado comunicável - dizemos que os estados i e j são comunicáveis se eles são acessíveis de um para o outro. Escreve-se então $i \leftrightarrow j$.

Classes de estados - cont.

Se o estado i é comunicável com o estado j e este é comunicável com k, então o estado i é comunicável com k. Ou seja, $i \leftrightarrow j$ e $j \leftrightarrow k$ então $i \leftrightarrow k$.

Diz-se que dois estados pertencem à mesma **classe** se eles são comunicáveis.

É preciso notar que duas classes de estados distintas devem ser disjuntas uma vez que a existência de um elemento comum implicaria que os estados das classes seriam comunicáveis.

Classes de estados - cont.

> Os estados de uma cadeia de Markov consistem de uma ou mais classes (comunicáveis) disjuntas.

Uma cadeia de Markov que consiste de uma única classe é dita ser **irredutível**.

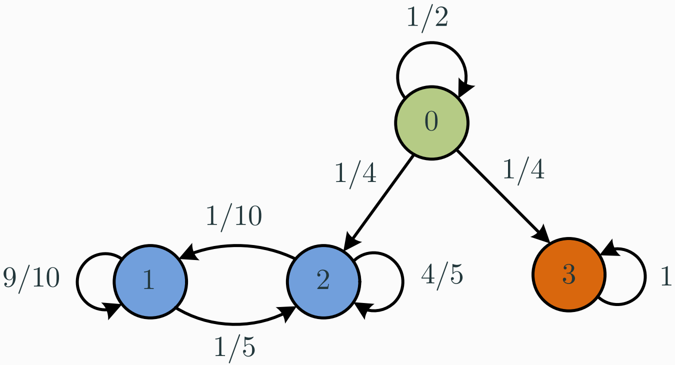


Figura 34: Três classes: $\{0\}, \{1, 2\}$ e $\{3\}$.

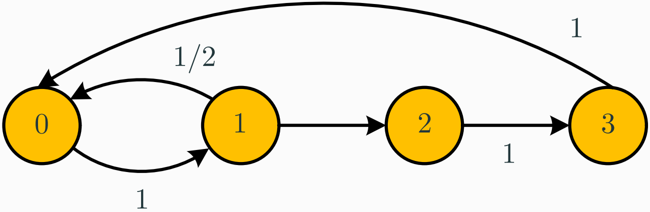


Figura 35: Uma classe - cadeia irredutível.

Propriedades de recorrência

Suponha uma cadeia de Markov no estado i. O estado i é dito ser **recorrente** se o processo retorna para o estado com probabilidade 1, ou seja

$$f_i = \Pr[\text{retornar alguma vez para estado } i] = 1 \tag{136}$$

O estado i é dito ser **transiente** se

$$f_i < 1 \tag{137}$$

Propriedades de recorrência - cont.

Recorrência e transiência são *propriedades de classe*:

- ✔ Se o estado i é recorrente (transiente) toda a sua classe é recorrente (transiente);
- ✔ Se uma cadeia de Markov é irredutível então todos os seus estados ou são transientes ou todos são recorrentes;
- ✔ Os estados de uma cadeia de Markov irredutível com estados finitos são todos recorrentes.

Propriedades de recorrência - cont.

Outro aspecto que pode ser analisado é que o diagrama de transição e as probabilidades de transição não nulas podem impor **periodicidade** na cadeia de Markov.

Dizemos que o estado i tem **período** d se ele apenas ocorre em instantes de tempo que são múltiplos de d.

Assim, também dizemos que um estado i é **aperiódico** se ele tem período $d = 1$.

Periodicidade é uma *propriedade de classe*.

- Todos os estados em uma classe têm o mesmo período;
- Uma cadeia de Markov irredutível é dita ser aperiódica se os estados na sua classe única tem período 1;
- Assim, uma cadeia de Markov irredutível é dita ser periódica ser seus estados possuem período $d > 1$.

Propriedades de recorrência - cont.

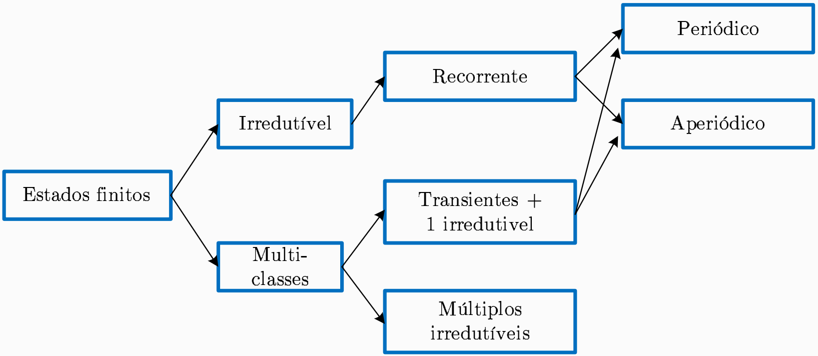


Figura 36: Possíveis estruturas para cadeias de Markov.

Um **processo de nascimento-e-morte** é uma cadeia de Markov nas quais as transições entre estados adjacentes ocorrem como representado na Figura 37.

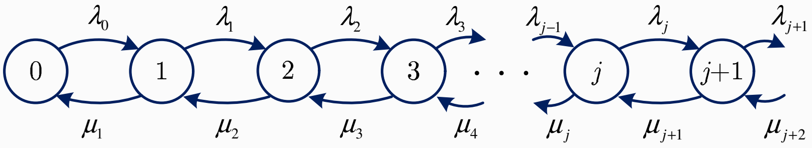


Figura 37: Diagrama de transição de um processo de nascimento-e-morte.

As **equações de equilíbrio** de um processo de nascimento-e-morte geral são dadas por:

$$\lambda_0 p_0 = \mu_1 p_1, \qquad j = 0 \tag{138a}$$

$$\lambda_j p_j - \mu_{j+1} p_{j+1} = \lambda_{j-1} p_{j-1} - \mu_j p_j, \qquad j = 1, 2, \ldots \tag{138b}$$

Com estas equações, para haver equilíbrio devemos ter

$$\lambda_{j-1} p_{j-1} - \mu_j p_j = \text{constante} \tag{139}$$

e pela equação para $j = 0$ temos que

$$\lambda_0 p_0 - \mu_1 p_1 = \text{constante} = 0 \tag{140}$$

Processo de nascimento-e-morte - cont.

Com isso, temos que

$$\lambda_{j-1}p_{j-1} - \mu_j p_j = 0 \tag{141}$$

o que implica em

$$p_j = r_j \cdot p_{j-1}, \quad j = 1, 2, \ldots \tag{142}$$

em que $r_j = \left(\frac{\lambda_{j-1}}{\mu_j}\right)$

E usando um argumento de indução simples podemos escrever

$$p_j = r_j r_{j-1} \ldots r_1 p_0, \quad j = 1, 2, \ldots \tag{143}$$

Se definirmos

$$R_j = r_j r_{j-1} \dots r_1 p_0 \quad \text{e} \quad R_0 = 1, \tag{144}$$

podemos encontrar p_0 como

$$\left(\sum_{j=0}^{\infty} R_j \right) p_0 = 1 \tag{145}$$

Se a série acima converge então a distribuição estacionária é dada por

$$\boxed{p_j = \frac{R_j}{\sum_{j=0}^{\infty} R_j}.} \tag{146}$$

Análise Espectral de Processos Estocásticos

Análise tempo × frequência

Em análise de sinais, uma importante ferramenta é a avaliação dos conteúdos de frequência que possuem informação relevante.

No caso dos sinais determinísticos, a ferramenta típica que é utilizada é a **transformada de Fourier** que pode ser escrita como:

$$\boxed{X(\omega) = \mathfrak{F}\{x(t)\} = \int_{-\infty}^{\infty} x(t) \cdot \exp(-j\omega t)\, dt,} \quad (147)$$

em que $\int_{-\infty}^{\infty} |x(t)| dt < \infty$ deve ser satisfeito para a transformação existir.

Problemas!

Para sinais aleatórios (estocásticos):

1. não conhecemos os valores dos mesmos pois eles possuem incerteza;
2. para a maioria das realizações dos processos, a condição de existência não é satisfeita;
3. mesmo quando uma realização que satisfaz a condição de existência, esta realização não representa todo o processo;
4. na maioria dos casos, as realizações de um processo possuem formas irregulares e não podem ser representadas na forma analítica.

Densidade espectral de potência

Para contornar estes problemas é que existe a **densidade espectral de potência (DEP)**. A DEP então considera a incerteza (aleatoriedade) dos processos e faz uma transformação tempo-frequência.

Mas como se define a DEP?

Como considerar a natureza estatística e uma representação geral do processo e não apenas de uma realização particular?

Densidade espectral de potência - cont.

Seja um processo estocástico $X(t)$ contínuo, estacionário e de média nula. A função **Densidade Espectral de Potência (DEP)** de $X(t)$ é definida como:

$$\begin{aligned} S_X(\omega) &= \mathfrak{F}\{r_X(\tau)\} \\ &= \int_{-\infty}^{\infty} r_X(\tau) \exp(-j\omega\tau)\, d\tau \end{aligned} \tag{148}$$

Com isso, a DEP é um indicador da distribuição da potência do sinal como uma função da freqüência. A Equação (148) é conhecida como relação de **Wiener-Khinchin**.

Densidade espectral de potência - cont.

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

$$r_X(\tau) = \mathfrak{F}^{-1}\{S_X(\omega\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(\omega)\exp(j\omega\tau)\ d\omega \qquad (149)$$

Densidade espectral de potência - cont.

Propriedades

❶ O valor médio quadrático de um processo WSS é dado por

$$\mathbb{E}\{X^2(t)\} = r_X(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(\omega)\, d\omega \tag{150}$$

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

$$S_X(\omega) \geq 0 \quad \text{para todo } \omega \tag{151}$$

Densidade espectral de potência - cont.

Propriedades - cont.

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

$$S_X(\omega) = S_X(-\omega) \tag{152}$$

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

$$S_X(0) = \int_{-\infty}^{\infty} r_X(\tau)\ d\tau \tag{153}$$

Exemplo

Considere o sinal aleatório dado por

$$X(t) = \sin(2\pi \cdot f_1 \cdot t) + A, \tag{154}$$

em que $A \sim N(0, \sigma^2)$. Uma realização de $X(t)$ para $f_1 = 150$ Hz e $\sigma^2 = 1$ é representada na Figura 38.

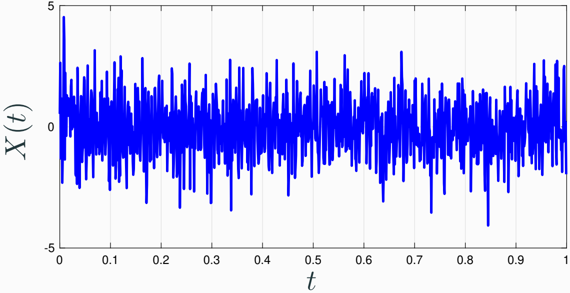


Figura 38: Sinal aleatório - senóide com ruído.

Exemplo - cont.

Do gráfico temporal é muito difícil ter alguma noção de quais frequências carregam informação.

Mas observarmos que a DEP de $X(t)$ é uma constante já que a correlação de um sinal de ruído branco é a função impulso. Além disso, a transformada de Fourier de uma função seno é uma função impulso na frequência de oscilação.

Finalmente, neste caso, a DEP seria uma combinação linear (soma) das representações em frequência do seno e do ruído branco.

Exemplo - cont.

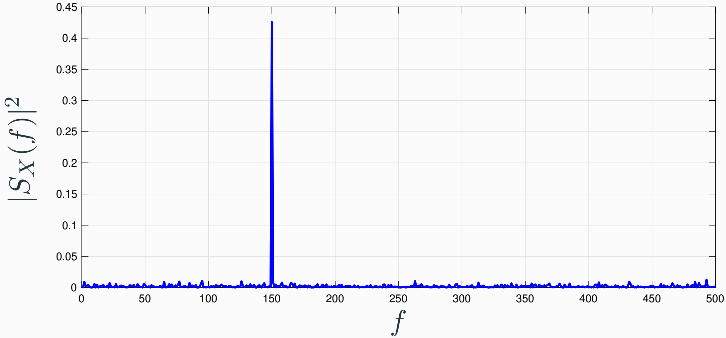


Figura 39: Representação da DEP. Frequência com mais informação bem visível.