Séries Temporais, Processos Estocásticos e Transformações de Variáveis

Prof. Caio Azevedo

Modelos para Séries Temporais

- Os modelos utilizados para descrever séries temporais são processos estocásticos, isto é, processos controlados por leis probabilísticas.
- Definição: Seja T um conjunto arbitrário. Um processo estocástico é uma família

$$Y = \{Y(t); t \in T\},$$

tal que, para cada $t \in T$, $Y(t)$ é uma variável aleatória (va ou v.a.).

- Como usual: $Y(t)$: variável aleatória e $y(t)$: valor observado da variável aleatória

Cont.

- Mais rigorosamente, $Y(t)$ é uma função de dois argumentos: $Y(t;\omega)$, para $t \in T$ e $\omega \in \Omega$.
- Para cada $t \in T$ fixado, $Y(t;\omega) \equiv Y(\omega)$ é uma variável aleatória definida sobre o espaço amostral Ω.
- Para cada $\omega \in \Omega$ fixado, $Y(t;\omega) = Y(t)$, é uma função de t, ou seja, uma realização ou trajetória do processo estocástico, ou ainda, uma série temporal.

Ilustração gráfica

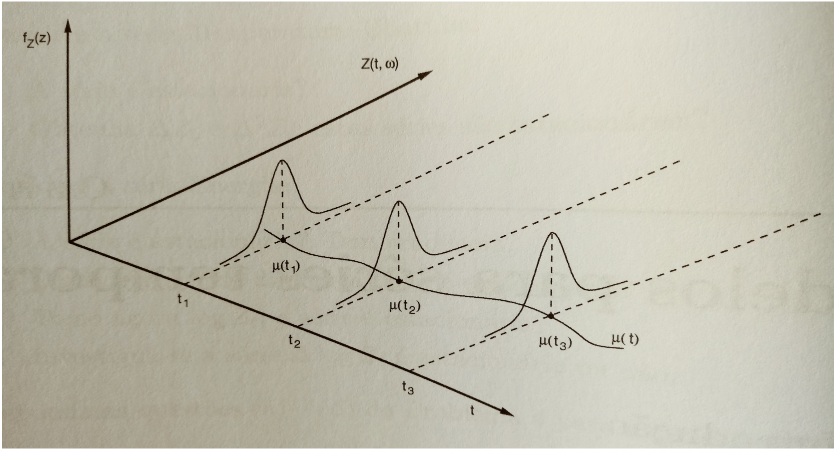
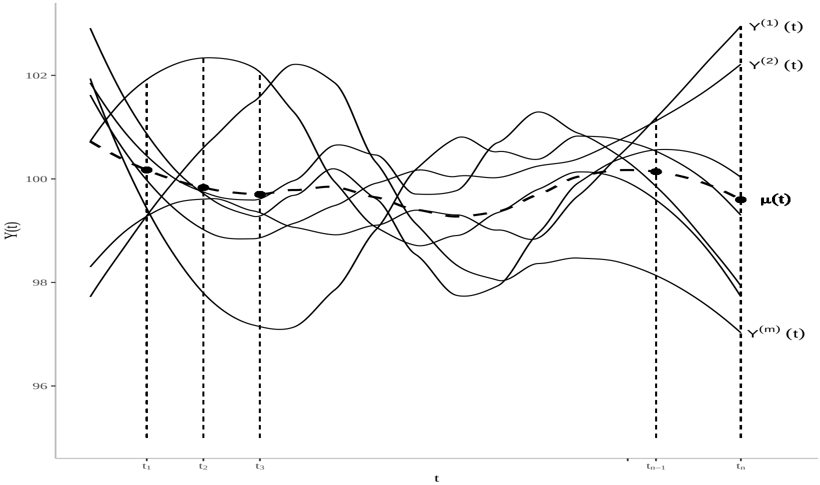


Figura: Um processo estocástico interpretado como uma família de variáveis aleatórias

Comentários (figura anterior)

- Olhando o gráfico em função de $t \times Z(t, \omega)$, para cada valor de ω, podemos traçar uma curva (representando um processo estocástico).
- Olhando o gráfico em função de $t \times f_Z(z)$, para cada valor de t, temos uma densidade ("aparentemente" Gaussiana).
- $\mu(t)$ seria a média do processo

Ilustração Gráfica 2



Comentários (figura anterior)

- Como na página 21 dos slides Introdução, podemos definir várias configurações de Processos Estocásticos (ST), dependendo dos objetivos/restrições :
  - Para $t_i, i = 1, 2, ..., n$ fixado. ST: por exemplo, os valores de $Y^{(j)}(t_4), j = 1, 2, .., m$.
  - Para $t_i, i = 1, 2, ..., n$ variando com em função de (j). ST: por exemplo, os valores de $\left(Y^{(2)}(t_1), Y^{(m-1)}(t_2), ..., Y^{(m-1)}(t_n)\right)$.
  - Para $j, i = 1, 2, ..., m$, considera-se o mesmo t_i. ST: por exemplo, os valores de $\left(Y^{(1)}(t_3), Y^{(2)}(t_3), ..., Y^{(m)}(t_3)\right)$.
  - Para $j, i = 1, 2, ..., m$, considera-se uma transformação de cada $Y^{(j)}(t_i)$. ST: por exemplo, tomar algum tipo de média de cada curva.

Cont.

- Ao se definir um processo estocástico é necessário introduzir três características:
  - O espaço onde está definido o processo.
  - O conjunto dos índices T.
  - A estrutura de dependência das v.a.'s, $Y(t), t \in T$.

Cont.

- O conjunto de valores $\{Y(t), t \in T\}$ é chamado de espaço de estados (S), do processo estocástico, e cada valor de $Y(t)$ é chamado de estado.
- O espaço de estados pode ser discreto ou contínuo:
  - No caso discreto $Y(t)$ pode representar uma contagem, como o número de transações de uma ação, durante um dia.
  - No caso contínuo, $Y(t)$ pode representar uma medida que varia continuamente, como o retorno de um ativo ou volume negociado, em cada dia.

Cont.

- O conjunto dos índices T também pode ser discreto ou contínuo.
  - Caso o conjunto seja finito ou enumerável, como por exemplo $\mathbb{Z}$, o processo diz-se com parâmetro discreto.
  - Caso o conjunto seja um intervalo dos $\mathbb{R}$, por exemplo, teremos um processo com parâmetro contínuo.

Cont.

- É de particular interesse a utilização das chamadas distribuições finito-dimensionais (DFD) (pois, na prática, temos um conjunto finito de observações de uma ST).
- As distribuições acima são uma forma alternativa de definirmos um processo estocástico $\{Y(t), t \in T\}$.
- Assim, dados $t_1, ..., t_k \in T$, então as DFD's correspondem as distribuições conjuntas de $(Y(t_1), ..., Y(t_k))'$, ou seja:

$$F(y_1, ..., y_k; t_1, ..., t_k) = P\left(Y(t_1) \leq y_1, ..., Y(t_k) \leq y_k\right)$$

Cont.

- Um problema é que na prática só dispomos (em geral) de uma única realização do processo. Assim, para que seja viável propor uma distribuição conjunta apropriada (segundo critérios estatísticos e em termos do problema), certas restrições tem de ser impostas.
- No caso dos processos estocásticos, as restrições que devem (podem) ser impostas são de dois tipos
  - Na heterogeneidade temporal (como as DFD's variam com o tempo).
  - Na memória do processo (estrutura de dependência).

Restrições na heterogeneidade temporal

- Pode-se assumir que a distribuição conjunta é invariante por translações. Ou seja, dado um subconjunto de índices de T, digamos $\{t_1 < t_2 < ... < t_m\}$, a restrição imposta à heterogeneidade temporal, em nosso caso, corresponde à assumir que as distribuições conjuntas de

$$\{Y(t_1), Y(t_2), ..., Y(t_m)\}$$

e

$$\{Y(t_1 + \tau), Y(t_2 + \tau), ..., Y(t_m + \tau)\}$$

- são idênticas $\forall\ \tau \in \mathbb{Z}^+$.

Restrições na heterogeneidade temporal

- Ou seja, $\forall\ \tau \in \mathbb{Z}^{+}$, temos que: (do slide anterior)

$$\begin{aligned} & P(Y(t_1) \leq y_1, ..., Y(t_k) \leq y_k) = \\ & P(Y(t_1 + \tau) \leq y_1, ..., Y(t_k + \tau) \leq y_k) \end{aligned} \quad (1)$$

- Tais processos são chamados de **estritamente estacionários**
- Em termos práticos é muito difícil ter-se e/ou conseguir verificar se um processo é estritamente estacionário.

Comentários

- Em geral, o que se faz é considerar restrições nos Momentos do processo Estocástico, i.e, $\mathcal{E}\left[Y^k(t)\right]$, $k \in \mathbb{Z}^+$, em particular, nos dois primeiros momentos, os quais são dados por:
  - Função média: $\mathcal{E}(Y(t)) = \mu(t)$.
  - Função de auto-covariância (facv):

$$\begin{aligned}\gamma(t_1, t_2) &= \mathcal{E}\left[(Y(t_1) - \mu(t_1))(Y(t_2) - \mu(t_2))\right] \\ &= \mathcal{E}\left[Y(t_1)Y(t_2)\right] - \mu(t_1)\mu(t_2)\end{aligned}$$

  $\forall t_1, t_2 \in T$.
  - Em particular, se $t_1 = t_2 = t$, então $\gamma(t,t) = \mathcal{E}\left[Y^2(t)\right] - \mu^2(t) = \mathcal{V}(Y(t)) = V(t) = \sigma^2(t)$ é a (função) variância.

Comentários

- A facv fornece a estrutura de dependência temporal do processo estocástico (Y). Contudo, assim como no caso usual probabilístico, a correlação é mais apropriada do que a covariância, para tal fim.
- Função de auto-correlação (fac):

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sigma(t_1)\sigma(t_2)},$$

em que $\sigma(.) = \sqrt{\sigma^2(.)}$.

Restrição na Memória do Processo

- Por outro lado, com relação a memória do processo estocástico ("grau e ordem de dependência"), um caminho simples seria assumir que o processo não tem memória. Isto é, que ele seja não (auto-) correlacionado ou independente.
- A restrição de independência é muito forte e pouco plausível, uma vez que séries temporais apresentam algum tipo de dependência temporal.
- Assim, uma forma mais apropriada é considerar que para instantes de tempo muito afastados a correlação seja nula. Dessa forma, o processo tem memória, mas esta vai diminuindo com o aumento dos intervalos entre os instantes de tempo.

Processos estacionários

- De modo semelhante ao contexto dos modelos de regressão (ME 613), o número de parâmetros tende a ser muito grande (maior do que o número de observações), caso restrições (modelagem) não sejam consideradas (heterogeneidade temporal e/ou memória do processo).
- Por exemplo, para t=1,2,...,n, teremos $y_1, ..., y_n$ (n observações) e, pelo menos n médias ($\mu(t)$) e n variâncias ($\sigma^2(t)$) para estimar.

Processos estacionários

- Portanto, faz-se mister reduzir o número de parâmetros para podermos realizar processos inferenciais de forma adequada.
- Uma das estruturas (hipóteses simplificadoras) mais comuns é a **estacionariedade** (que, à rigor, tem relação com a heterogeneidade temporal e a memória do processo).

Cont.

- Como dito anteriormente, a suposição de estacionariedade estrita (Equação (1)) é difícil de ser avaliada (ser apropriada).
- Uma alternativa, é considerar uma estrutura de estacionariedade menos rígida, restringindo somente alguns momentos.
- Consideraremos (ao menos na parte inicial do curso) ST que (mesmo que sob alguma transformação) sejam fracamente estacionárias.

Processo fracamente estacionário

- Um processo estocástico $\{Y(t), t = 1, 2, ...\}$, com segundo momento finito $\mathcal{E}\left\{Y^2(t)\right\} < \infty$ é dito ser fracamente estacionário (estacionário de segunda ordem, estacionário no sentido amplo ou estacionário em covariância) se, se somente se:
  1. $\mathcal{E}[Y(t)] = \mu$ (constante), $\forall t \in T$.
  2. $\mathcal{V}[Y(t)] = \mathcal{E}\left[(Y(t) - \mu)^2\right] = \sigma^2$ (constante), $\forall t \in T$.
  3. A função de autocovariância (facv) é tal que $(\forall t_1, t_2, \tau)$

$$
\begin{aligned}
\gamma(t_1, t_2) &= Cov\left[Y(t_1), Y(t_2)\right] = Cov\left[Y(t_1 + \tau), Y(t_2 + \tau)\right] \\
&= \gamma(t_1 + \tau, t_2 + \tau) = f(t_2 - t_1)
\end{aligned}
$$

Ou seja, $\forall t_1, t_2, \tau \in T$ depende apenas da distância entre as observações e não dos instantes em si.

Processo fracamente estacionário

- Com efeito, se $t_1 = t_2 + \tau$, temos que

$$\begin{aligned}\gamma(t_2+\tau, t_2) &= Cov\left[Y(t_2+\tau), Y(t_2)\right] = Cov\left[Y(\tau), Y(0)\right] \\ &= \gamma(\tau, 0) \equiv \gamma(\tau),\end{aligned}$$

pois $|(t_2+\tau) - t_2| = |\tau - 0| = |\tau|$.

Cont.

- Se $\mu = 0$, então $\gamma(\tau) = \mathcal{E}\left[Y(t)Y(t+\tau)\right]$.
- Assim, é possível provar que (exercício):
  1. $\gamma(0) > 0$.
  2. $\gamma(-\tau) = \gamma(\tau)$.
  3. $|\gamma(\tau)| \leq \gamma(0)$.
- Uma outra importância (além da modelagem da dependência temporal) da correta especificação da função média e da facv deve-se ao fato de que, se as distribuições finito-dimensionais de $Y(t)$ são normais (multivariadas), o processo estocástico será totalmente conhecido, em se conhecendo μ e γ.

Cont.

- No caso geral, em que $\mu \neq 0$, temos que:

$$\gamma(\tau) = Cov\left[Y(t), Y(t+\tau)\right] = \mathcal{E}\left[(Y(t)-\mu)(Y(t+\tau)-\mu)\right]$$

$\forall t, \tau \in T$, dependa somente de τ e não de t.

- Existem formas empíricas e inferenciais de verificar se uma ST é fracamente estacionária (veremos algumas ao longo do curso).
- Muitas das classes de modelos que serão vistas são apropriadas para ST fracamente estacionárias.

Cont.

- Se os dois primeiros momentos existirem, então

$$\text{Estacionariedade forte} \rightarrow \text{Estacionariedade fraca}$$

- A volta só vale se as distribuições finito-dimensionais de $Y(t)$ (ou seja, de $(Y(t_1), ..., Y(t_k))'$) forem normais (multivariadas) (Processo Gaussiano).
- Sob estacionariedade fraca ($\mathcal{V}[Y(t)] = \mathcal{V}[Y(t+\tau)]$), a fac torna-se

$$\rho(\tau) = \frac{Cov[Y(t), Y(t+\tau)]}{\sqrt{\mathcal{V}([Y(t)]}\sqrt{\mathcal{V}([Y(t+\tau)]}} = \frac{\gamma(\tau)}{\gamma(0)}$$

- A representação gráfica da expressão anterior (fac) é chamada de correlograma.

Cont.

- Como na prática, temos somente uma única realização de um processo estocástico (isto é, uma única série temporal observada), podemos calcular apenas as funções amostrais (estimadores/estimativas), considerando uma ST de tamanho n:
  - Função média amostral: $\widehat{\mathcal{E}}(Y(t)) = \widehat{\mu} = \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.
  - Função variância amostral: $\widehat{\mathcal{V}}(Y(t)) = \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$.
  - Função de auto-covariância amostral: $\widehat{\gamma}(\tau) = \frac{1}{n}\sum_{i=1}^{n-\tau}\left(Y_i - \overline{Y}\right)\left(Y_{i+\tau} - \overline{Y}\right)$.

Cont.

- Cont. (estimadores)
  - Função de autocorrelação amostral:

$$\widehat{\rho}(\tau) = \frac{\widehat{\gamma}(\tau)}{\widehat{\gamma}(0)} = \frac{\frac{1}{n}\sum_{i=1}^{n-\tau}\left(Y_i - \overline{Y}\right)\left(Y_{i+\tau} - \overline{Y}\right)}{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}$$

- As respectivas estimativas são dadas por:
  - Função média amostral: $\widetilde{\mathcal{E}}(Y(t)) = \widetilde{\mu}(t) = \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.
  - Função variância amostral: $\widetilde{\mathcal{V}}(Y(t)) = \widetilde{\sigma}^2(t) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$.

Cont.

- Cont. (estimativas)
  - Função de auto-covariância amostral:

$$\widetilde{\gamma}(\tau) = \frac{1}{n}\sum_{i=1}^{n-\tau}(y_i - \overline{y})(y_{i+\tau} - \overline{y}).$$

  - Função de auto-correlação amostral:

$$\widetilde{\rho}(\tau) = \frac{\widetilde{\gamma}(\tau)}{\widetilde{\gamma}(0)} = \frac{\frac{1}{n}\sum_{i=1}^{n-\tau}(y_i - \overline{y})(y_{i+\tau} - \overline{y})}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

  - Na prática, calculamos $\widetilde{\rho}(\tau), \tau = 1, 2, ..., l$, em que $l < n$ é chamado de lag (defasagem) máxima.

Ilustração do cálculo das auto-correlações ($\tau = 0$)

Y_1	.	.	$\cdots$	.	.
Y_2	Y_2	.	$\cdots$	.	.
Y_3	Y_3	Y_3	$\cdots$	.	.
$\vdots$	$\vdots$	$\vdots$	$\ddots$	.	.
Y_{n-2}	Y_{n-2}	Y_{n-2}	$\cdots$	.	.
Y_{n-1}	Y_{n-1}	Y_{n-1}	$\cdots$	Y_{n-1}	.
Y_n	Y_n	Y_n	$\cdots$	Y_n	Y_n

Ilustração do cálculo das auto-correlações ($\tau = 1$)

Y_1	.	.	$\cdots$	.	.
Y_2	Y_2	.	$\cdots$	.	.
Y_3	Y_3	Y_3	$\cdots$	.	.
$\vdots$	$\vdots$	$\vdots$	$\ddots$	.	.
Y_{n-2}	Y_{n-2}	Y_{n-2}	$\cdots$	.	.
Y_{n-1}	Y_{n-1}	Y_{n-1}	$\cdots$	Y_{n-1}	.
Y_n	Y_n	Y_n	$\cdots$	Y_n	Y_n

Ilustração do cálculo das auto-correlações ($\tau = 2$)

Y_1	.	.	$\cdots$	.	.
Y_2	Y_2	.	$\cdots$	.	.
Y_3	Y_3	Y_3	$\cdots$	.	.
$\vdots$	$\vdots$	$\vdots$	$\ddots$	.	.
Y_{n-2}	Y_{n-2}	Y_{n-2}	$\cdots$	.	.
Y_{n-1}	Y_{n-1}	Y_{n-1}	$\cdots$	Y_{n-1}	.
Y_n	Y_n	Y_n	$\cdots$	Y_n	Y_n

Observação (Processo Estacionário)

- É possível provar que ($a_i, i = 1, ..., n$ não aleatórios):

$$\mathcal{V}\left(\sum_{i=1}^{n} a_i Y_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(i-j) \geq 0.$$

- Dessa forma, um estimador para a variância acima, tem de gerar estimativas ≥ 0 com probabilidade 1, ou seja

$$\widehat{\mathcal{V}}\left(\sum_{i=1}^{n} a_i Y_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \widehat{\gamma}(i-j) \geq 0.$$

- O estimador $\widehat{\gamma}(\tau)$ garante a restrição acima, mas o estimador $\widehat{\gamma}^*(\tau) = \frac{n}{n-\tau}\widehat{\gamma}(\tau)$, não. Contudo, ambos são viciados.

Cont.

- Bartlett (1946) mostrou que se uma ST apresentar um comportamento puramente aleatório (essencialmente, as observações tem correlação nula), então os estimadores dos coeficientes de auto-correlação $\widehat{\rho}(\tau)$ terão distribuição aproximadamente $N\left(0, \frac{1}{n}\right)$, em que n é o tamanho da amostra (tamanho da ST).

Cont.

- Assim, não é difícil demonstrar que o intervalo de confiança (assintótico) bicaudal, com por exemplo $\delta\%$ de confiança, para qualquer $\rho(\tau)$ será dado por:

$$\widehat{\rho}(\tau) \pm z_{\delta}\frac{1}{\sqrt{n}}$$

em que $P\left(Z \leq z_{\delta}\right) = \dfrac{1+\delta}{2}$, $Z \sim N(0,1)$.

Ruído branco (RB)

- Um outro processo estocástico bastante útil em ST é o chamado ruído branco ("white noise"). Esse termo é emprestado da engenharia de processamento de sinais.
- Dizemos que um processo $\{\epsilon(t), t \in \mathbb{Z}\}$, é um RB se as variáveis aleatórias $(\epsilon(t))$ forem não correlacionadas, ou seja, se $Cov\left[\epsilon(t), \epsilon(t-\tau)\right] = 0$, $\forall t \in \mathbb{Z}$ e $\forall \tau \neq 0$.
- Tal processo será estacionário se $\mathcal{E}[\epsilon(t)] = \mu$ e $\mathcal{V}[\epsilon(t)] = \sigma^2$, $\forall t$. Esse tipo de RB será bastante utilizado ao longo do curso.

FAC e estacionariedade

- Um processo estacionário pode ser identificado por sua FAC, à medida que esta vai para zero "rapidamente" (com o aumento de τ) para processos desse tipo.
- O que é "rapidamente"? A dependência em relação ao passado diminui exponencialmente, por exemplo.

Cont.

- A função de autocorrelação (FAC) proporciona evidências sobre a estacionariedade de uma ST. Tipicamente, séries temporais não-estacionárias apresentam FAC's com valores altos e significativos para muitas defasagens.
- Nos quatro exemplos a seguir, simulamos n $= 1.000$ valores das ST em que $\sigma^2 = 4$ e os demais parâmetros mencionados nos respectivos slides.

Exemplo 1

- Considere o seguinte processo (RB):

$$\begin{aligned} Y(t) &= \epsilon(t), t = 1, 2, 3, ... \\ \epsilon(t) &\sim NID(0, \sigma^2) \end{aligned}$$

em que $NID(\mu, \sigma^2)$ significa iid (independente e identicamente distribuídas) $N(\mu, \sigma^2)$.

- $Y(t)$ é fracamente estacionária? Notemos que, $\mathcal{E}(Y(t)) = 0, \forall t$, $\mathcal{V}(Y(t)) = \sigma^2$, $Cov(Y(t), Y(t-\tau)) = 0$. Logo, $Y(t)$ é fracamente estacionária.

Ilustração gráfica

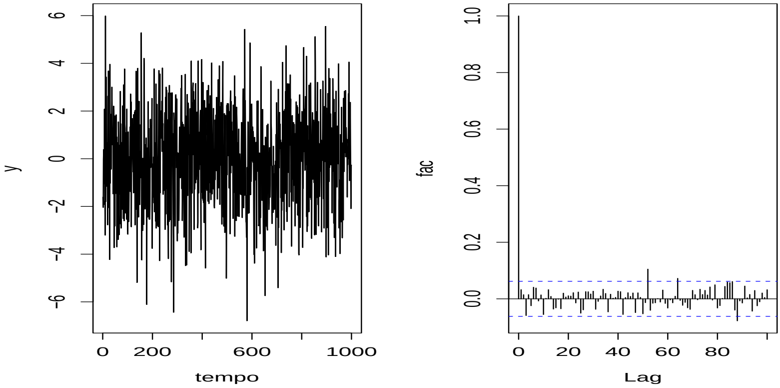


Figura: Exemplo 1

Exemplo 2

- Considere o seguinte processo:

$$
\begin{aligned}
Y_t &= \delta_0 + \delta_1 t + \epsilon_t, t = 1, 2, 3, \ldots \\
\epsilon_t &\sim NID(0, \sigma^2), \delta_1 \neq 0
\end{aligned}
$$

1. Y_t é fracamente estacionária?
2. $Y_t - E(Y_t)$ é fracamente estacionária?
3. $Y_t - Y_{t-1}$ é fracamente estacionária?
4. Qual a intuição dos procedimentos (2) e (3)?

- Nos próximos dois slides discutiremos os quatro pontos acima.

Cont.

1. Temos que $\mathcal{E}(Y_t) = \delta_0 + \delta_1 t$, que não é constante. Assim, apesar de que $\mathcal{V}(Y_t) = \sigma^2$ e $Cov(Y_t, Y_{t-h}) = 0$, $\forall t, h \geq 1$, o processo não é fracamente estacionário.

2. Defina $Z_t = Y_t - E(Y_t)$. Assim, pelo item 1), $Z_t = \epsilon_t$ e, pelo exemplo 1, o processo em questão é fracamente estacionário.

Cont.

3. Defina $W_t = Y_t - Y_{t-1}$. Pelo enunciado, temos que $W_t = \delta_1 + \epsilon_t - \epsilon_{t-1}$. Assim, $\mathcal{E}(W_t) = \delta_1$, $\mathcal{V}(W_t) = 2\sigma^2$ e $Cov(W_t, W_{t-h}) = Cov(\epsilon_t - \epsilon_{t-1}, \epsilon_{t-h} - \epsilon_{t-h-1}) = -\mathcal{V}(\epsilon_{t-1}) = -\sigma^2$, se $h = 1$ e 0, caso contrário. Assim, o processo em questão é fracamente estacionário.
4. A intuição dos procedimentos (2) e (3) é buscar transformações que gerem processos que dependam apenas do RB e, eventualmente, também de quantidades não aleatórias.

Ilustração gráfica ($\delta_0 = 1, \delta_1 = 0,05$)

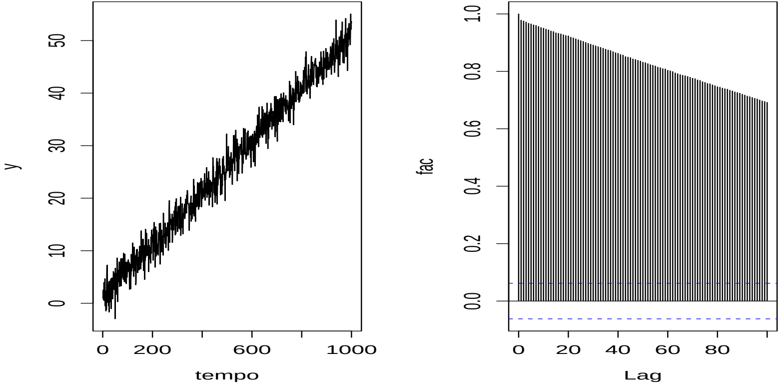


Figura: Exemplo 2

Exemplo 3

- Considere o seguinte processo:

$$
\begin{aligned}
Y_t &= Y_{t-1} + \epsilon_t, t = 1, 2, 3, \ldots \\
\epsilon_t &\sim NID(0, \sigma^2), Y_0 \equiv 0
\end{aligned}
$$

1. Y_t é fracamente estacionária?
2. $Y_t - Y_{t-1}$ é fracamente estacionária?
3. Qual a intuição do procedimento (2)?

- No próximo slide discutiremos os quatro pontos acima.

Cont.

- Primeiramente, note que (exercício) $Y_t = \sum_{i=1}^{t} \epsilon_i$ (representação de médias móveis). Assim, $\mathcal{E}(Y_t) = 0$, que é constante $\forall t$. Mas, $\mathcal{V}(Y_t) = \sigma^2 t$, o qual não é constante em t. Além disso, $Cov(Y_t, Y_{t-\tau}) = Cov\left(\sum_{i=1}^{t} \epsilon_i, \sum_{i=1}^{t-\tau} \epsilon_i\right) = \mathcal{V}\left(\sum_{i=1}^{t-\tau} \epsilon_i\right) = \sigma^2(t-\tau)$. Portanto o processo não é fracamente estacionário.

- Defina $W_t = Y_t - Y_{t-1}$. Pelo enunciado, temos que $W_t = \epsilon_t$. Assim, $\mathcal{E}(W_t) = 0$, $\mathcal{V}(W_t) = \sigma^2$ e $Cov(W_t, W_{t-h}) = Cov(\epsilon_t, \epsilon_{t-h}) = 0$, caso contrário. Assim, o processo em questão é fracamente estacionário.

Ilustração gráfica

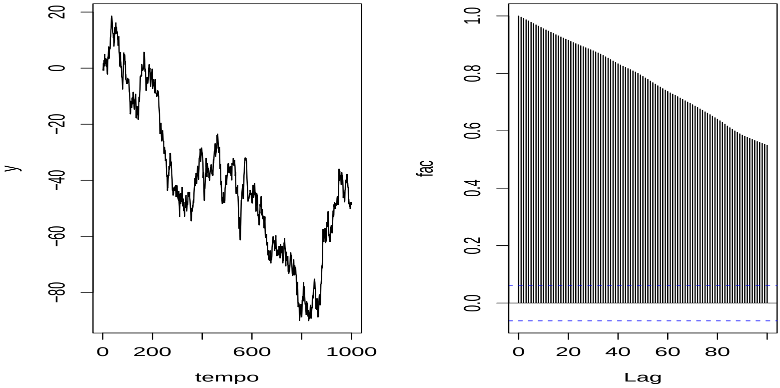


Figura: Exemplo 3

Exemplo 4

- Considere o seguinte processo:

$$
\begin{aligned}
Y_t &= \epsilon_t - \alpha\epsilon_{t-1}, t = 1, 2, 3, ... \\
\epsilon_t &\sim NID(0, \sigma^2), \alpha \neq 0, \epsilon_0 \equiv 0
\end{aligned}
$$

- Y_t é fracamente estacionária? Note que $\mathcal{E}(Y_t) = 0$, $\mathcal{V}(Y_t) = \sigma^2\left(1 + \alpha^2\right)$ e $Cov(Y(t), Y(t-\tau)) = Cov(\epsilon_t - \alpha\epsilon_{t-1}, \epsilon_{t-\tau} - \alpha\epsilon_{t-\tau-1}) = -\alpha\sigma^2$, se $\tau = 1$ e 0, caso contrário. Assim, o processo em questão é fracamente estacionário.

Ilustração gráfica ($\alpha = 1$)

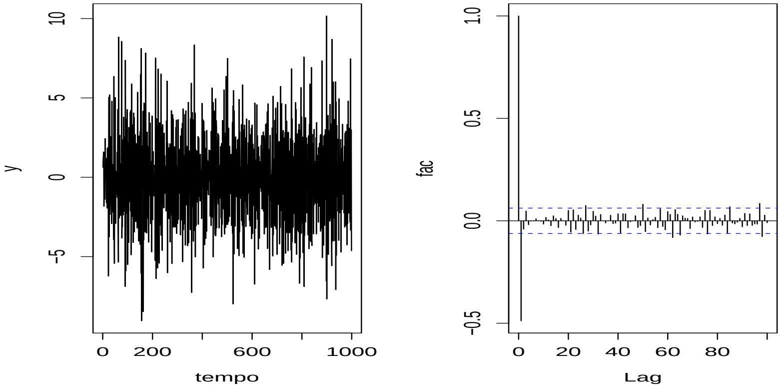


Figura: Exemplo 4

Exercício 1

- Considere o seguinte processo:

$$
\begin{aligned}
Y_t &= \alpha + Y_{t-1} + \epsilon_t, t = 1, 2, 3, ... \\
\epsilon_t &\sim NID(0, \sigma^2), \alpha \neq 0, Y_0 = 0 \,(\text{constante})
\end{aligned}
$$

1. Y_t é fracamente estacionária?
2. $Y_t - Y_{t-1}$ é fracamente estacionária?
3. Simule t=1.000 observações com $\alpha = 1$ e $\sigma^2 = 4$, apresentando os gráficos da série gerada e do respectivo ACF.

Exercício 2

- Considere o seguinte processo:

$$\begin{aligned} Y_t &= \phi Y_{t-1} + \epsilon_t, t = 1, 2, 3, ... \\ \epsilon_t &\sim NID(0, \sigma^2),\, Y_0 = 0\,(\text{constante}) \end{aligned}$$

1. Y_t é fracamente estacionária?
2. $Y_t - \phi Y_{t-1}$ é fracamente estacionária?
3. Para $\phi = 1,1$, simule t=1.000 observações com $\alpha = 1$ e $\sigma^2 = 4$, apresentando os gráficos da série gerada e do respectivo ACF.
4. Para $\phi = -1,1$, simule t=1.000 observações com $\alpha = 1$ e $\sigma^2 = 4$, apresentando os gráficos da série gerada e do respectivo ACF.

Cont.

5. Para $\phi = 0,8$, simule t=1.000 observações com $\alpha = 1$ e $\sigma^2 = 4$, apresentando os gráficos da série gerada e do respectivo ACF.
6. Para $\phi = -0,8$, simule t=1.000 observações com $\alpha = 1$ e $\sigma^2 = 4$, apresentando os gráficos da série gerada e do respectivo ACF.
7. Sob quais condições você acredita que $Y_{t \geq 1}$ seja estacionária? Justifique adequadamente a sua resposta.

Transformações

- Como mencionado, uma das principais suposições a serem consideradas é a estacionariedade, uma vez que muitas das abordagens que serão vistas, dependem da validade de tal suposição.
- Comumente, a ST original não apresenta tal propriedade.
- Uma forma de lidar com essa situação é através de transformações apropriadas.

Transformações: operador diferença

- Um operador bastante útil e comum em ST é o chamado operador diferença(Δ), o qual é definido por:

$$\Delta Y_t = Y_t - Y_{t-1}$$

- Aplicando novamente o operador diferença na equação acimam vem que:

$$\begin{aligned} \Delta^2 Y_t &= \Delta \left(Y_t - Y_{t-1}\right) = Y_t - Y_{t-1} - \left(Y_{t-1} - Y_{t-2}\right) \\ &= Y_t - 2Y_{t-1} + Y_{t-2} \end{aligned}$$

Transformações: operador diferença

- A n-ésima diferença é obtida recursivamente por:

$$\Delta^n Y_t = \Delta\left(\Delta^{n-1} Y_t\right)$$

- Normalmente, uma ou duas diferenças são suficientes para tornar uma dada série estacionária.

Transformação potência

- Proposta por Tukey (1957), dada por:

$$Y_t^{(\lambda)} = g(Y_t) = \begin{cases} Y_t^{\lambda}, \lambda \neq 0 \\ \ln Y_t, \lambda = 0 \end{cases} \tag{2}$$

- Casos particulares (além da transformação ln(.)).
  - $\lambda = 1$, então $g(Y_t) = Y_t$.
  - $\lambda = 2$, então $g(Y_t) = Y_t^2$.
  - $\lambda = 1/2$, então $g(Y_t) = \sqrt{Y_t}$.

Transformação Box-Cox

- Para mitigar problemas de descontinuidade (para $\lambda = 0$), Box & Cox (1964) propuseram:

$$Y_t^{(\lambda)} = g(Y_t) = \begin{cases} \dfrac{Y_t^{\lambda} - 1}{\lambda}, \lambda \neq 0 \\ \ln Y_t, \lambda = 0 \end{cases} \tag{3}$$

- As transformações dadas em (2) e (3), são mais apropriadas quando se busca obter normalidade da ST (não se prestam, à rigor, para a obtenção de estacionariedade). Veja também aqui.

Ilustração gráfica: Exemplo 2 ($t \times \Delta Y_t$)

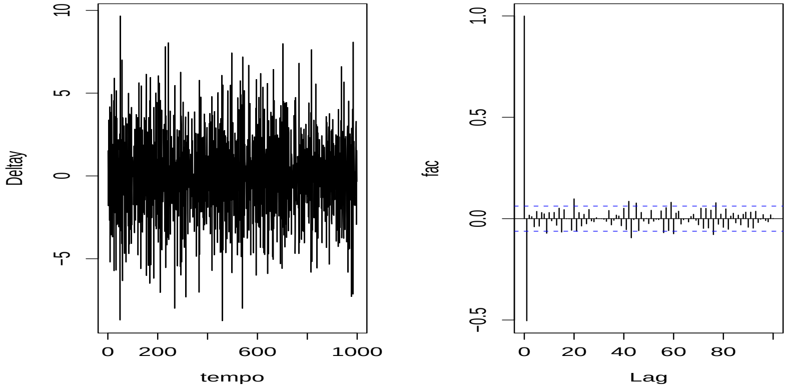


Figura: Exemplo 2 (transformação diferença de ordem 1)

Ilustração gráfica: Exemplo 3 ($t \times \Delta Y_t$)

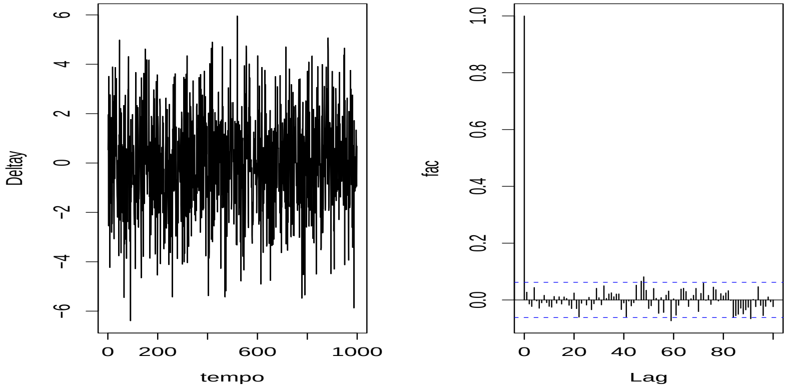


Figura: Exemplo 2: (transformação diferença de ordem 1)