Modelos ARIMA

Ricardo Ehlers
ehlers@icmc.usp.br

Departamento de Matemática Aplicada e Estatística
Universidade de São Paulo

Alguns operadores úteis

- Operador defasagens (*backshift*),

$$
\begin{aligned}
BX_t &= X_{t-1}, \\
B^2X_t &= X_{t-2}, \\
&\vdots \quad , \\
B^mX_t &= X_{t-m}
\end{aligned}
$$

- Operador translação para o futuro (*forward*),

$$
\begin{aligned}
FX_t &= X_{t+1}, \\
F^2X_t &= X_{t+2}, \\
&\vdots \\
F^mX_t &= X_{t+m}
\end{aligned}
$$

- Operador diferença,

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$$

- Operador soma,

$$\begin{aligned} SX_t &= \sum_{j=0}^{\infty} X_{t-j} = X_t + X_{t-1} + \cdots \\ &= (1 + B + B^2 + \cdots)X_t \end{aligned}$$

Modelos Lineares Estacionários

Processo linear geral,

$$\begin{aligned} X_t &= \mu + \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots \\ &= \mu + (1 + \psi_1 B + \psi_2 B^2 + \cdots)\epsilon_t \\ &= \mu + \psi(B)\epsilon_t \end{aligned}$$

sendo que

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots$$

é denominada *função de transferência* e μ representa o nivel da série.

Os erros ϵ_t são tais que,

$$\begin{aligned} E(\epsilon_t) &= 0, \ \forall t, \\ Var(\epsilon_t) &= \sigma_\epsilon^2, \ \forall t, \\ E(\epsilon_t \epsilon_s) &= 0, \ \forall s \neq t, \end{aligned}$$

- O processo X_t será estacionário e $E(X_t) = \mu$ se $\sum_{j=1}^{\infty} \psi_j$ converge.
- Função de autocovariância,

$$\gamma_j = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+j}.$$

  com $\psi_0 = 1$.
- Função de variância,

$$\gamma_0 = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i^2.$$

Portanto a condição de existencia é $\sum_{i=0}^{\infty} \psi_i^2 < \infty$.

Escrevendo $\tilde{X}_t = X_t - \mu$ como uma soma ponderada de seus valores passados,

$$\tilde{X}_t = \pi_1 \tilde{X}_{t-1} + \pi_2 \tilde{X}_{t-2} + \cdots + \epsilon_t.$$

Equivalentemente,

$$(1 - \pi_1 B - \pi_2 B^2 - \cdots)\tilde{X}_t = \epsilon_t \quad \text{ou} \quad \pi(B)\tilde{X}_t = \epsilon_t.$$

Substituindo $\tilde{X}_t = \psi(B)\epsilon_t$ segue que,

$$\pi(B)\psi(B)\epsilon_t = \epsilon_t,$$

e finalmente,

$$\pi(B) = \psi^{-1}(B).$$

Portanto, os coeficientes $\pi_1, \pi_2, \ldots$ podem ser obtidos a partir de $\psi_1, \psi_2, \ldots$ (e vice-versa).

Exemplo. Considere o filtro linear geral com $\psi_j = \phi^j$ sendo $|\phi| < 1$. Temos que,

$$\sum_{j=0}^{\infty} \psi_j = \sum_{j=0}^{\infty} \phi^j = \frac{1}{1-\phi}.$$

Portanto, $E(X_t) = \mu$ e como $\sum_{j=0}^{\infty} \psi_j^2 < \infty$,

$$\begin{aligned} \gamma_j &= \sigma_\epsilon^2 \frac{\phi^j}{1-\phi^2} \\ \gamma_0 &= \sigma_\epsilon^2 \frac{1}{1-\phi^2} \end{aligned}$$

Exemplo. No exemplo anterior se $\phi = 1$ e $\mu = 0$ segue que o processo

$$X_t = \epsilon_t + \epsilon_{t-1} + \cdots$$

não é estacionário pois $\sum_{j=0}^{\infty} \psi_j$ não converge.

Exemplo. Um caso particular do filtro linear geral é,

$$\tilde{X}_t = \epsilon_t - \theta\epsilon_{t-1}.$$

Note que $\sum_{j=0}^{\infty} \psi_j = 1 - \theta$ e assim o processo é estacionário para qualquer θ.

Podemos reescrever o processo como,

$$\tilde{X}_t = (1 - \theta B)\epsilon_t$$

ou equivalentemente,

$$\epsilon_t \;=\; (1 - \theta B)^{-1}\tilde{X}_t = (1 + \theta B + \theta^2 B^2 + \cdots)\tilde{X}_t$$

Finalmente, temos que

$$\pi(B) = (1 + \theta B + \theta^2 B^2 + \cdots),\; \pi_j = -\theta^j.$$

- A sequência de pesos π_j converge se $|\theta| < 1$.
- Neste caso dizemos que o processo é *inversivel*.

Processos de Médias Móveis

Definição

Seja $\{\epsilon_t\}$ um processo discreto puramente aleatório com $E(\epsilon_t) = 0$ e $Var(\epsilon_t) = \sigma_\epsilon^2$.

$\{X_t\}$ é chamado de processo de médias móveis de ordem q, ou MA(q), se

$$X_t = \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}, \tag{1}$$

sendo $\beta_i \in \mathbb{R}$, $i = 1, \ldots, q$.

Verifique como ficam a média e a variância deste processo,

$$\begin{aligned} E(X_t) &= E(\epsilon_t) + \sum_{j=1}^{q} \beta_j E(\epsilon_{t-j}) = 0 \\ Var(X_t) &= Var(\epsilon_t) + \sum_{j=1}^{q} \beta_j^2 Var(\epsilon_{t-j}) \\ &= (1 + \beta_1^2 + \cdots + \beta_q^2)\ \sigma_\epsilon^2. \end{aligned}$$

Verifique a estrutura de autocovarância,

$$
\begin{aligned}
\gamma(k) &= Cov(X_t, X_{t+k}) \\
&= Cov(\epsilon_t + \beta_1\epsilon_{t-1} + \cdots + \beta_q\epsilon_{t-q}, \epsilon_{t+k} + \beta_1\epsilon_{t+k-1} + \cdots + \beta_q\epsilon_{t+k-q}) \\
&= \begin{cases} 0 & k > q \\ \sigma_\epsilon^2 \sum_{j=0}^{q-k} \beta_j\beta_{j+k} & k = 0, \ldots, q \\ \gamma(-k) & k < 0 \end{cases}
\end{aligned}
$$

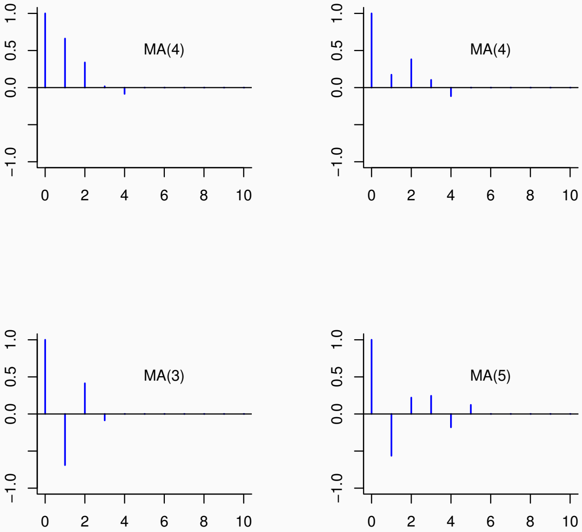
com $\beta_0 = 1$.

- Como a média e a variância são constantes e $\gamma(k)$ não depende de t o processo é (fracamente) estacionário para todos os possíveis valores de $\beta_1, \ldots, \beta_q$.
- Se os ϵ_t's forem normalmente distribuidos os X_t's também serão e portanto o processo será estritamente estacionário.

Função de Autocorrelação

$$\rho(k) = \begin{cases} 1 & k = 0 \\ \displaystyle\sum_{j=0}^{q-k} \beta_j \beta_{j+k} \Bigg/ \sum_{j=0}^{q} \beta_j^2 & k = 1, \dots, q \\ 0 & k > q \\ \rho(-k) & k < 0. \end{cases}$$

Autocorrelações teóricas de processos MA(q).



Para $q = 1$,

$$\rho(k) = \begin{cases} \beta_1/(1+\beta_1^2) & k = 1 \\ 0 & k > 1. \end{cases}$$

O processo é estacionário $\forall\ \beta_1 \in \mathbb{R}$ mas não é *inversivel*.

$$\begin{aligned} X_t &= \epsilon_t + \theta\ \epsilon_{t-1} \\ X_t &= \epsilon_t + \frac{1}{\theta}\ \epsilon_{t-1} \end{aligned}$$

levam a mesma estrutura de autocorrelação $\rho(1) = \theta/(1+\theta^2)$.

Reescrevendo as equações acima como,

$$\begin{aligned}
\epsilon_t &= X_t - \theta \ \epsilon_{t-1} \\
\epsilon_t &= X_t - \frac{1}{\theta} \ \epsilon_{t-1}
\end{aligned}$$

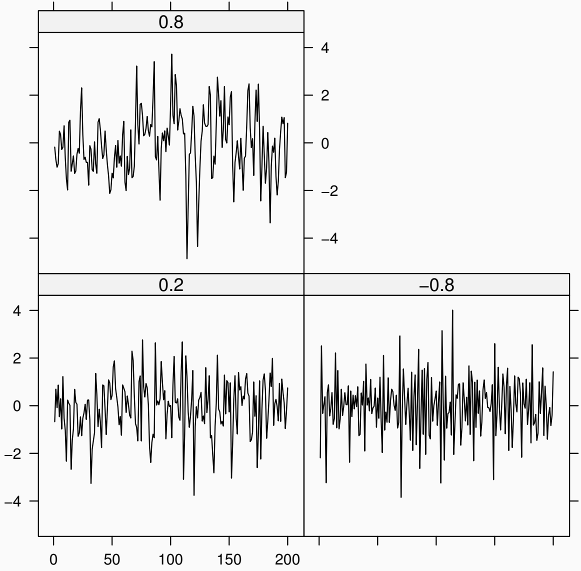
e fazendo substituições sucessivas obtemos,

$$\begin{aligned}
\epsilon_t &= X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \theta^3 X_{t-3} + \dots \\
\epsilon_t &= X_t - \frac{1}{\theta} X_{t-1} + \frac{1}{\theta^2} X_{t-2} - \frac{1}{\theta^3} X_{t-3} + \dots
\end{aligned}$$

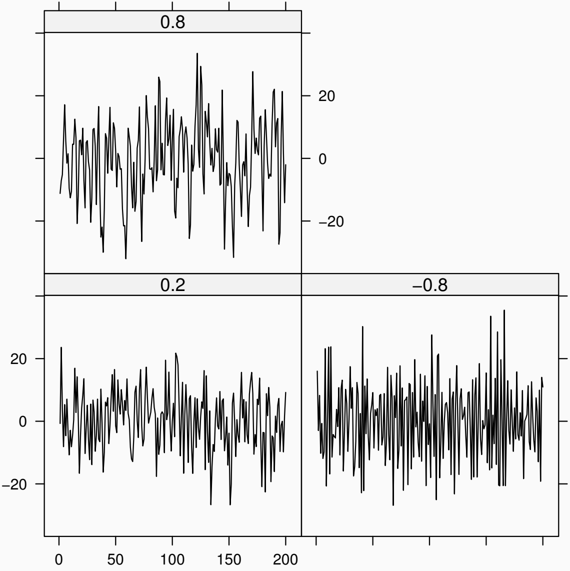
- Se $|\theta| < 1$ a primeira série converge e o modelo é dito ser *inversível* mas a segunda não converge e o modelo é *não inversível*.
- A condição de inversibilidade garante que existe um único processo MA(1) para uma dada função de autocorrelação.
- O processo MA(1) pode ser reescrito como uma regressão de ordem infinita nos seus próprios valores defasados.

$$X_t = \theta X_{t-1} - \theta^2 X_{t-2} + \theta^3 X_{t-3} + \cdots + \epsilon_t$$

Processos MA(1) simulados com $\theta \in \{0.2, 0.8, -0.8\}$ e erros N(0,1).



Processos MA(1) simulados com $\theta \in \{0.2, 0.8, -0.8\}$ e erros N(0,100).



Caso geral, processo $MA(q)$,

$$\begin{aligned} X_t &= \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q} \\ &= (1 + \beta_1 B + \cdots + \beta_q B^q)\epsilon_t = \theta(B)\epsilon_t \end{aligned}$$

é inversível se as raízes $\delta_1, \ldots, \delta_q$ de $\theta(B) = 0$ são tais que,

$$\delta_i < -1 \quad \text{ou} \quad \delta_i > 1, \; i = 1, \ldots, q.$$

Dizemos que as raízes de $\theta(B) = 0$ estão fora do círculo unitário.

- Teremos então 2^q modelos com a mesma função de autocorrelação mas somente um deles será inversível.
- Uma constante μ qualquer pode ser adicionada ao lado direito dando origem a um processo com média μ,

$$X_t = \mu + \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}.$$

- O processo continuará sendo estacionário com $E(X_t) = \mu$ e a função de autocorrelação não será afetada.
- Como escolher o valor de q?

Exemplo. Considere o modelo MA(2), $X_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$.

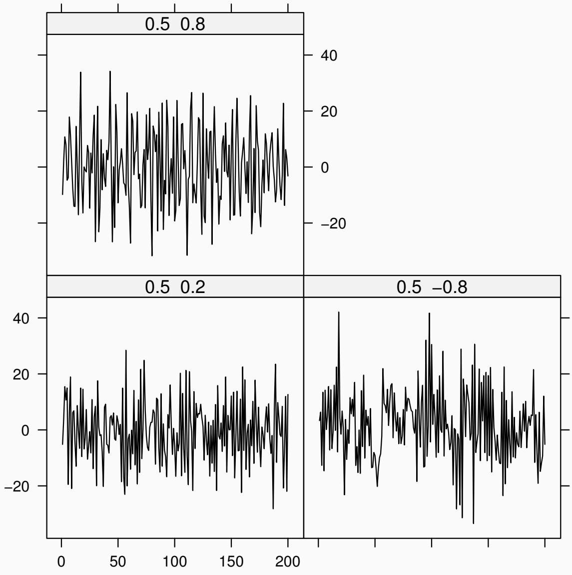
- O processo é estacionário para valores quaisquer de θ_1 e θ_2.
- O processo é inversivel se as raízes de

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 = 0$$

  estiverem fora do circulo unitário.
- Equivalentemente temos que,

$$\theta_1 + \theta_2 < 1, \quad \theta_2 - \theta_1 < 1, \quad -1 < \theta_2 < 1.$$

Processos MA(2) simulados com erros N(0,100).



Processos Autoregressivos

Seja $\{\epsilon_t\}$ um processo puramente aleatório com $E(\epsilon_t) = 0$ e $Var(\epsilon_t) = \sigma_\epsilon^2$.

Então $\{X_t\}$ é um processo autoregressivo de ordem p, ou AR(p), se

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \epsilon_t$$

Em termos de operador de retardo temos,

$$\begin{aligned}(1-\alpha_1 B-\cdots-\alpha_p B^p)X_t &= \epsilon_t \\ \phi(B)X_t &= \epsilon_t\end{aligned}$$

Condição de estacionariedade do processo AR(p):

raizes de $\phi(B)=0$ fora do círculo unitário.

O Processo AR(1)

Para $p = 1$,

$$
\begin{aligned}
X_t &= \alpha X_{t-1} + \epsilon_t \\
&= \alpha(\alpha X_{t-2} + \epsilon_{t-1}) + \epsilon_t = \alpha^2 X_{t-2} + \alpha\epsilon_{t-1} + \epsilon_t \\
&= \alpha^2(\alpha X_{t-3} + \epsilon_{t-2}) + \alpha\epsilon_{t-1} + \epsilon_t \\
&= \alpha^3 X_{t-3} + \alpha^2\epsilon_{t-2} + \alpha\epsilon_{t-1} + \epsilon_t \\
&\vdots \\
&= \alpha^{r+1} X_{t-r-1} + \sum_{j=0}^{r} \alpha^j \epsilon_{t-j}
\end{aligned}
$$

$$
\begin{aligned}
X_t - \sum_{j=0}^{r} \alpha^j \epsilon_{t-j} &= \alpha^{r+1} X_{t-r-1} \\
(X_t - \sum_{j=0}^{r} \alpha^j \epsilon_{t-j})^2 &= \alpha^{2(r+1)} X_{t-r-1}^2 \\
E(X_t - \sum_{j=0}^{r} \alpha^j \epsilon_{t-j})^2 &= \alpha^{2(r+1)} E(X_{t-r-1}^2)
\end{aligned}
$$

Se $|\alpha| < 1$ e $r \to \infty$, podemos escrever o processo X_t como,

$$X_t = \sum_{j=0}^{\infty} \alpha^j \epsilon_{t-j}.$$

Portanto,

$$\begin{aligned} E(X_t) &= 0 \\ Var(X_t) &= \sigma_\epsilon^2(1+\alpha^2+\alpha^4+\dots) = \frac{\sigma_\epsilon^2}{1-\alpha^2} \end{aligned}$$

$$Cov(X_t, X_{t+k}) =$$
$$Cov(\epsilon_t + \alpha\epsilon_{t-1} + \alpha^2\epsilon_{t-2} + \dots,$$
$$\epsilon_{t+k} + \dots + \alpha^{k-1}\epsilon_{t+1} + \alpha^k\epsilon_t + \alpha^{k+1}\epsilon_{t-1} + \alpha^{k+2}\epsilon_{t-2} + \dots)$$

$$\begin{aligned} &= \alpha^k Var(\epsilon_t) + \alpha^{k+2} Var(\epsilon_{t-1}) + \dots \\ &= \alpha^k\sigma_\epsilon^2(1+\alpha^2+\alpha^4+\dots) = \frac{\alpha^k\sigma_\epsilon^2}{1-\alpha^2}. \end{aligned}$$

A função de autocorrelação é,

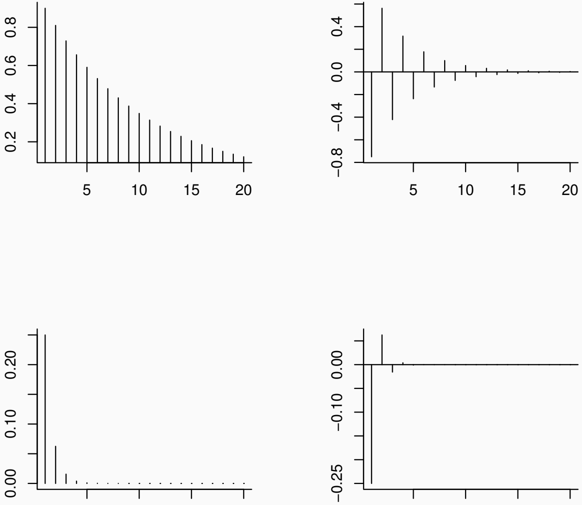
$$\rho(k) = \alpha^k$$

e o processo pode ser escrito como um MA infinito em $t + k$,

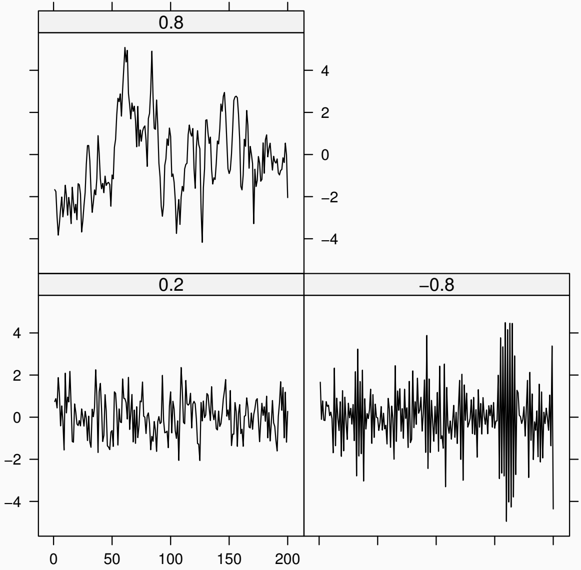
$$X_{t+k} = \epsilon_{t+k} + \alpha\epsilon_{t+k-1} + \cdots + \alpha^k\epsilon_t + \ldots$$

O efeito de ϵ_t sobre X_{t+k} diminui a medida que k aumenta e por isso é chamado *efeito transitório*.

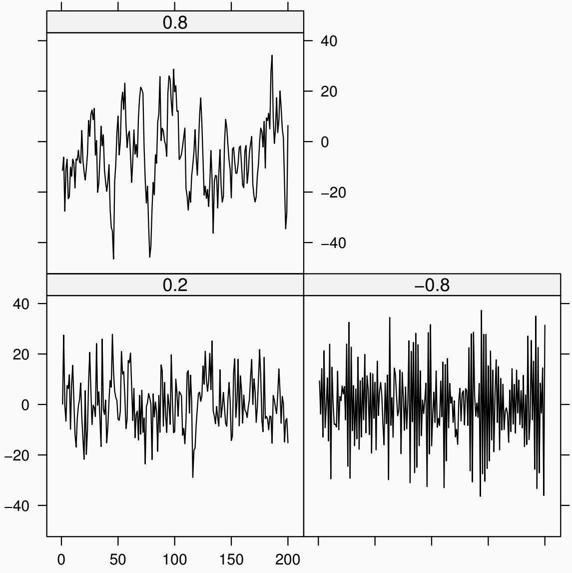
Autocorrelações teoricas de processos AR(1) com $\alpha \in \{0.9, -0.75, 0.25, -0.25\}$.



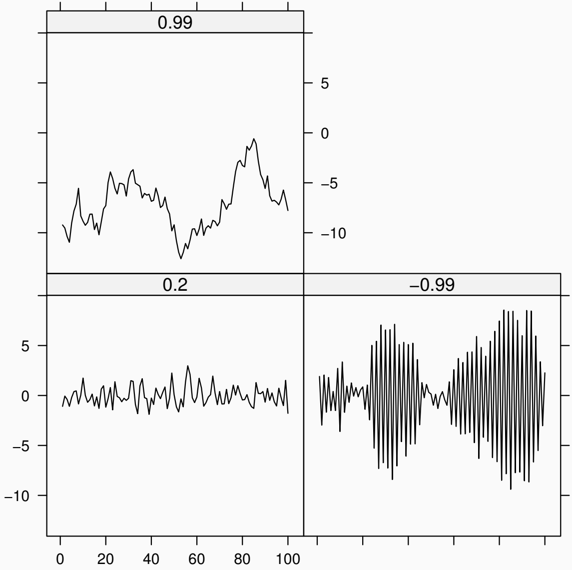
Processos AR(1) simulados com $\alpha \in \{0.2, 0.8, -0.8\}$ e erros N(0,1).



Processos AR(1) simulados com $\alpha \in \{0.2, 0.8, -0.8\}$ e erros N(0,100).



Processos AR(1) simulados com $\alpha \in \{0.2, 0.99, -0.99\}$ e erros N(0,1).



No processo $AR(p)$ temos que,

$$\phi(B)X_t = \epsilon_t.$$

Reescrevendo o processo como um MA(∞),

$$X_t \;=\; \psi_0\epsilon_t + \psi_1\epsilon_{t-1} + \psi_2\epsilon_{t-2} + \cdots = \psi(B)\epsilon_t.$$

que é estacionário se $\sum_j \psi_j < \infty$.

Como obter os coeficientes $\psi_0, \psi_1, \ldots$ a partir dos coeficientes $\alpha_1, \ldots, \alpha_p$?

$$
\begin{aligned}
\phi(B)X_t &= \epsilon_t \\
\phi(B)\psi(B)\epsilon_t &= \epsilon_t \\
(1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p)(\psi_0 + \psi_1 B + \psi_2 B^2 + \dots) &= \\
1 + 0 + 0 + \dots
\end{aligned}
$$

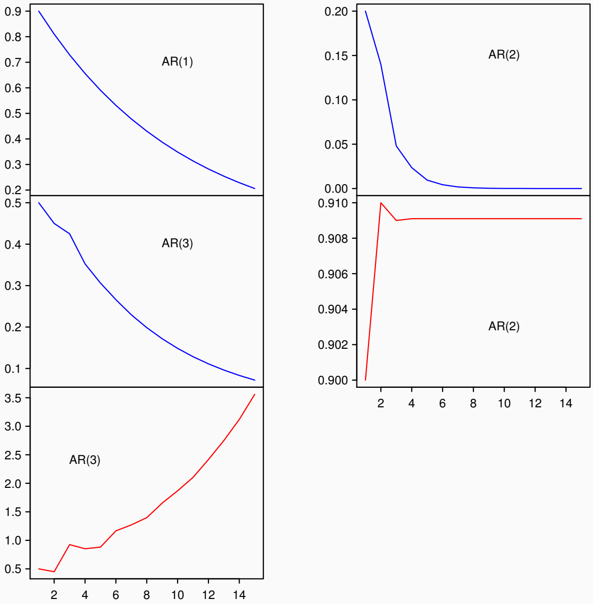
$$
\begin{aligned}
\psi_0(1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p) & + \\
\psi_1(B - \alpha_1 B^2 - \alpha_2 B^3 - \cdots - \alpha_p B^{p+1}) & + \\
\psi_2(B^2 - \alpha_1 B^3 - \alpha_2 B^4 - \cdots - \alpha_p B^{p+2}) & + \\
& \vdots \\
= 1 + 0 + 0 + \dots
\end{aligned}
$$

$$
\psi_0 + (\psi_1 - \psi_0\alpha_1)B + (\psi_2 - \psi_1\alpha_1 - \psi_0\alpha_2)B^2 + \cdots = 1 + 0B + 0B^2 + \dots
$$

$$
\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= \psi_0\alpha_1 \\
\psi_2 &= \psi_1\alpha_1 + \psi_0\alpha_2 \\
\psi_3 &= \psi_2\alpha_1 + \psi_1\alpha_2 + \psi_0\alpha_3 \\
&\vdots \\
\psi_i &= \sum_{j=1}^{i} \psi_{i-j}\alpha_j. \\
X_{t+k} &= \epsilon_{t+k} + \psi_1\epsilon_{t+k-1} + \cdots + \psi_k\epsilon_t + \dots
\end{aligned}
$$

Um gráfico $\psi_k \times k$ é chamado *função resposta a impulso*.

Exemplo. Função resposta a impulso, AR(1) com $\alpha = 0.9$, AR(2) com $\alpha = (0.2, 0.1)$ e $\alpha = (0.9, 0.1)$, AR(3) com $\alpha = (0.5, 0.2, 0.1)$ e $\alpha = (0.5, 0.2, 0.6)$.


AR(1)
AR(2)
AR(3)
AR(2)
AR(3)

- Os 2 ultimos processos não são estacionários.
- As raizes do processo AR(2) são 1 e 10
- As raizes do processo AR(3) são

$$0.8783, \quad -0.6058 + 1.2371i, \quad -0.6058 - 1.2371i.$$

Autocorrelações teóricas de um processo $AR(p)$

Seja $\{X_t\}$ um processo AR(p) estacionário com,

$$E(X_t) = 0,\ Var(X_t) = \sigma_X^2,\ Cov(X_t, X_{t-k}) = \gamma(k).$$

- Multiplique o AR(p) por X_{t-k},

$$X_t X_{t-k} = \alpha_1 X_{t-1} X_{t-k} + \cdots + \alpha_p X_{t-p} X_{t-k} + \epsilon_t X_{t-k}.$$

- Tome o valor esperado,

$$\begin{aligned} E(X_t X_{t-k}) = \gamma(k) &= \alpha_1 E(X_{t-1} X_{t-k}) + \cdots + \alpha_p E(X_{t-p} X_{t-k}) \\ &= \alpha_1 \gamma(k-1) + \cdots + \alpha_p \gamma(k-p), \quad k > 0. \end{aligned}$$

- Divida ambos os lados por σ_X^2,

$$\rho(k) = \alpha_1 \rho(k-1) + \cdots + \alpha_p \rho(k-p), \quad k > 0$$

Equações de Yule-Walker,

$$
\begin{aligned}
\rho(1) &= \alpha_1\ \rho(0) + \alpha_2\ \rho(-1) + \cdots + \alpha_p\ \rho(1-p) \\
\rho(2) &= \alpha_1\ \rho(1) + \alpha_2\ \rho(0) + \cdots + \alpha_p\ \rho(2-p) \\
&\ \vdots \\
\rho(p) &= \alpha_1\ \rho(p-1) + \alpha_2\ \rho(p-2) + \cdots + \alpha_p\ \rho(0) \\
&\ \vdots
\end{aligned}
$$

- Processo $AR(1)$,

$$\begin{aligned}
\rho(1) &= \alpha_1 \\
\rho(2) &= \alpha_1 \rho(1) = \alpha^2 \\
&\vdots \\
\rho(k) &= \alpha_1^k.
\end{aligned}$$

- Processo $AR(2)$,

$$\begin{aligned}
\rho(1) &= \alpha_1 + \alpha_2\ \rho(1) \Rightarrow \rho(1) = \alpha_1/(1-\alpha_2) \\
\rho(k) &= \alpha_1\ \rho(k-1) + \alpha_2\ \rho(k-2),\ k = 2, 3, \ldots.
\end{aligned}$$

- Conclusão: O correlograma não informa sobre a ordem p do modelo.

Variância de X_t no processo AR(p)

- No processo AR(p) multiplique os dois lados por X_t,

$$X_t^2 = \alpha_1 X_t X_{t-1} + \cdots + \alpha_p X_t X_{t-p} + X_t \epsilon_t.$$

- Tome o valor esperado,

$$\begin{aligned}
E(X_t^2) &= \alpha_1 E(X_t X_{t-1}) + \cdots + \alpha_p E(X_t X_{t-p}) + E(X_t \epsilon_t) \\
&= \alpha_1 \gamma(1) + \cdots + \alpha_p \gamma(p) + E(\psi(B)\epsilon_t \epsilon_t) \\
&= \alpha_1 \gamma(1) + \cdots + \alpha_p \gamma(p) + E(\epsilon_t^2) \\
1 &= \alpha_1 \rho(1) + \cdots + \alpha_p \rho(p) + \sigma_\epsilon^2 / \sigma_X^2
\end{aligned}$$

$$\sigma_X^2 = \frac{\sigma_\epsilon^2}{1 - \alpha_1 \rho(1) - \cdots - \alpha_p \rho(p)}$$

As p primeiras equações de Yule-Walker podem ser escritas na forma matricial,

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & 1 & \cdots & \rho(p-2) \\ \vdots & \vdots & & \vdots \\ \rho(p-1) & \rho(p-2) & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}$$

Autocorrelações amostrais

Dados os valores observado $x_1, x_2, \ldots, x_n$ de uma série temporal as autocorrelações $\rho(j)$ são estimadas como,

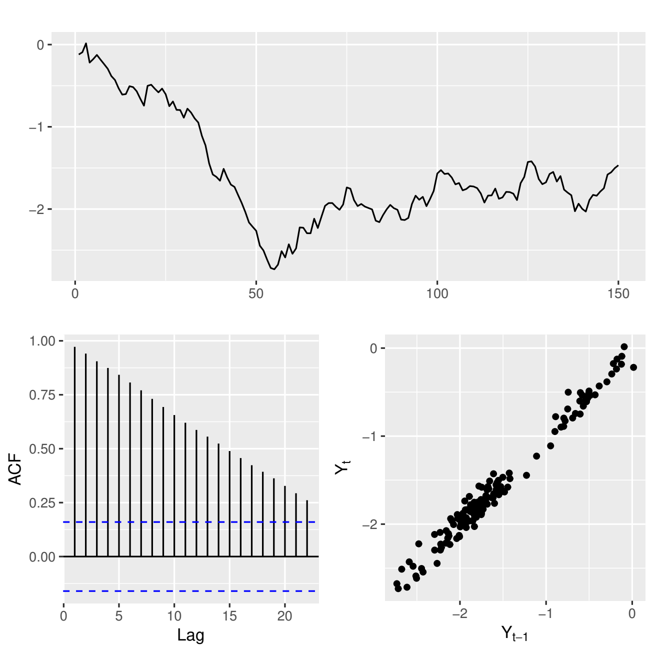
$$r_k = \frac{c_k}{c_0},$$

sendo c_k estimativas das autocovariâncias,

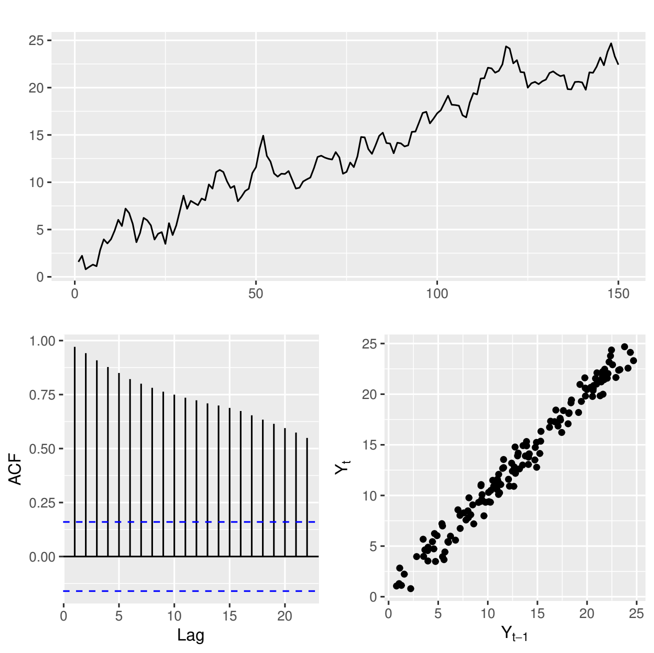
$$c_k = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \overline{x})(x_{t+k} - \overline{x}),\ \ k = 0, 1, \ldots .$$

com $r_{-k} = r_k$.

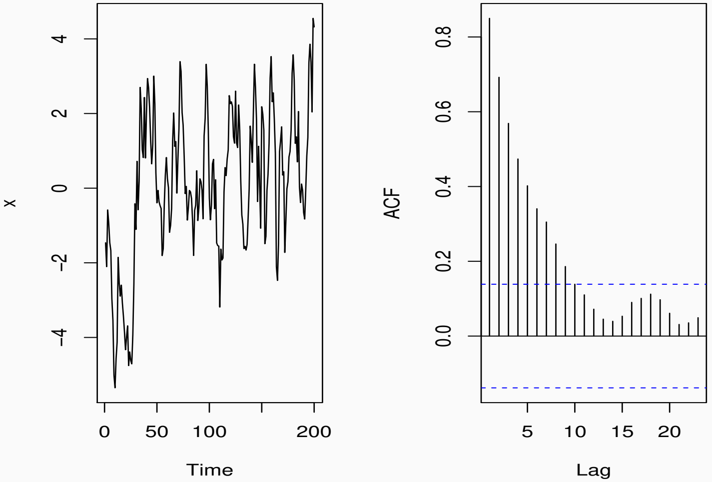
Passeio aleatorio com $\sigma^2 = 0.1$.



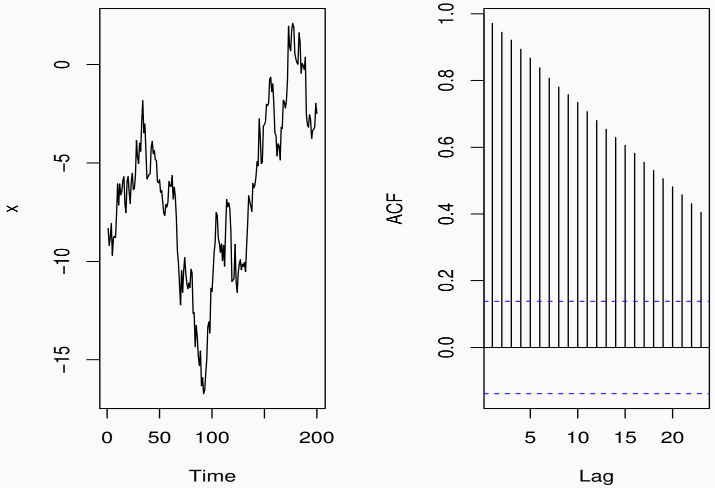
Passeio aleatorio com $\sigma^2 = 1$.



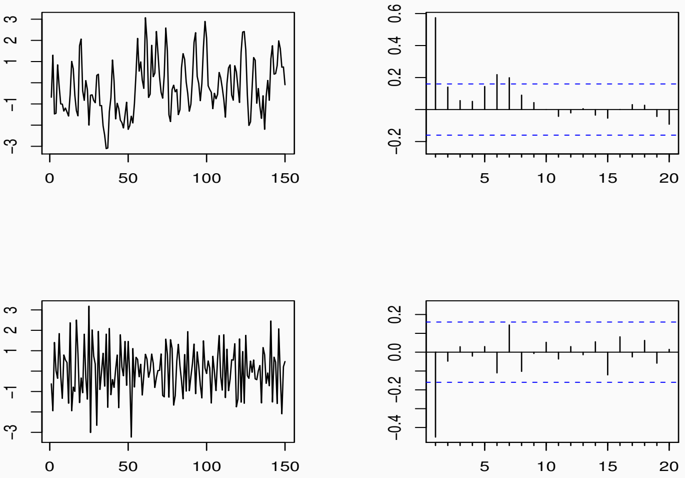
Processo AR(1) com $\alpha = 0.8$ e $\sigma_\epsilon^2 = 1$.



Processo AR(1) com $\alpha = 0.99$ e $\sigma_\epsilon^2 = 1$.



Processos MA(1) simulados com $\theta \in \{0.8, -0.8\}$ e erros N(0,1).



Autocorrelações Parciais

- Para um processo AR(p), o coeficiente α_p mede o "excesso de correlação" na defasagem p que não é levado em conta por um modelo AR($p-1$).
- Este é chamado de p-ésimo *coeficiente de autocorrelação parcial*.
- Variando $k = 1, 2, \ldots$ temos a chamada *função de autocorrelação parcial* (FACP).

- Em um processo AR(p) não existe correlação direta entre X_t e $X_{t-p-1}, X_{t-p-2}, \ldots$
- Substituindo $k = p+1, p+2, \ldots$ nas equações de Yule-Walker obtem-se que todos os coeficientes de correlação parcial serão nulos para $k > p$.
- Por exemplo, substituindo-se $k = p+1$ segue que

$$\rho(p+1) = \alpha_1 \rho(p) + \cdots + \alpha_p \rho(1) + \alpha_{p+1}.$$

- Este fato é sugerido em como uma ferramenta para determinar a ordem p do processo autoregressivo para séries temporais observadas.

Modelos Mistos ARMA

Combinando-se processos AR e MA pode-se obter uma representação adequada com um número menor de parâmetros.

O modelo ARMA(p, q) é dado por

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}$$

sendo $\{\epsilon_t\}$ é um processo puramente aleatório com média zero e variância σ_ϵ^2.

- Note que, modelos AR ou MA podem ser obtidos como casos especiais quando $p = 0$ ou $q = 0$.
- Usando o operador de retardo o modelo pode ser reescrito como

$$(1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p) X_t = (1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q) \epsilon_t$$

ou

$$\phi(B) X_t = \theta(B) \epsilon_t.$$

- Os valores de $\alpha_1, \ldots, \alpha_p$ que tornam o processo estacionário são tais que as raízes de $\phi(B) = 0$ estão fora do círculo unitário.
- Analogamente, os valores de $\beta_1, \ldots, \beta_q$ que tornam o processo inversível são tais que as raízes de $\theta(B) = 0$ estão fora do círculo unitário.

- As funções de autocorrelação e autocorrelação parcial ficam consideravelmente mais complicadas em processos ARMA.
- De um modo geral, para um processo ARMA(p, q) estacionário a função de autocorrelação tem um decaimento exponencial ou oscilatório após a defasagem q enquanto que a facp tem o mesmo comportamento após a defasagem p.
- Na prática pode ser bastante difícil distinguir entre decaimentos exponenciais e oscilatórios através das estimativas destas funções.

Propriedades teóricas da fac e facp.

Processo	FAC	FACP
série aleatória	0	0
AR(1), $\alpha > 0$	decaimento exponencial	0, $k \geq 2$
AR(1), $\alpha < 0$	decaimento oscilatório	idem
AR(p)	decaimento para zero	0, $k > p$
MA(1)	0, $k > 1$	decaimento oscilatório
ARMA(p, q)	decaimento a partir de q	decaimento a partir de p

Modelos ARMA Integrados

Um modelo ARMA no qual X_t é substituido pela sua d-ésima diferença $\nabla^d X_t$ é capaz de descrever alguns tipos de séries não estacionárias.

Denotando a série diferenciada por

$$W_t = \nabla^d X_t = (1 - B)^d X_t$$

o processo autoregressivo integrado médias móveis é dado por,

$$W_t = \alpha_1 W_{t-1} + \cdots + \alpha_p W_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}.$$

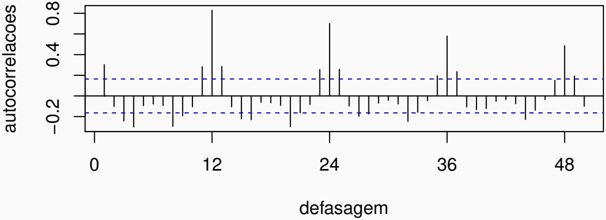
- Denotamos ARIMA(p, d, q)
- Equivalentemente,

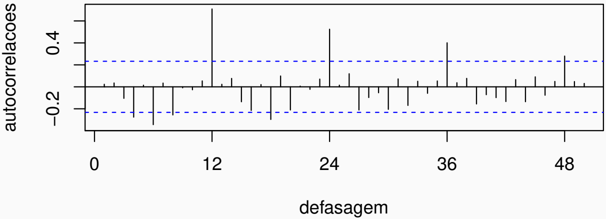
$$\phi(B)(1-B)^d X_t = \theta(B)\epsilon_t. \tag{2}$$

- O modelo para X_t é claramente não estacionário já que o polinômio autoregressivo $\phi(B)(1-B)^d$ tem exatamente d raízes sobre o círculo unitário, ou d *raízes unitárias*.
- Na prática valores pequenos são em geral especificados para d, sendo $d = 1$ o valor mais frequentemente utilizado e excepcionalmente $d = 2$. Note também que o passeio aleatório pode ser considerado um processo ARIMA(0,1,0).

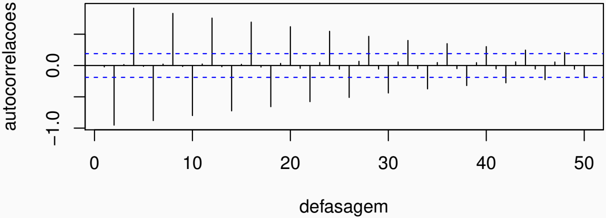
- Na prática pode ser bem difícil distinguir entre um processo estacionário com memória longa (e.g. AR(1) com $\alpha \approx 1$) e um processo não estacionário homogêneo.
- Existe uma vasta literatura econométrica sobre testes de raíz unitária.
- Mais recentemente, modelos da classe ARFIMA (ou ARIMA fracionários) tem sido utilizados para analisar séries com memória longa.

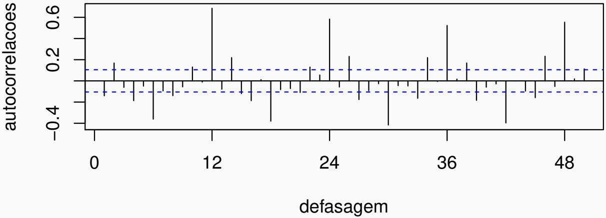
Correlogramas das primeiras diferenças: Totais mensais de passageiros em linhas aéreas e mortes por acidente nos EUA,



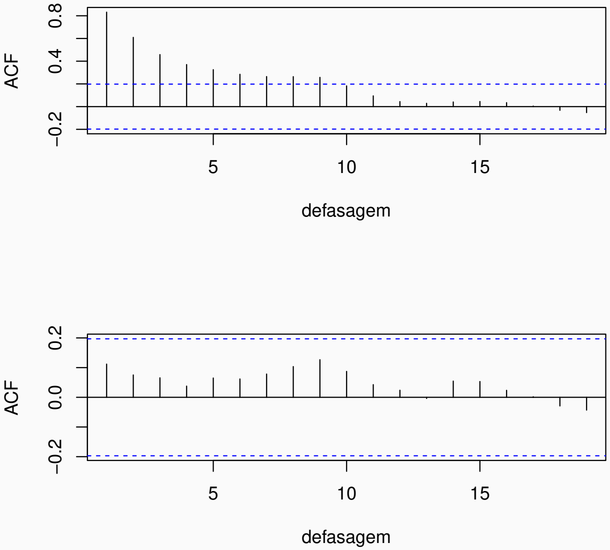


Correlogramas das primeiras diferenças: consumo de gás no Reino Unido e Produção Industrial do Brasil.

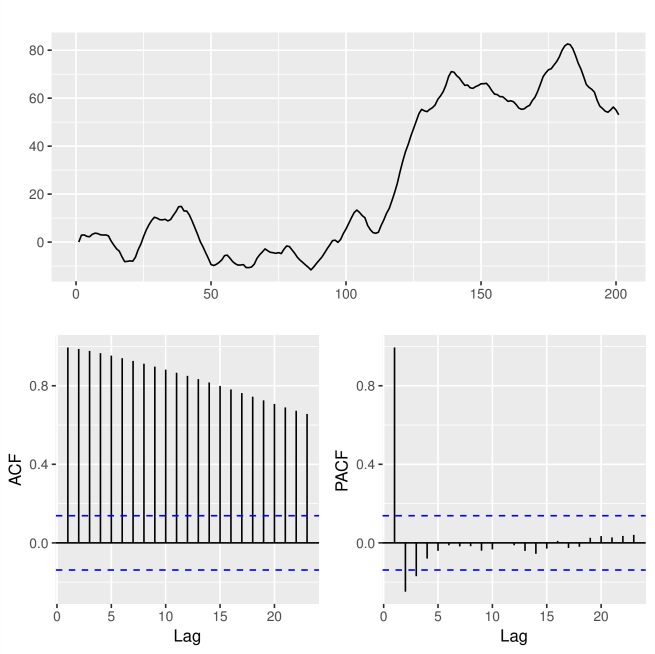




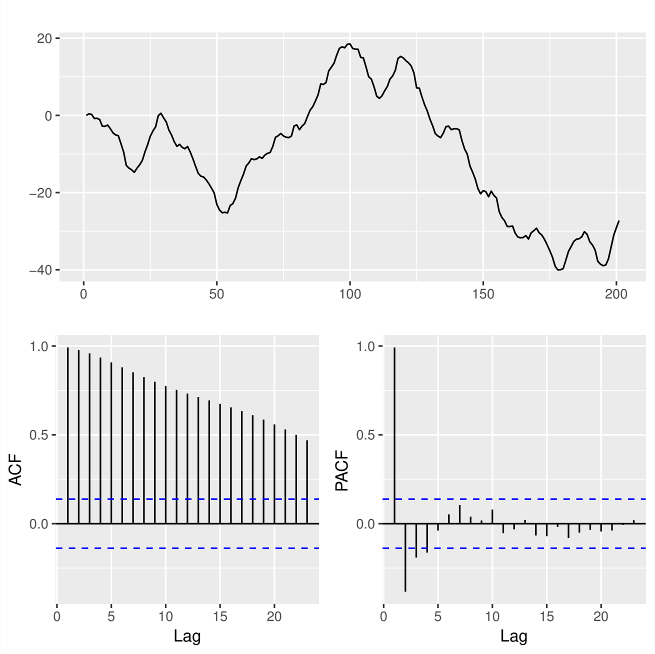
Correlogramas da série original de medições do nivel do Lago Huron e série contaminada com outlier.



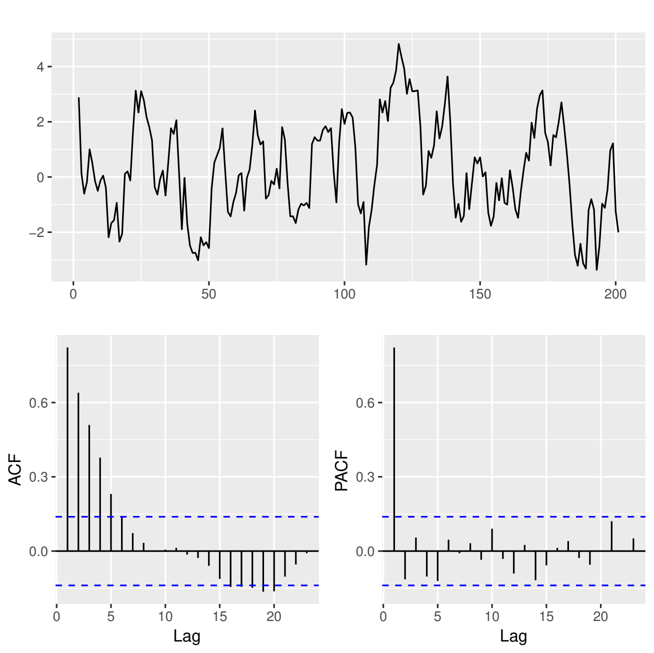
ARIMA(1,1,0), $\alpha = 0.8$.



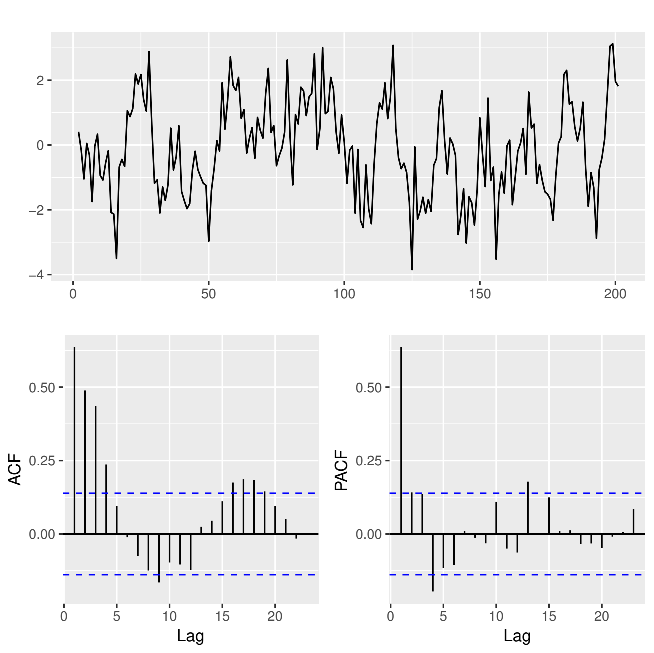
ARIMA(1,1,1), $\alpha = 0.8$, $\beta = 0.3$.



Primeiras diferenças, FAC e FACP (ARIMA(1,1,0), $\alpha = 0.8$).



Primeiras diferenças, FAC e FACP (ARIMA(1,1,1), $\alpha = 0.8$, $\beta = 0.3$).



Forma de choques aleatórios

Queremos escrever o processo ARIMA(p, d, q),

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_{p+d} X_{t-p-d} + \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}$$

como um processo MA(∞),

$$X_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \cdots = \psi(B)\epsilon_t$$

que pode ser reescrito como,

$$\begin{aligned} \varphi(B) X_t &= \varphi(B)\psi(B)\epsilon_t \\ \theta(B)\epsilon_t &= \varphi(B)\psi(B)\epsilon_t \\ \theta(B) &= \varphi(B)\psi(B) \end{aligned}$$

Os pesos ψ_j podem ser obtidos a partir das relações,

$$
\begin{aligned}
& (1 - \varphi_1 B - \cdots - \varphi_{p+d} B^{p+d})(1 + \psi_1 B + \psi_2 B^2 + \cdots) \\
= \; & (1 + \beta_1 B + \cdots \beta_q B^q)
\end{aligned}
$$

Exemplo. Seja o processo ARIMA(0,1,1),

$$(1-B)X_t = (1-\theta B)\epsilon_t.$$

Na representação MA(∞) temos que,

$$\begin{aligned}
(1-B)(1+\psi_1 B+\psi_2 B^2+\cdots) &= 1-\theta B\\
(1+\psi_1 B+\psi_2 B^2+\cdots)-(B+\psi_1 B^2+\psi_2 B^3+\cdots) &= 1-\theta B\\
1+(\psi_1-1)B+(\psi_2-\psi_1)B^2+\cdots &= 1-\theta B
\end{aligned}$$

Obtemos então que,

$$\psi_j = 1-\theta,\ j=1,2,\ldots$$

e o processo (não estacionário) é escrito como,

$$X_t = \epsilon_t + (1-\theta)\epsilon_{t-1} + (1-\theta)\epsilon_{t-2} + \cdots$$

Critérios de Informação

- Uma forma de "discriminar" entre estes modelos competidores é utilizar os chamados critérios de informação que levam em conta não apenas a qualidade do ajuste mas também penalizam a inclusão de parâmetros extras.
- Em muitas aplicações vários modelos podem ser julgados adequados.
- Um modelo com mais parâmetros pode ter um ajuste melhor mas não necessariamente será preferível em termos de critério de informação.
- A regra básica consiste em selecionar o modelo cujo critério de informação calculado seja mínimo.

- Suponha que existem k diferentes modelos $M_1, \ldots, M_k$ candidatos.
- Para cada modelo existe um vetor de parâmetros $\boldsymbol{\theta}_i \in \mathbb{R}^{n_i}$, i=1,...,k
- Para cada modelo existe uma função de verossimilhança $p(\mathbf{x}|\boldsymbol{\theta}_i, M_i)$.
- Como comparar estes modelos competidores?
- Como fazer inferência média usando todos os modelos (ou um subconjunto deles)?

Critério de informação de Akaike (AIC)

$$
\begin{aligned}
AIC &= -2\log \text{verossimilhança maximizada} + 2m \\
&= -2\log p(\mathbf{x}|\hat{\boldsymbol{\theta}}_i, M_i) + 2m
\end{aligned}
$$

Critério de informação Bayesiano (BIC)

$$
\begin{aligned}
BIC &= -2\log \text{verossimilhança maximizada} + m\log n \\
&= -2\log p(\mathbf{x}|\hat{\boldsymbol{\theta}}_i, M_i) + m\log n
\end{aligned}
$$

sendo m o número de parâmetros no modelo. E.g. num modelo ARMA(p, q) com erros normais, $m = p + q + 1$.

- Dado um conjunto de modelos, quanto menor o valor do critério melhor.
- Se a inclusão de termos adicionais no modelo não melhorar sensivelmente o ajuste, o critério terá valor maior.
- BIC penaliza bem mais a inclusão de parâmetros e tende a selecionar modelos mais parcimoniosos.
- Critérios de informação assumem valores na reta e não têm nenhum significado individualmente.
- Para comparar modelos a estimação deve ser feita no mesmo período amostral.
- Critérios de informação podem ser usados para comparar modelos não encaixados.

Identificação Revisitada

- Dado um conjunto de modelos competidores usar os critérios como ferramentas de identificação.
- E.g. estimar modelos ARMA(p, q) variando os valores de p e q.
- Após selecionar um modelo fazer análise de residuos (checar as hipóteses do modelo).
- Pode haver dois ou mais modelos com AIC e/ou BIC muito similares não sendo trivial discriminar entre eles.

Pesos AIC/BIC

A comparação fica mais fácil usando os desvios em relação ao valor minimo.

Pesos AIC

$$
\begin{aligned}
w_i &\propto \frac{\exp(-\frac{1}{2}AIC_i)}{\exp(-\frac{1}{2}AIC_{\min})} \\
&\propto \exp\left\{-\frac{1}{2}(AIC_i - AIC_{\min})\right\} \\
&\propto \exp\left(-\frac{1}{2}\Delta AIC_i\right)
\end{aligned}
$$

- Pesos normalizados,

$$w_i^* = \frac{w_i}{\sum_{j=1}^{M} w_j}, \quad i = 1, \ldots, M.$$

- Agora temos que $0 < w_k^* < 1$ e $\sum_{j=1}^{M} w_j^* = 1$.