ARMA modely I. - autoregresné modely (AR), časť 1

Beáta Stehlíková

2-PMS-101 Časové rady

Fakulta matematiky, fyziky a informatiky, UK v Bratislave

ARMA modely: plán prednášok

- Terminológia:
  - AR - autoregresný model - tieto slajdy
  - MA - kĺzavé priemery, *moving average*
  - ARMA - ich kombinácia
- Najskôr: autoregresný model prvého rádu, AR(1)
  - definícia
  - podmienky stacionarity
  - výpočet momentov a ACF
  - simulované dáta
  - praktický príklad s reálnymi dátami
- Potom:
  - autoregresné procesy vyšších rádov
  - ako určiť vhodný rád procesu pre dané dáta
- V ďalších slajdoch: MA a ARMA modely

Autoregresný proces prvého rádu - AR(1)

Definícia, podmienky stacionarity, výpočet momentov a autokorelačnej funkcie (ACF)

Rekurentná definícia a explicitné vyjadrenie

- AR(1) proces
$$x_t = \delta + \alpha x_{t-1} + u_t,$$
kde δ, α sú konštanty a $\{u_t\}$ je biely šum
- Nech pre $t = t_0$ je daná hodnota x_{t_0}:

$$\begin{aligned}
x_{t_0+1} &= \delta + \alpha x_{t_0} + u_{t_0+1}, \\
x_{t_0+2} &= \delta + \alpha x_{t_0+1} + u_{t_0+2} = \\
&\quad \delta(1+\alpha) + \alpha^2 x_{t_0} + (\alpha u_{t_0+1} + u_{t_0+2}) \\
x_{t_0+3} &= \dots
\end{aligned}$$

- Vo všeobecnosti:
$$x_{t_0+\tau} = \frac{1-\alpha^\tau}{1-\alpha}\delta + \alpha^\tau x_{t_0} + \sum_{j=0}^{\tau-1} \alpha^j u_{t_0+\tau-j}$$

AR(1) - stacionarita

- Prepíšeme si explicitné vyjadrenie do tvaru

$$x_t = \frac{1-\alpha^{t-t_0}}{1-\alpha}\delta + \alpha^{t-t_0} x_{t_0} + \sum_{j=0}^{t-t_0-1} \alpha^j u_{t-j}$$

- *Deterministická začiatočná podmienka*
  - stredná hodnota závisí od začiatočnej podmienky x_{t_0} $\rightarrow$ proces nie je stacionárny
- *Náhodná začiatočná podmienka*
  - proces je generovaný aj pred začiatkom našich pozorovaní $\rightarrow$ naša prvá pozorovaná hodnota je náhodná
  - ak $-1 < \alpha < 1$, tak pre $t_0 \rightarrow -\infty$ dostaneme

$$x_t = \frac{1}{1-\alpha}\delta + \sum_{j=0}^{\infty} \alpha^j u_{t-j}$$

  - to je Woldova reprezentácia s $\psi_j = \alpha^j$ $\rightarrow$ **stacionarita**

Stredná hodnota

- Ďalej pracujeme so stacionárnym procesom, teda $-1 < \alpha < 1$
- Pripomeňme si explicitné vyjadrenie procesu:

$$x_t = \frac{1}{1-\alpha}\delta + \sum_{j=0}^{\infty} \alpha^j u_{t-j}$$

- Stredná hodnota:

$$\begin{aligned}
\mathbb{E}(x_t) &= \mathbb{E}\left(\frac{1}{1-\alpha}\delta + \sum_{j=0}^{\infty} \alpha^j u_{t-j}\right) \\
&= \frac{1}{1-\alpha}\delta + \sum_{j=0}^{\infty} \alpha^j \mathbb{E}(u_{t-j}) = \frac{1}{1-\alpha}\delta
\end{aligned}$$

- Teda vo všeobecnosti $\mathbb{E}(x_t) \neq \delta$ (rovnosť je len pre $\delta = 0$), ale $\mathbb{E}(x_t)$ a δ majú rovnaké znamienko (lebo $|\alpha| < 1$)

Disperzia

$$\begin{aligned}\mathbb{D}(x_t) &= \mathbb{D}\left(\frac{1}{1-\alpha}\delta + \sum_{j=0}^{\infty}\alpha^j u_{t-j}\right) \\ &= \sum_{j=0}^{\infty}\mathbb{D}\left(\alpha^j u_{t-j}\right) = \sum_{j=0}^{\infty}\alpha^{2j}\mathbb{D}\left(u_{t-j}\right) = \frac{\sigma^2}{1-\alpha^2},\end{aligned}$$

kde

- sme využili, že disperzia súčtu nekorelovaných náhodných premenných je súčet ich disperzií
- σ^2 je disperzia bieleho šumu $\{u_t\}$

Autokovariancie

$$
\begin{aligned}
\mathrm{Cov}(x_t, x_{t-s}) &= \mathbb{E}\left[\left(\sum_{i=0}^{\infty} \alpha^i u_{t-i}\right)\left(\sum_{j=0}^{\infty} \alpha^j u_{t-s-j}\right)\right] \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \alpha^{i+j}\mathbb{E}\left(u_{t-i}u_{t-s-j}\right) \\
&= \sigma^2 \sum_{j=0}^{\infty} \alpha^{s+2j} = \alpha^s \frac{\sigma^2}{1-\alpha^2},
\end{aligned}
$$

kde sme využili, že

- $\mathrm{Cov}(u_k, u_l) = 0$ pre $k \neq l$
- $\mathrm{Cov}(u_k, u_l) = \sigma^2$ pre $k = l$

Autorelácie

- Autokorelačná funkcia AR(1) procesu teda je

$$\mathrm{Cor}(x_t, x_{t-s}) = \frac{\mathrm{Cov}(x_t, x_{t-s})}{\sqrt{\mathbb{D}(x_t)}\sqrt{\mathbb{D}(x_{t-s})}} = \alpha^s$$

- Napríklad pre proces $x_t = 10 + 0.4x_{t-1} + u_t$ je ACF rovná 0.4^s; numericky prvé členy:

```
## [1] 0.40000 0.16000 0.06400 0.02560 0.01024 0.00410
```

- *Otázka na opakovanie:* Aká je stredná hodnota tohto procesu?

Simulované dáta

Postup

- Budeme pracovať s AR(1) procesom
$$x_t = \delta + \alpha x_{t-1} + u_t,$$
kde $\delta = 0$ a $\{u_t\}$ je biely šum s normálnym rozdelením a disperziou 10.
- Parameter $\alpha \in (-1, 1), \alpha \neq 0$ zoberieme postupne z množiny $\{0.9, 0.5, -0.9\}$ - uvidíme vplyv znamienka a absolútnej hodnoty
- Zobrazíme:
  - realizáciu procesu dĺžky 250 (funkcia `arima.sim` z balíka `stats`)
  - odhadnutú ACF z vygenerovaných dát - prvých 10 hodnôt (už poznáme funkciu `acf`)
  - presnú ACF - takisto prvých 10 hodnôt (máme odvodený vzorec)

Prípad 1: $\alpha = 0.9$ - simulácia

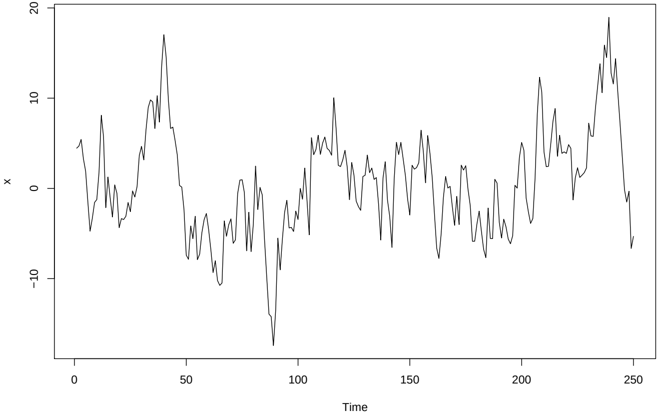
```
set.seed(123) # kvoli reprodukovatelnosti
x <- arima.sim(model = list(ar = c(0.9)),
               n = 250, sd = sqrt(10))
```

Poznámky:

- `model` je typu `list`, obsahuje vektory `ar` a `ma` členov (zatiaľ máme len jeden AR člen)
- `n` je dĺžka časového radu
- `sd` je štandardná odchýlka bieleho šumu (defaultne `sd = 1`)

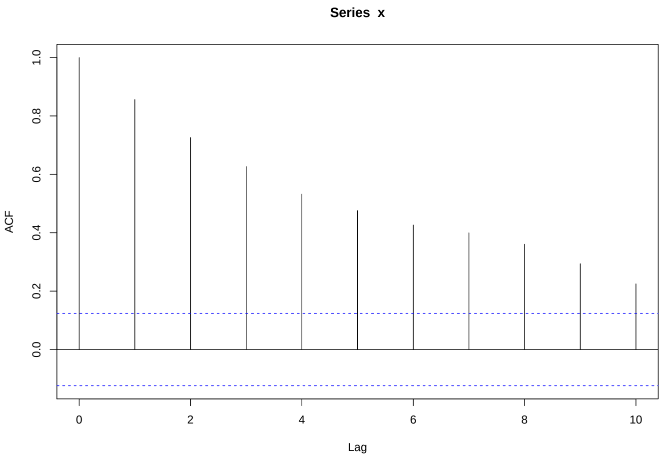
Prípad 1: $\alpha = 0.9$, priebeh

```
plot(x)
```



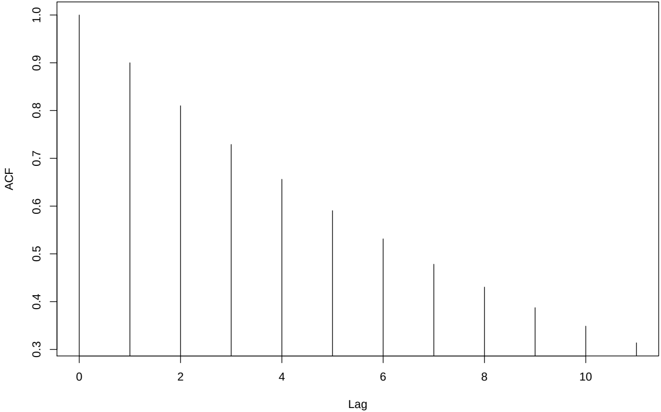
Prípad 1: $\alpha = 0.9$, odhadnutá ACF z dát

```
acf(x, lag.max = 10)
```



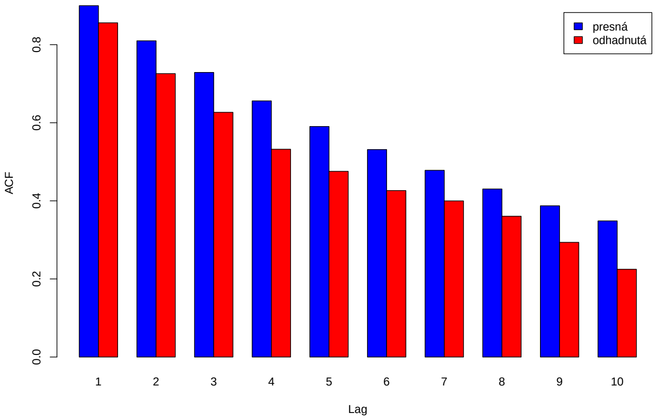
Prípad 1: $\alpha = 0.9$, presná ACF

```
plot(0:11, 0.9^(0:11), type = "h", xlab = "Lag", ylab = "AC
```

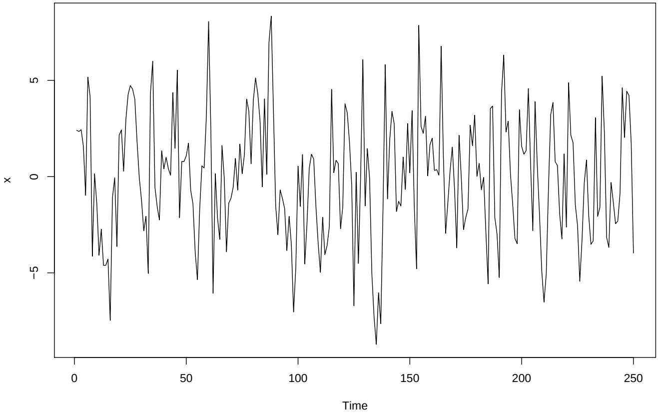


Cvičenie: Práca v R-ku

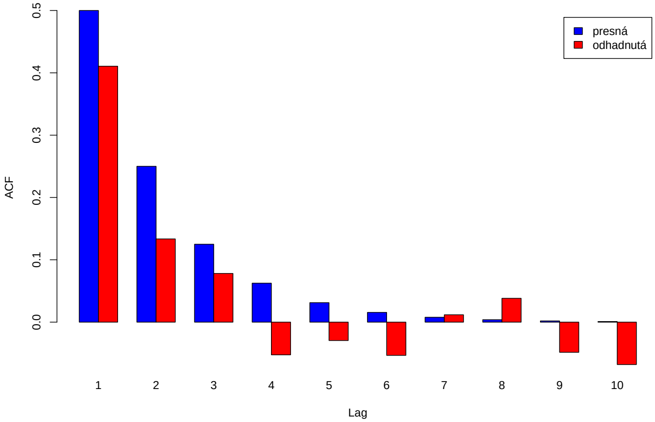
Porovnajte graficky presnú a odhadnutú ACF, pričom vynecháte lag 0 (zbytočný - korelácia so sebou je rovná vždy 1)



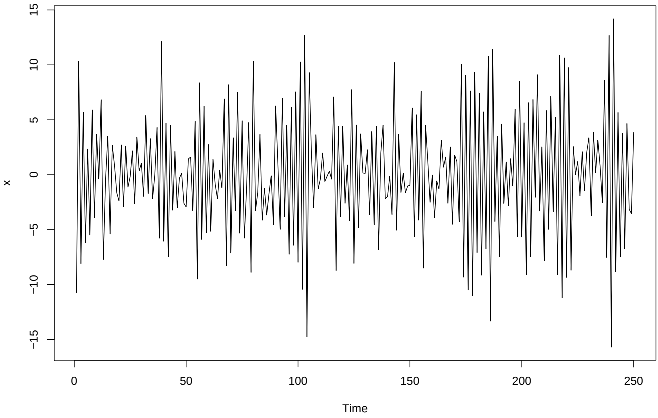
Prípad 2: $\alpha = 0.5$, priebeh



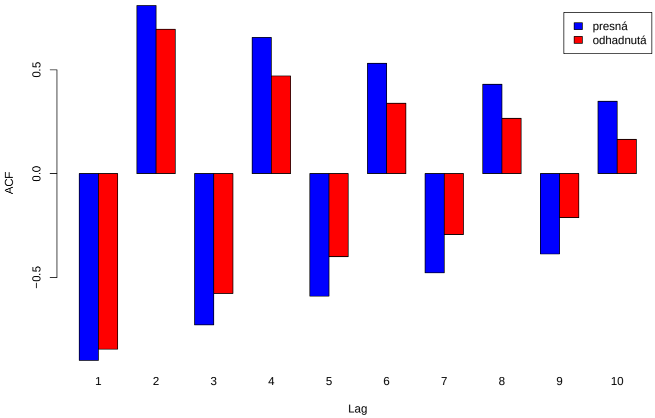
Prípad 2: $\alpha = 0.5$, odhadnutá a presná ACF



Prípad 3: $\alpha = -0.9$, priebeh



Prípad 3: $\alpha = -0.9$, odhadnutá a presná ACF



Cvičenie: Proces s nenulovou strednou hodnotou

Cvičenie 1. Proces $x_t = \delta + 0.9x_{t-1} + u_t$ simulujeme nasledovným kódom:

```
x <- 10 + arima.sim(model = list(ar = c(0.9)), n = 50)
```

Vyberte správnu hodnotu δ :

- $\delta = 10$
- $\delta = 10 \times (1 - 0.9) = 1$
- $\delta = \frac{10}{1-0.9} = 100$

Cvičenie 2. Vygenerujte simuláciu procesu $x_t = -1 + 0.6x_{t-1} + u_t$

Odhadovanie modelu v R-ku

Funkcia sarima z balíka astsa

- Na odhadovanie modelu použijeme funkciu `sarima` v tvare:

```
# AR(1) model pre k-te diferencie
model <- sarima(data, 1, k, 0, details = FALSE)
```

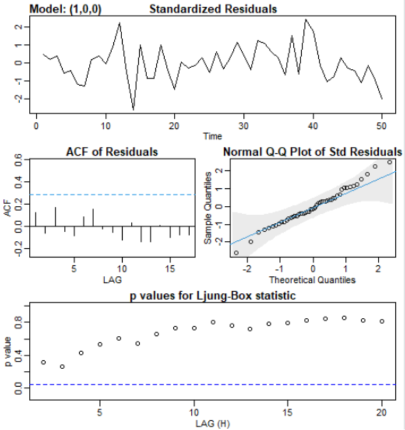
- Napríklad pre simulované dáta:

```
# vygenerujeme simulaciu AR(1) procesu
set.seed(123)
x <- 10 + arima.sim(model = list(ar = c(0.9)), n = 50)

# odhadneme pre ziskane data AR(1) model
library(astsa)
model <- sarima(x, 1, 0, 0, details = FALSE)
```

Kontrola rezíduí

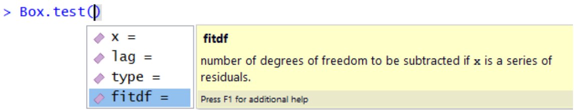
Čo znázorňuje ACF a interval na nej? Čo testuje Ljung-Boxov test? S akými výsledkami?



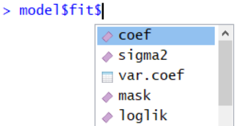
Všimnime si, že LB test začína pri lagu 2 (a nie 1) a pripomeňme si zo slajdov o LB teste: "Počet stupňov voľnosti sa zmení, ak ide o rezíduá z modelu". O čo ide: Počet stupňov voľnosti sa zníži o počet AR (a neskôr aj MA) členov modelu.

Ljung-Boxov test pre rezíduá

- Funkcia `Box.test` obsahuje parameter `fitdf`, ktorý zabezpečí správny počet stupňov voľnosti



- Pomocou `str(model)` si pozrieme štruktúru objektu `model`, aby sme vedeli pristupovať k jeho zložkám, napríklad `model$fit$residuals` (časový rad rezíduí)
- Pomoc *R Studia*:



- Máme AR(1) model, testujme na ukážku pre jeho rezíduá hypotézu $\rho(1) = \rho(2) = \rho(3) = \rho(4) = 0$:

```
Box.test(model$fit$residuals,
         lag = 4,     # testujeme 4 autokor.
         type = "Ljung-Box",
         fitdf = 1)  # jeden AR koeficient
```

```
##
##  Box-Ljung test
##
## data:  model$fit$residuals
## X-squared = 2.7667, df = 3, p-value = 0.429
```

- Môžeme porovnať s výstupom z funkcie `sarima` aj s tým, čo by vyšlo, keby sme zabudli na parameter `fitdf`.

Ďalšie zložky odhadnutého modelu

```
model$BIC # Bayesovo informacne kriterium
```

```
## [1] 2.86438
```

```
model$ttable   # odhady, SE, t statistky, p hodnoty
```

```
##       Estimate     SE t.value p.value
## ar1     0.8671 0.0704 12.3209       0
## xmean  10.5917 0.8525 12.4249       0
```

```
model$fit$coef # odhadnute parametre ako vektor
```

```
##      ar1    xmean
##  0.86707 10.59174
```

Zápis odhadnutého modelu: z vektora parametrov `model$fit$coef` vidíme, že odhadnutý model je

$$x_t = \delta + \alpha x_{t-1} + u_t,$$

kde α je parameter `ar1` (0.86707) a δ je taká, že stredná hodnota procesu $\mathbb{E}(x_t)$ je rovná parametru `xmean` (10.59174).

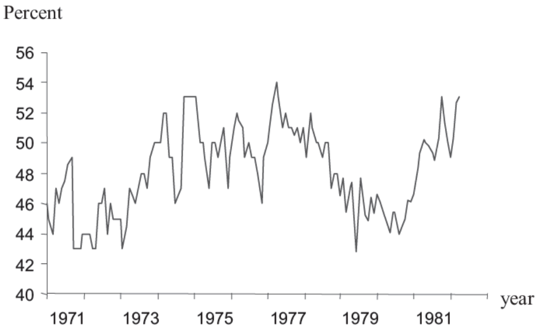
Cvičenia:

- Dopočítajte hodnotu parametra δ pomocou uvedených zaokrúhlených hodnôt
- Dopočítajte hodnotu parametra δ pomocou prístupu k presným hodnotám odhadnutých parametrov `xmean` a `ar1`.

Reálne dáta: Volebné preferencie v Nemecku

Dáta

- Nemecko, január 1971 - apríl 1982
- CDU_t - volebné preferencie CDU/CSU



Prebraté z učebnice *Kirchgässner & Wolters*, example 2.2

Citovaný pôvodný zdroj dát: G. Kirchgässner: *Causality Testing of the Popularity Function: An Empirical Investigation for the Federal Republic of Germany, 1971-1982*, Public Choice 45 (1985), p. 155-173.

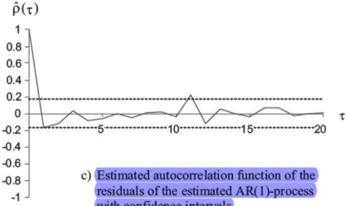
Odhadnutý AR(1) model

V knihe sa píše:

$$CDU_t = \underset{(3.43)}{8.053} + \underset{(17.10)}{0.834}\, CDU_{t-1} + \hat{u}_t,$$

$$\bar{R}^2 = 0.683, \quad SE = 1.586, \quad Q(11) = 12.516 \ (p = 0.326).$$

The estimated t values are given in parentheses. The autocorrelogram, which is also given in *Figure 2.4*, does not indicate any higher-order process. Moreover, the Box-Ljung Q Statistic with 12 correlation coefficients (i.e. with 11 degrees of freedom) gives no reason to reject this model.



c) Estimated autocorrelation function of the residuals of the estimated AR(1)-process with confidence intervals

Odhadnutý AR(1) model - otázky

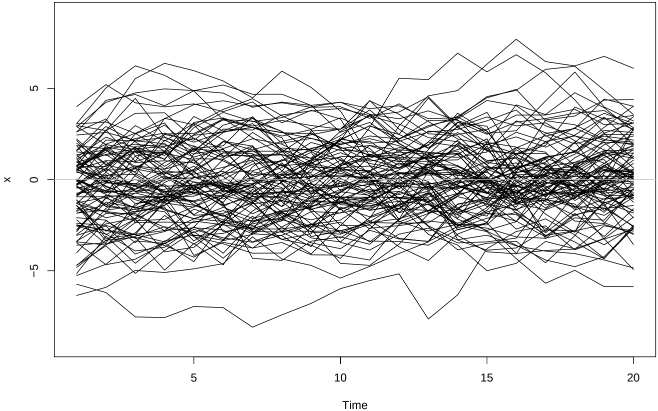
- *Je odhadnutý model stacionárny?* Z čoho to vyplýva?
- *Rezíduá modelu by mali byt' bielym šumom:*
  - Na grafe sú pri autokoreláciách zostrojené intervaly. Na čo slúžia? Vypočítajte pomocou známych údajov ich hranice.
  - V texte sa spomínajú autokorelácie rezíduí a Ljung-Boxova Q štatistika - aké hypotézy sa testujú (a prečo), akým spôsobom, s akými závermi?
- Čomu sa rovná *stredná hodnota* premennej CDU_t?

Pripomeňme si znovu zo slajdov o LB teste: *"Počet stupňov voľnosti sa zmení, ak ide o rezíduá z modelu"* aj vysvetlenie: *Počet stupňov voľnosti sa zníži o počet AR (a neskôr aj MA) členov modelu.*

Predikcie

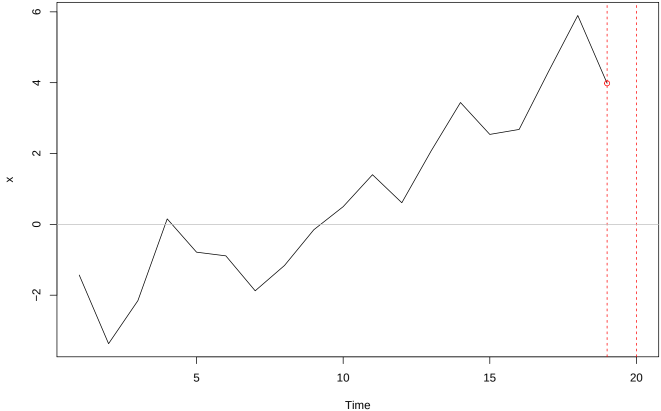
Podmienená a nepodmienená stredná hodnota (simulácie)

- Generujeme proces $x_t = 0.9x_{t-1} + u_t$ a zaujíma nás očakávaná hodnota v čase 20 - *je nulová*

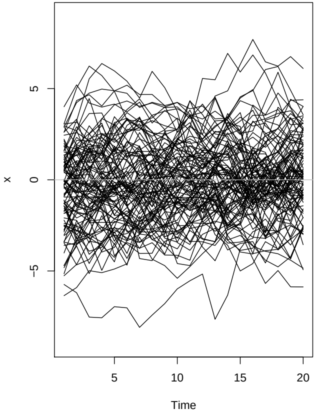


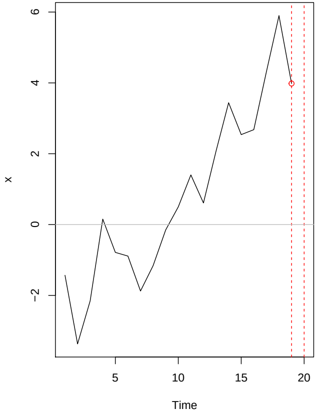
Podmienená a nepodmienená stredná hodnota (simulácie)

- Ak už máme prvých 19 hodnôt a pýtame sa na očakávanú hodnotu v čase 20 - *je to iná situácia*



Podmienená a nepodmienená stredná hodnota (simulácie)





- Vľavo: *nepodmienená* stredná hodnota procesu
- Vpravo: *podmienená* stredná hodnota procesu (podmienená doterajším priebehom) - toto nás zaujíma pri **predikciách**

Podmienená a nepodmienená stredná hodnota (dáta)

- Máme stacionárny proces

$$x_t = 8.053 + 0.834 x_{t-1} + u_t$$

ako model pre volebné preferencie $x_t := CDU_t$

- Vieme nájsť *nepodmienenú strednú hodnotu procesu* - je samozrejme konštantná
- Môžeme sa však pýtať na **predikcie**:
  - Aká je očakávaná hodnota preferencií budúci mesiac, ak terajšie preferencie sú 40 percent?
  - Aká je očakávaná hodnota preferencií budúci mesiac, ak terajšie preferencie sú 55 percent?
- Odpovede budú **rôzne**. Pri týchto otázkach hľadáme *podmienenú strednú hodnotu*.

Intuitívne postup

- Pri AR modeloch zostaneme pri intuitívnom postupe (presnejšie a formálnejšie potom pri tých modeloch, kde postup konštrukcie predikcií nebude zrejmý)
- Pripomeňme si, že pre $x_t := CDU_t$ máme model

$$x_t = 8.053 + 0.834 x_{t-1} + u_t$$

- Pri predikciách biely šum u_t nahradíme jeho strednou hodnotou - nulou
- Za x_{t-1} dosadíme
  - skutočnú hodnotu x_{t-1}, ak ju máme k dispozícii
  - predikciu hodnoty x_{t-1}, ak sa ešte nerealizovala

Numerická realizácia

- Postup je dobre viditeľný pri použití tabuľkového editora:

	A	B
1	delta	8.053
2	alfa	0.834
3		
4	40	
5	=$B$1+$B$2*A4	

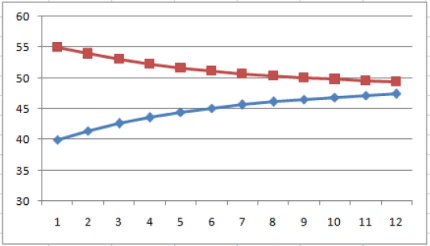
	A	B
1	delta	8.053
2	alfa	0.834
3		
4	40	
5	41.413	
6	=$B$1+$B$2*A5	

	A	B
1	delta	8.053
2	alfa	0.834
3		
4	40	
5	41.413	
6	42.59144	
7	43.57426	
8	44.39394	
9	45.07754	
10	45.64767	
11	46.12316	
12	46.51971	
13	46.85044	
14	47.12627	
15	=$B$1+$B$2*A14	

Numerická realizácia

- Predikcie pre začiatočné hodnoty 40 a 55 percent:

	A	B
1	delta	8.053
2	alfa	0.834
3		
4	40	55
5	41.413	53.923
6	42.59144	53.02478
7	43.57426	52.27567
8	44.39394	51.65091
9	45.07754	51.12986
10	45.64767	50.6953
11	46.12316	50.33288
12	46.51971	50.03062
13	46.85044	49.77854
14	47.12627	49.5683



- Konvergujú k spoločnej hodnote, ktorá sa rovná nepodmienenej strednej hodnote procesu
- Prakticky - treba si zvážiť, na aké dlhé obdobie má zmysel použiť model pri predikovaní

V R-ku: funkcia sarima.for z balíka astsa

- Naše simulované dáta:

```
set.seed(123)
x <- 10 + arima.sim(model = list(ar = c(0.9)), n = 50)
```

- Najskôr odhadneme a otestujeme model pomocou funkcie sarima:

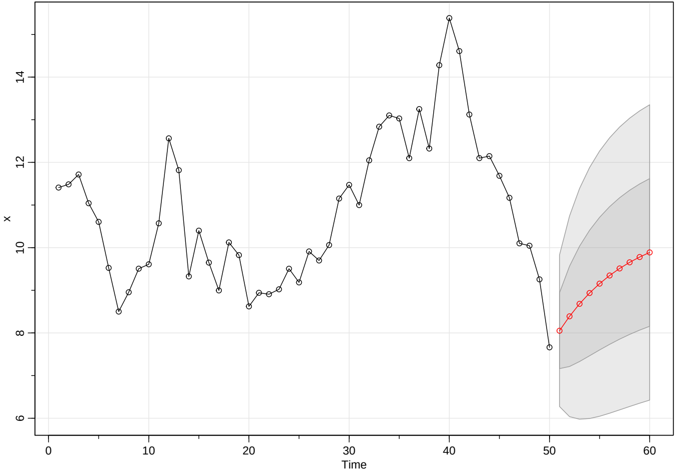
```
sarima(x, 1, 0, 0) # odhadli sme AR(1) model - je OK
```

- Z dobrého modelu môžeme robiť predikcie:

```
sarima.for(x, n.ahead = 10, 1, 0, 0) # predikcie
                                      # 10 pozorovani
sarima.for(x, 10, 1, 0, 0) # to iste (treba dat pozor na
                           # spravne poradie parameterov)
```

Predikcie a intervaly spoľahlivosti (+/- 1 a 2 štandardné odchýlky):

```
sarima.for(x, n.ahead = 10, 1, 0, 0)
```



Autoregresný proces vyššieho rádu - AR(p)

Motivácia - prečo nestačí AR(1)

Prečo nestačí AR(1) - niekoľko pohľadov

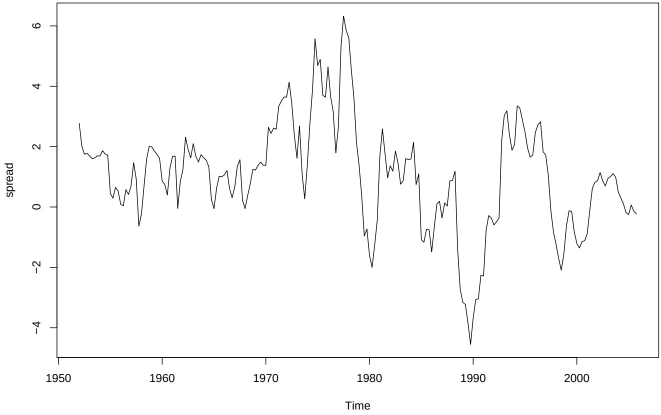
- Celkom prirodzene môžeme očakávať, že na dobré popísanie vývoja x_t nám nebude stačíť x_{t-1} , ale budeme potrebovať aj x_{t-2} (prípadne aj x_{t-3} a x_{t-4} a pod.)
- AR(1) proces má dosť obmedzené možnosti pri zachytení priebehu ACF - napríklad neumožňuje modelovať periodický charakter
- Niekedy to nemusí byť dopredu zrejmé z dát, ani z odhadnutej ACF, ale AR(1) nebude vyhovovať kvôli rezíduám - *toto uvidíme na príklade*

Príklad: Úrokové miery

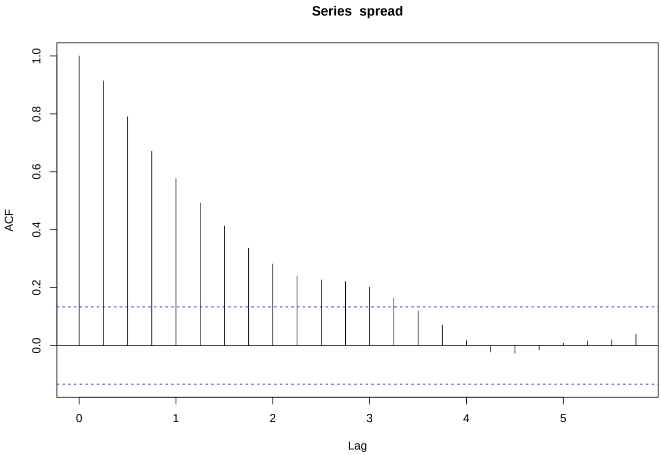
- Štvrťročné dáta, 1952Q1 - 2005Q4
- Premenné:
  - krátkodobá úroková miera (3 mesiace)
  - dlhodobá úroková miera (20 rokov)
- Budeme modelovať *spread*, teda rozdiel dlhodobej a krátkodobej úrokovej miery

Mills, Markellos: *The Econometric Modelling of Financial Time Series.* Cambridge University Press, 2008

Príklad: Úrokové miery - priebeh dát



Príklad: Úrokové miery - odhadnutá ACF



Podobá sa na AR(1) proces s kladným parametrom α.

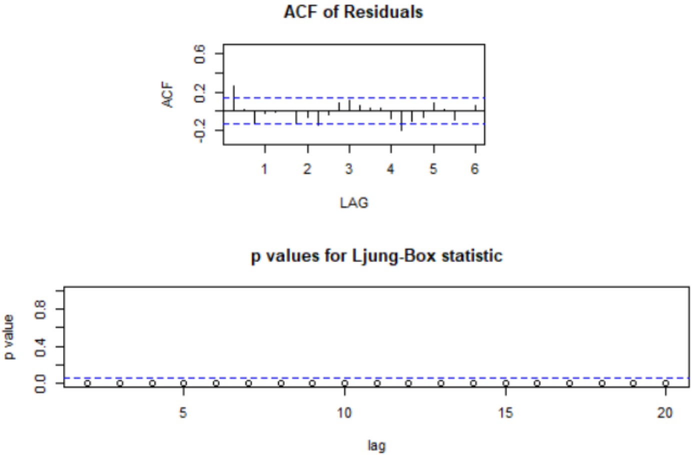
Príklad: Úrokové miery - parametre AR(1) modelu

Parameter α AR(1) modelu (vo výstupe označený ako `ar1`) je medzi -1 a 1, teda získaný proces je stacionárny - toto je ok.

```
##         Estimate     SE t.value p.value
## ar1       0.9156 0.0266 34.4589  0.0000
## xmean     1.0473 0.5491  1.9075  0.0578
```

Príklad: Úrokové miery - rezíduá AR(1) modelu

Rezíduá sa nesprávajú ako biely šum, model je nevyhovujúci.



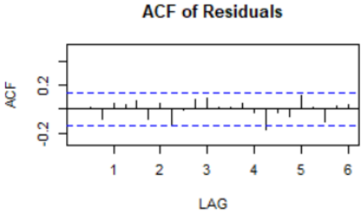
Príklad: Úrokové miery - parametre AR(2) modelu

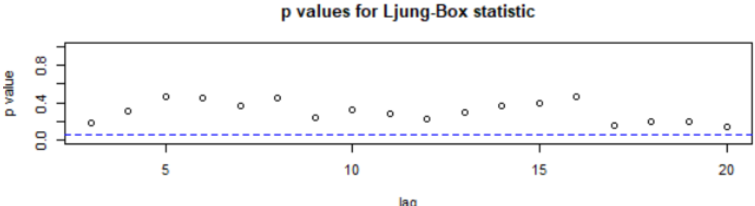
Zatiaľ sa na ne pozrime, analyzovať ich budeme vedieť o chvíľu.

```
##        Estimate     SE t.value p.value
## ar1      1.1809 0.0650 18.1566  0.0000
## ar2     -0.2886 0.0651 -4.4321  0.0000
## xmean    1.0449 0.4212  2.4807  0.0139
```

Príklad: Úrokové miery - rezíduá AR(2) modelu

Vieme však už teraz zhodnotiť rezíduá - tie sú v poriadku. AR(2) ako model je dobrý.





Definícia autoregresného procesu vyššieho rádu

AR(2) a všeobecný AR(p) proces

- **AR(2) proces** modeluje x_t pomocou x_{t-1} a x_{t-2}:

  $$x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + u_t$$

  Analogicky, **AR(p) proces** modeluje x_t pomocou p predchádzajúcich hodnôt $x_{t-1}, \ldots, x_{t-p}$:

  $$x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + u_t$$

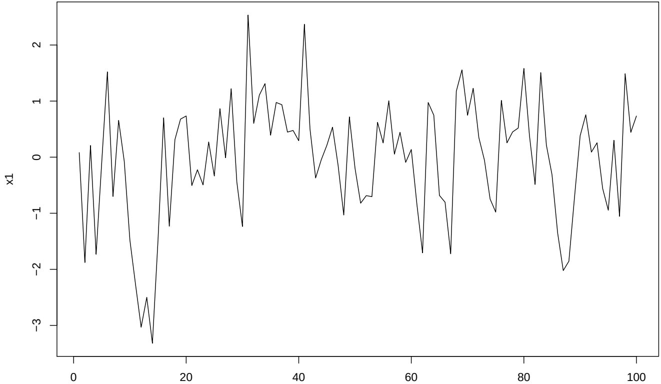
- Odhadovanie AR(p) modelu pre k-te diferencie v R-ku:

```
sarima(data, p, k, 0)
```

Simulácie AR(2) procesov

Príklad 1. Proces $x_t = 0.5x_{t-1} + 0.2x_{t-2} + u_t$

```
x1 <- arima.sim(model = list(ar = c(0.5, 0.2)), n = 100)
```



Príklad 2. Proces $x_t = 0.5x_{t-1} + 0.6x_{t-2} + u_t$ spôsobí chybu - nie je stacionárny

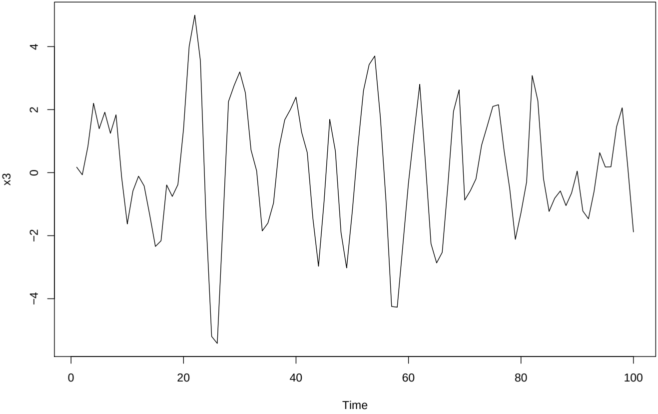
```
> x2 <- arima.sim(model = list(ar = c(0.5, 0.6)), n = 100, sd = 1)
Error in arima.sim(model = list(ar = c(0.5, 0.6)), n = 100, sd = 1) :
  'ar' part of model is not stationary
```

- NEPLATÍ teda "zovšeobecnenie" z AR(1) prípadu, že ak je α_1, α_2 v absolútnej hodnote menej ako 1, tak je proces $x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + u_t$ stacionárny.

Príklad 3. Proces $x_t = 1.2x_{t-1} - 0.8x_{t-2} + u_t$ chybu nespôsobí - je stacionárny

- NEPLATÍ teda ani také "zovšeobecnenie", že ak je niektorá z hodnôt α_1, α_2 v absolútnej hodnote väčšia ako 1, tak je proces $x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + u_t$ nestacionárny.

```
x3 <- arima.sim(model = list(ar = c(1.2, -0.8)), n = 100)
plot(x3)
```



- Nemôžeme však čakať také jednoduché explicitné vyjadrenie AR(p) procesu ako pre AR(1), lebo je zložitejšie aj bez bieleho šumu (diferenčná rovnica) $\rightarrow$ **proces prepíšeme inak, aby sa s ním lepšie pracovalo** $\rightarrow$ definujeme tzv. **operátor posunu**.

Operátor posunu

Operátor posunu (lag operator) L - vráti hodnotu procesu o jedno pozorovanie dozadu:

$$Lx_t = x_{t-1}$$

Niektoré vlastnosti:

- dajú sa robiť mocniny: $L^2 x_t = L(Lx_t) = x_{t-2}$, $L^3 x_t = x_{t-3}$ a pod.
- počítanie s mocninami: $L^2(L^3) = L^5$
- $L^0 = 1$ je identita
- násobenie: $(1 - 0.5L)(1 + 0.2L) = 1 - 0.3L - 0.1L^2$
- konštanta je vlastne konštatný proces, posunom sa nezmení: $(1 - 0.2L + 0.3L^2)c = c - 0.2c + 0.3c = 1.1c$

Definícia a podmienky stacionarity pre AR(2)

Rekurentná definícia a zápis pomocou operátora posunu

- Definíciu sme už videli:

$$x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + u_t$$

- Teraz proces prepíšeme pomocou operátora posunu L :

$$x_t = \delta + \alpha_1 L x_t + \alpha_2 L^2 x_t + u_t$$

a teda

$$(1 - \alpha_1 L - \alpha_2 L^2) x_t = \delta + u_t$$

Podmienky stacionarity

- Potrebujeme proces zapísať v tvare Woldovej reprezentácie
- Chceli by sme spraviť:

$$x_t = (1 - \alpha_1 L - \alpha_2 L^2)^{-1}\delta + (1 - \alpha_1 L - \alpha_2 L^2)^{-1} u_t$$

- Treba teda zistiť, kedy existuje inverzný operátor $(1 - \alpha_1 L - \alpha_2 L^2)^{-1}$ a čomu sa rovná

Podmienky stacionarity

- Použijeme metódu neurčitých koeficientov:

$$(1 - \alpha_1 L - \alpha_2 L^2)^{-1} = \psi_0 + \psi_1 L + \psi_2 L^2 + \dots$$

- Z toho:

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

- Porovnáme koeficienty pri L^j na oboch stranách:

$$\psi_0 = 1, \psi_1 = \alpha_1$$

$$\psi_j - \alpha_1 \psi_{j-1} - \alpha_2 \psi_{j-2} = 0$$

Podmienky stacionarity

- Podmienka stacionarity: Kvôli splneniu podmienky $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ musia byť korene charakteristickej rovnice
$$\lambda^2 - \alpha_1 \lambda - \alpha_2 = 0$$
v absolútnej hodnote menšie ako 1.
- Inak povedané (obvyklá formulácia v súvislosti s časovými radmi): korene rovnice
$$1 - \alpha_1 L - \alpha_2 L^2 = 0$$
musia byť v absolútnej hodnote väčšie ako 1, teda mimo jednotkového kruhu
- Všimnime si, že to isté vyšlo pre AR(1) proces $\alpha(L)x_t = \delta + u_t$, kde $\alpha(L) = 1 - \alpha L$ - koreň polynómu $\alpha(L)$ musí byť mimo jednotkového kruhu

Podmienky stacionarity v R-ku

- Funkcia `polyroot`:

A polynomial of degree *n - 1*,

*p(x) = z1 + z2 * x + … + z[n] * x^(n-1)*

is given by its coefficient vector `z[1:n]`. `polyroot` returns the *n-1* complex zeros of *p(x)*

Podmienky stacionarity v R-ku: príklad

Overíme stacionaritu procesu

$$x_t = 1.2 + 0.3x_{t-1} - 0.8x_{t-2} + u_t,$$

teda

$$(1 - 0.3L + 0.8L^2)x_t = 1.2 + u_t$$

Je stacionárny, lebo všetky absolútne hodnoty sú väčšie ako 1:

```
polyroot(c(1, -0.3, 0.8)) # korene
```

```
## [1] 0.1875+1.1022i 0.1875-1.1022i
```

```
abs(polyroot(c(1, -0.3, 0.8))) # abs. hodnoty
```

```
## [1] 1.118034 1.118034
```

AR(2) model pre spread: stacionarita

- Pripomeňme si výstup:

```
##        Estimate     SE t.value p.value
## ar1      1.1809 0.0650 18.1566  0.0000
## ar2     -0.2886 0.0651 -4.4321  0.0000
## xmean    1.0449 0.4212  2.4807  0.0139
```

- Teda model je

$$x_t = \delta + 1.1809 x_{t-1} - 0.2886 x_{t-2} + u_t$$

parameter δ je taký, aby platilo $\mathbb{E}(x_t) = 1.0449$

- Prepíšeme pomocou polynómu v L:

$$(1 - 1.1809L + 0.2886L^2)x_t = \delta + u_t$$

AR(2) model pre spread: stacionarita

- Nájdeme presné hodnoty koeficientov modelu

$$(1 - 1.1809L + 0.2886L^2)x_t = \delta + u_t :$$

- Pripomeňme si štruktúru objektu ar2 pomocou `str(ar2)`,potrebujeme vytiahnuť hodnoty koeficientov:

```
ar2$fit$coef
```

```
##       ar1        ar2      xmean
##  1.1808724 -0.2885893  1.0449355
```

```
polyroot(c(1, -1.1809, 0.2886))    # priblizne
polyroot(c(1, -ar2$fit$coef[1:2])) # presne
```

AR(2) model pre spread: stacionarita

- Rozdiel pri použití približných a presných koeficientov je malý, ale druhý postup sa dá používať automaticky, bez kopírovania výstupu

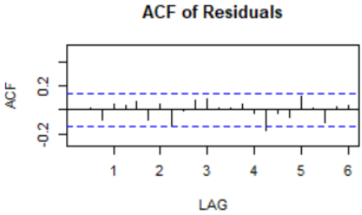
```
abs(polyroot(c(1, -ar2$fit$coef[1:2])))
```

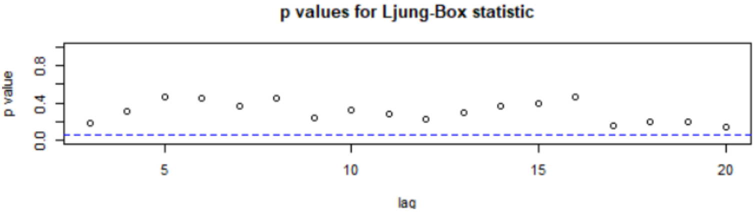
```
## [1] 1.196978 2.894901
```

- Absolútne hodnoty sú všetky väčšie ako 1 $\rightarrow$ **stacionarita**

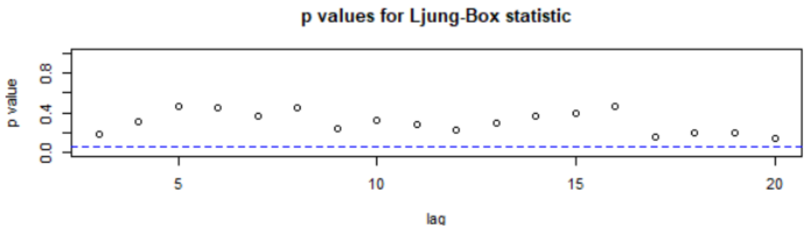
AR(2) model pre spread: rezíduá

V rezíduách nie je signifikantná autokorelácia , model je vyhovujúci:





AR(2) model pre spread: rezíduá



- Všimnime si, že LB štatistika začína od lagu 3
- Počet stupňov voľnosti je počet testovaných korelácií mínus 2 (tá 2 pochádza z toho, že máme AR(2) model)

Cvičenia

Cvičenie 1. Uvažujme proces $x_t = 5 - 0.4x_{t-1} + 0.1x_{t-2} + u_t$

- Ukážte, že je stacionárny.
- Odvoďte diferenčnú rovnicu a začiatočné podmienky pre koeficienty Woldovej reprezentácie. Vypočítajte rekurentne niekoľko prvých členov. Potom odvoďte explicitný predpis pre všeobecný člen.

```
# na kontrolu Woldova reprezentacia z R-ka
ARMAtoMA(ar = c(-0.4, 0.1), lag.max = 5)
```

```
## [1] -0.40000  0.26000 -0.14400  0.08360 -0.04784
```

Cvičenie 2. Zopakujte pre proces $x_t = 5 + 0.4x_{t-1} - 0.1x_{t-2} + u_t$

Definícia a podmienka stacionarity pre AR(p)

- AR(p) proces:

$$x_t = \delta + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + u_t, \tag{1}$$

teda $\alpha(L)x_t = \delta + u_t$, kde $\alpha(L) = 1 - \alpha_1 L - \cdots - \alpha_p L^p$

- Woldova reprezentácia a stacionarita:

$$x_t = \alpha(L)^{-1}(\delta + u_t),$$

inverzný operátor $\alpha(L)^{-1}$ hľadáme v tvare

$$\alpha(L)^{-1} = 1 + \phi_1 L + \phi_2 L^2 + \ldots$$

- Pre koeficienty ϕ_j dostaneme diferenčnú rovnicu

$$\phi_k - \alpha_1 \phi_{k-1} - \cdots - \alpha_p \phi_{k-p} = 0$$

$\Rightarrow$ kvôli konvergencii $\sum \phi_j^2$ musia byť **korene polynómu** $\alpha(L)$ **mimo jednotkového kruhu**

Príklad 1

- Proces $x_t = 0.2x_{t-1} + 0.3x_{t-1} - 0.6x_{t-3} + u_t$
- Je stacionárny, všetky korene $1 - 0.2L - 0.3L^2 + 0.6L^3$ sú mimo jednotkového kruhu

```
abs(polyroot(c(1, -0.2, -0.3, 0.6)))
```

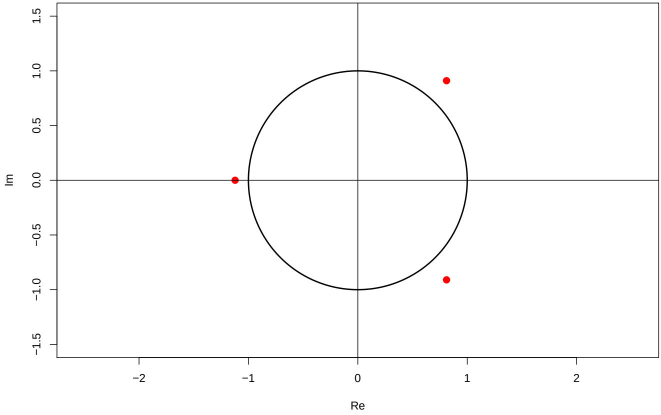
```
## [1] 1.218935 1.121728 1.218935
```

- Korene:

```
##          Re      Im    Mod
## 1  0.8109  0.9101 1.2189
## 2 -1.1217  0.0000 1.1217
## 3  0.8109 -0.9101 1.2189
```

Príklad 1

- Grafické znázornenie koreňov



Príklad 2

- Proces $x_t = 0.2x_{t-1} + 0.3x_{t-1} - 0.6x_{t-3} + 0.5x_{t-4} + u_t$
- Nie je stacionárny, jeden z koreňov polynómu $1 - 0.2L - 0.3L^2 + 0.6L^3 - 0.5L^4$ je vnútri jednotkového kruhu

```
abs(polyroot(c(1, -0.2, -0.3, 0.6, -0.5)))
```

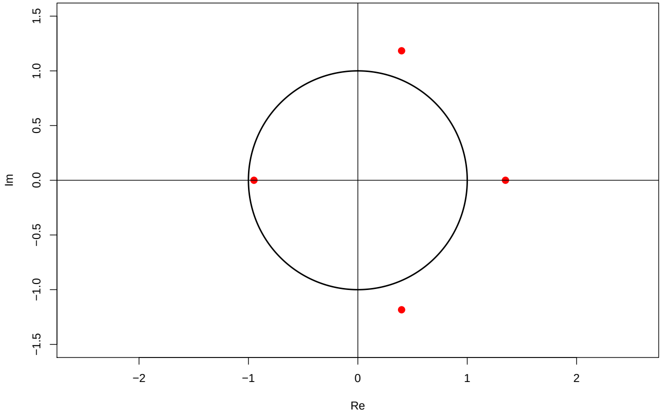
```
## [1] 1.2494352 0.9493448 1.2494352 1.3495174
```

- Korene:

```
##          Re      Im    Mod
## 1  0.3999  1.1837 1.2494
## 2 -0.9493  0.0000 0.9493
## 3  0.3999 -1.1837 1.2494
## 4  1.3495  0.0000 1.3495
```

Príklad 2

- Grafické znázornenie koreňov:



Užitočnosť pridávania ďalších AR členov

Predikovanie dopytu po elektrine:

Vu, D. H., Muttaqi, K. M., Agalgaonkar, A. P., & Bouzerdoum, A. (2016). **Intra-hour and hourly demand forecasting using selective order autoregressive model.** In 2016 IEEE International Conference on Power System Technology (POWERCON) (pp. 1-6). IEEE.

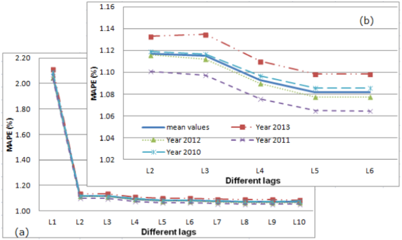


Fig. 5. Performance of the model with half hourly lags

When applying the autoregressive model given in (1) in load demand forecasting for the half hourly demand in NSW, the average performance of the model with different values of order p is recorded as shown in Fig. 5(a) and (b). It is noted that Fig. 5(b) is a zoom of Fig.5(a) from lag 2 to lag 6.

It can be seen from this Fig. 5(a) and (b) that the MAPE value reduces when adding more lag to the model until $P = 5$. After this critical value, the MAPE values get to the stationary state. Consequently, this critical value is selected as the order of the AR model for further development.