# AR(1) model: CVIČENIA

Beáta Stehlíková

2-PMS-10 Časové rady

Fakulta matematiky, fyziky a informatiky, UK v Bratislave

# Cvičenie 1: Opakovanie, odhadnutie AR(1) modelu

```
x <- window(Nile, start = 1905)
```

- Pre dáta o prietokoch Nílu sme Ljung-Boxovým testom zistili, že majú signifikantné autokorelácie. Nejde preto iba o biely šum posunutý o konštantu, ale treba modelovať závislosť pozorovaných hodnôt.
- Odhadnite pre tieto dáta AR(1) model a na základe Ljung-Boxovho testu pre rezíduá (ktorý je súčasťou výstupu) rozhodnite, či ide o dobrý model pre tieto dáta.
- Prečo sa Ljung-Boxov test začína až od testovania prvých dvoch autokorelácií (pri testovaní pôvodných dát sme testovali aj prvú koreláciu)?

# Cvičenie 2: Predikcie a ich porovnanie s dátami

- Vychajme z dát posledných 5 pozorovaní.
- Zo zostávajúcich dát odhadneme AR(1) model.
- Funkciou `sarima.for` spravíme predikcie pre nasledujúcich 5 rokov.
- Doplníme pre tieto časy reálne pozorované dáta (dá sa to napr. pomocou `lines` a nastavenia hodnôt na x-ovej osi)

- Pri kreslení grafov a pri výpočtoch je niekedy užitočné mať k dispozícii číselné hodnoty predikcií a štandardných odchýlok:

```
# AR(1) pre vsetky data x, 3 predikcie
predikcie <- sarima.for(x, n.ahead = 3, 1, 0, 0)

str(predikcie)  # pozrite si strukturu
predikcie$pred  # casovy rad predikcii
```

# Cvičenie 3: AR modely vyššieho rádu

- Analogicky ako AR(1) sa definuje AR(p):

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

- V balíku `astsa`

```
model <- sarima(data, p, k, 0) # AR(p) pre k-te dif.
model <- sarima(data, p, k, 0
```

- Odhadnite AR(p) pre $p \in \{1, 2, 3, 4, 5\}$ model pre dáta o prietokoch Nílu. Pre modely s dobrými rezíduami vyberte model s najnižším Bayesovym informačným kritériom.

```
model$BIC # ziskanie hodnoty BIC
```

## Cvičenie 4: Volebné preferencie v Nemecku

- Zrekonštruujeme (do istej miery) model z prednášky o volebných preferenciách v Nemecku.
- Dáta sú dostupné na webe (odkaz na stránke tohto predmetu), ale vo formáte `wf1` (pre softvér EViews). Načítanie do R spravíme pomocou balíka `hexView`

```
library(hexView)
data <- readEViews("chapter_2_example_2_2.wf1")
```

- Pozrite si štruktúru získaných dát a vytvorte na základe nich časový rad s pozorovaniami.
- Odhadnite AR(1) model pre tieto dáta a zhodnoťte jeho rezíduá

- Uložte do objektu ar1 výsledok odhadovania tohto modelu

```
ar1 <- sarima(...)

 # ak nechceme vypisat a vykreslit vystupy
ar1 <- sarima(..., details = FALSE)
```

```
str(ar1)
```

```
## List of 6
##  $ fit                      :List of 14
##   ..$ coef      : Named num [1:2] 0.835 48.212
##   .. ..- attr(*, "names")= chr [1:2] "ar1" "xmean"
##   ..$ sigma2    : num 2.48
##   ..$ var.coef : num [1:2, 1:2] 0.00227 0.00095 0.00095
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : chr [1:2] "ar1" "xmean"
```

- Napríklad odhady koeficientov:

```
ar1$fit$coef
```

```
##         ar1       xmean
##   0.8349926 48.2122716
```

```
ar1$fit$coef[1]     # prva zlozka vektora
```

```
##       ar1
## 0.8349926
```

```
ar1$fit$coef["ar1"] # zlozka s nazvom ar1
```

```
##       ar1
## 0.8349926
```

**Cvičenie.** Zapíšeme model v tvare $x_t = \delta + \alpha_1 x_{t-1} + u_t$.

- Koeficient $\alpha_1$ je vo výstupe označený ako `ar1`
- Koeficient $\delta$ dopočítame pomocou `xmean` a `ar1` (s presnými hodnotami, nie odpísanými zo zaokrúhleného výstupu)

- Rezíduá modelu

```
ar1$fit$residuals
```

```
## Time Series:
## Start = 1
## End = 137
## Frequency = 1
##   [1] -1.76758832  0.46995143 -1.36504117 -1.53004857
##   [7]  0.63495883  0.29996622  0.88246992  0.54747731
##  [13] -0.86006336  0.13993664 -0.69505596 -0.69505596
##  [19]  2.13993664 -0.36504117  0.63495883 -3.20003378
##  [25] -0.53004857 -0.53004857 -2.53004857  0.63993664
##  [31] -0.78253748  0.63495883  0.79996622 -0.03502639
##  [37]  1.12998101  0.29498840  0.29498840  2.29498840
##  [43]  0.12998101 -2.87001899  0.13495883  0.21746252
##  [49]  0.79001058  0.79001058  0.79001058 -2.20998942
```

**Cvičenie:** Chceme získať číselnú hodnotu Ljung-Boxovej štatistiky a príslušnú p-hodnotu z testovania prvých 12 autokorelácií. (V texte boli číselné hodnoty, nielen obrázok.)

- Pri použití funkcie `Box.test` potrebujeme dodať informáciu o tom, že ide o rezíduá z AR(1) modelu.
- Spraví sa to parametrov `fitdf`, ktorého hodnota sa rovná počtu odhadovaných ARMA parametrov (teraz počet AR parametrov)
- Vo výstupe si všimnite počet stupňov voľnosti `df`

```
Box.test(..., fitdf = ...)
```

```
##
##  Box-Ljung test
##
## data:  ar1$fit$residuals
## X-squared = 12.762, df = 11, p-value = 0.3092
```

# Cvičenie 5: Balík priceR, modelovanie inflácie

- Z balíka `priceR` použijeme `retrieve_inflation_data` na získanie dát z *World Bank API*
- Vytvorte z týchto dát časový rad s hodnotami inflácie.

```
library(priceR) # ak treba, nainstalujte
data <- retrieve_inflation_data("US")
```

```
## Validating iso2Code for US
## Generating URL to request all 297 results
## Retrieving inflation data for US
## Generating URL to request all 63 results
```

```r
str(data)
```

```
## 'data.frame':    63 obs. of  8 variables:
##  $ indicator      :'data.frame': 63 obs. of  2 variables
##   ..$ id   : chr  "FP.CPI.TOTL.ZG" "FP.CPI.TOTL.ZG" "FP.
##   ..$ value: chr  "Inflation, consumer prices (annual %)
##  $ country        :'data.frame': 63 obs. of  2 variables
##   ..$ id   : chr  "US" "US" "US" "US" ...
##   ..$ value: chr  "United States" "United States" "Unite
##  $ countryiso3code: chr  "USA" "USA" "USA" "USA" ...
##  $ date           : chr  "2022" "2021" "2020" "2019" ...
##  $ value          : num  8 4.7 1.23 1.81 2.44 ...
##  $ unit           : chr  "" "" "" "" ...
##  $ obs_status     : chr  "" "" "" "" ...
##  $ decimal        : int  1 1 1 1 1 1 1 1 1 1 ...
```

```
plot(x)
```

- Neskôr uvidíme, že takéto dáta treba diferencovať, aj keď v nich nie je trend.
- Preto teraz **budeme pracovať s diferenciami**
- Pri modelovaní si treba uvedomiť, že **konštantný člen v modeli pre diferencie znamená lineárny trend v pôvodných dátach** a tento trend sa dostane aj do predikcií

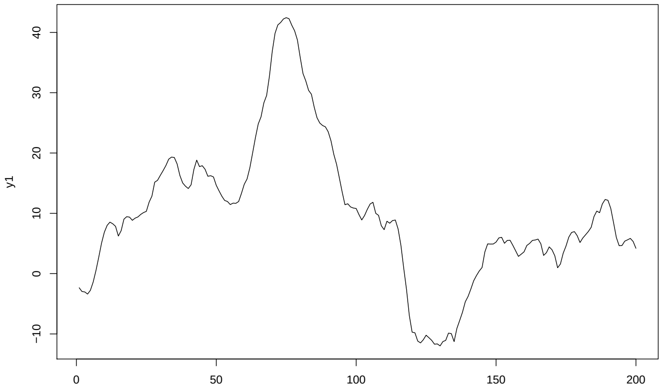
**Príklad 1.** Diferencie sú "ar(0)" s danou hodnotou $x_0 = u_0$:

$$x_t - x_{t-1} = c + u_t \Rightarrow x_t = ct + \sum_{i=0}^{t} u_i$$

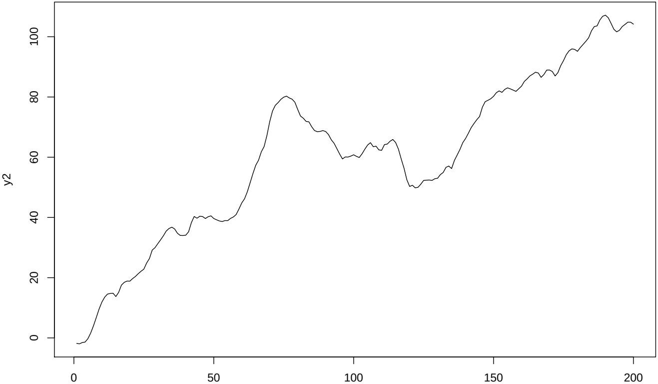
$$x_t - x_{t-1} = u_t \Rightarrow x_t = \sum_{i=0}^{t} u_i$$

**Príklad 2.** Simulácie priebehu, ak sú diferencie AR(1)

```r
set.seed(123)
x1 <- arima.sim(model = list(ar = c(0.8)), n = 200)
y1 <- ts(cumsum(x1)) # diferencie su ar(1) bez konst.
plot(y1)
```



```
set.seed(123)
x2 <- 0.5 + arima.sim(model = list(ar = c(0.8)), n = 200)
y2 <- ts(cumsum(x2)) # diferencie su ar(1) s konst.
plot(y2)
```



- V dátach o inflácii nie je trend $\Rightarrow$ chceme odhadnúť model pre diferencie bez konštanty
- Vo funkcii `sarima` pridáme parameter `no.constant` na `TRUE`
- Ak sa diferencuje viac ako raz, spraví sa to automaticky

```
# ar(p) model bez konstanty pre 1. diferencie
sarima(data, p, 1, 0, no.constant = TRUE)
```

- Odhadnite model bez konštanty a pozrite si vektor koeficientov.

- Pre diferencie zistite či sú pre ne na základe rezíduí dobrými modelmi AR(p) pre $p \in \{0, 1, 2, 3, 4, 5\}$ bez konštanty
- Z vyhovujúcich modelov vyberte ten, ktorý má najnižšie BIC.

Zopakujte pre infláciu v štátoch (všade treba diferencovať):

- Francúzsko (FR)
- Japonsko (JP)
- Kanada (CA)

## Cvičenie 6: Teoretický príklad

Nech $x$ je stacionárny AR(1) proces. Definujme

$$y_t = x_t - x_{t-1}.$$

Vypočítajte strednú hodnotu, disperziu a autokorelačnú funkciu procesu $y$ a ukážte, že je stacionárny.