AR, MA a ARMA procesy

Beáta Stehlíková
FMFI UK Bratislava

Overovanie stacionarity a invertovateľnosti

Opakovanie - stacionarita AR procesu

Zistite, či je proces

$$x_t = 1.2x_{t-1} + 0.5x_{t-2} + 0.3x_{t-3} + u_t$$

stacionárny.

- Napíšte polynóm, ktorho korene idete počítať.
- Aké musia byť, aby bol proces stacionárny?
- Vypočítajte korene a rozhodnite o stacionarite.

Invertovateľnosť MA procesu

Zistite, či je proces

$$x_t = 0.2 + u_t - 0.2u_{t-1} + 0.1u_{t-2}$$

invertovateľný.

- Napíšte polynóm, ktorho korene idete počítať.
- Aké musia byť, aby bol proces invertovateľný?
- Vypočítajte korene a rozhodnite o invertovateľnosti. **Pozor na znamienka pri použití funkcie `armaRoots`**

Stacionarita a invertovateľnosť ARMA procesu

Zistite, či sú nasledovné procesy stacionárne a invertovateľné:

- $x_t = 5 + 0.2x_{t-1} - 0.3x_{t-2} + u_t - 0.15u_{t-1} - 0.1u_{t-2}$
- $x_t = 1 + 0.7x_{t-1} - 0.6x_{t-2} - 0.1x_{t-3} + u_t + 0.3u_{t-1}$

Ak je proces stacionárny, vypočítajte jeho strednú hodnotu.

ACF a PACF

Autokorelačná funkcia

Opakovanie: Už sme robili: ACF napríklad pre AR proces

$$x_t = 1 + 0.5x_{t-1} + u_t :$$

```
acf1 <- ARMAacf(ar=c(0.5), lag.max=10)
barplot(acf1)
```

Analogicky: ACF pre MA proces

$$x_t = 1 + u_t + 0.5u_{t-1} :$$

```
acf2 <- ARMAacf(ma=c(0.5), lag.max=10)
barplot(acf2)
```

Aké sú presné hodnoty (teda nie numericky z R) týchto ACF?

Parciálna autokorelačná funkcia

Do funkcie `ARMAacf` doplníme parameter `pacf=TRUE`.

Napríklad: PACF pre AR proces

$$x_t = 1 + 0.5x_{t-1} + u_t :$$

```
pacf1 <- ARMAacf(ar=c(0.5), pacf=TRUE, lag.max=10)
barplot(pacf1)
```

PACF pre moving average proces

$$x_t = 1 + u_t + 0.5u_{t-1} :$$

```
pacf2 <- ARMAacf(ma=c(0.5), pacf=TRUE, lag.max=10)
barplot(pacf2)
```

ACF a PACF

ACF a PACF pre ARMA proces

$$x_t = 1 + 0.3x_{t-1}u_t + 0.5u_{t-1}$$

```
acf3 <- ARMAacf(ar=c(0.3), ma=c(0.5), lag.max=10)
barplot(acf3)
```

```
pacf3 <- ARMAacf(ar=c(0.3), ma=c(0.5), pacf=TRUE,
                 lag.max=10)
barplot(pacf3)
```

ACF a PACF

Zobrazte ACF a PACF pre nasledujúce stacionárne procesy:

- $x_t = 1 + u_t + 0.3u_{t-1}$
- $x_t = 3 + 0.2x_{t-1} + u_t - 0.3u_{t-1}$
- $x_t = 1 + 0.2x_{t-1} + 0.1x_{t-2} + u_t$

Identifikácia a odhadovanie ARMA modelov

Postup

- Zobrazíme výberovú ACF a PACF, naraz sa to dá spraviť funkciou `acf2(data)` z balíka `astsa`
- "Podozrenie" na AR proces máme vtedy, ak sa PACF vynuluje, počet signifikantných hodnôt je rád procesu.
- "Podozrenie" na MA proces máme vtedy, ak sa ACF vynuluje, počet signifikantných hodnôt je rád procesu.
- Môžeme odhadovať aj zmiešané ARMA procesy s AR aj MA členmi.
- `ARMA(p,q)` model pre `k`-te diferencie sa odhadne funkciou `sarima(data,p,k,q)`, analogicky predikcie `sarima.for`
- Treba overiť nekorelovanosť rezíduí (ACF a Ljung-Boxovu štatistiku)
- Ak existujú blízke korene z AR a MA časti, treba skúsiť `ARMA(p-1,q-1)` model

Príklad 1

Počet oviec v tisícoch v Anglicku a Walese, ročné dáta, 1867-1939
Na stránke v súbore `ovce.txt`

Klesajúci trend, preto budeme pracovať s diferenciami.

Príklad 2

Ceny kakaa z prednásky Na stránke v súbore `kakao.txt`

Budeme modelovať s diferencie logaritmov.

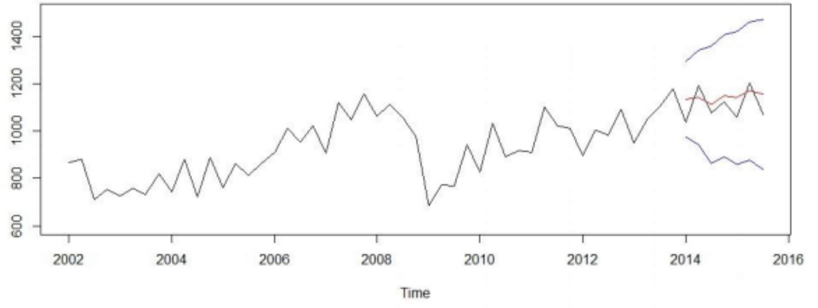
Ukážte, že sa tieto dáta dajú dobre modelovať aj pomocou AR aj pomocou MA procesu.

Porovnávanie reálnych dát s predikciami

- Zvoľte si dáta z niekorého z predchádzajúcich príkladov.
- Vynechajte niekoľko posledných pozorovaní a na zostavenie modelu pouzite len tie zostávajúce.
- Spravte predikcie a porovnajte ich so skutocným vývojom na základe dát, ktoré ste si na začiatku odložili bokom.

Porovnávanie reálnych dát s predikciami

Príklad - z domácej úlohy odovzdanej v roku 2015:



Porovnanie Woldovych reprezentácií

Modely pre `diff(gas)`

Ukážte, že AR(4) aj MA(3) sú vyhovujúce modely pre dáta `diff(gas)`.

Porovnajme ich Woldove reprezentácie.

Postup:

```
wold <- expand.grid(k=as.factor(1:10),
                    model=c("ar4","ma3"))
```

Pozrite si *data frame* `wold`, aby ste videli, akú má štruktúru.

Potom pridajte hodnoty premennej `psi` s koeficientami `psi.k` pre príslušný model.

```
wold$psi <- ...
```

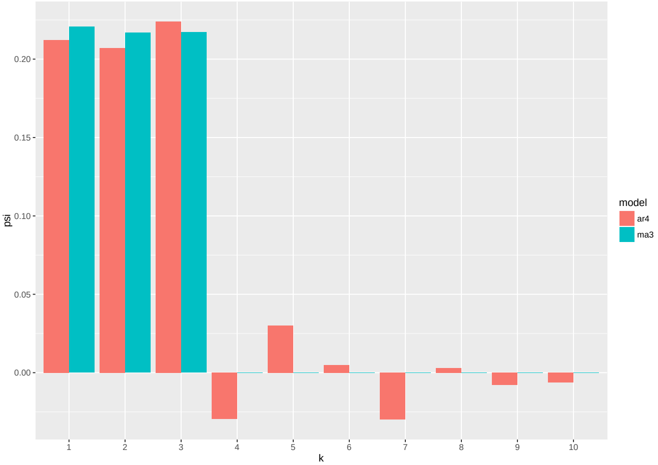
Modely pre diff(gas)

Porovnanie pomocou ggplot2:

```
library(ggplot2)

ggplot(wold, aes(x=k, y=psi, fill=model)) +
  geom_bar(stat="identity", position="dodge")
```

Modely pre `diff(gas)`



Cvičenia

CV.1 - AR(1) proces pozorovaný s chybou

Vygenerujeme AR(1) proces y:

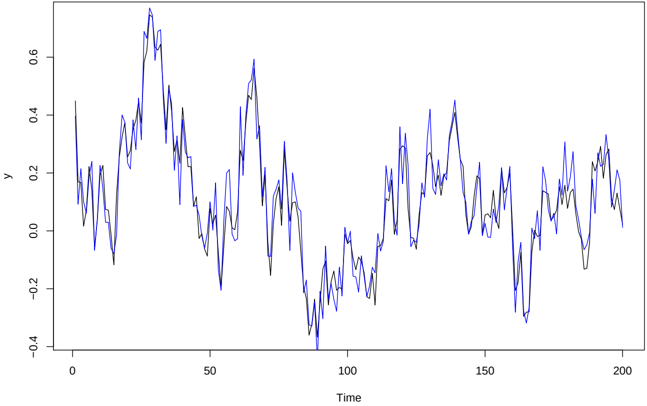
```
set.seed(12345)
y <- arima.sim(n=200, list(ar = c(0.9)), sd=0.1)
```

Pororovať ho ale budeme s chybou eps, vidí?me teda proces x:

```
eps <- rnorm(200)/15
x <- y+eps
```

Zobrazte vygenerované procesy y (pôvodný - čiernou) a x (pozorovaný - modrou).

CV.1 - AR(1) proces pozorovaný s chybou



CV.1 - AR(1) proces pozorovaný s chybou

Ukážte, že:

- AR(1) je dobrý model pre dáta `y` - to je prirodzené, tak sme ich generovali
- Ale AR(1) nie je dobrý model pre dáta `x`.
- ARMA(1,1) je dobrý model pre dáta `x`.

Poslednú vlastnosť dokážte analyticky (nielen pre vygenerované dáta) - teda ak `y` je AR(1) a `x=y+eps`, kde `eps` je proces definovaný v predch. slajdoch, tak `x` sa dá zapísať ako ARMA(1,1) proces.

CV.2 - Stacionarita a parameter procesu

Zadanie:

Určte všetky hodnoty parametra `k`, pre ktoré je stacionárny AR(2) proces

$$x_t = x_{t-1} + kx_{t-2} + u_t.$$

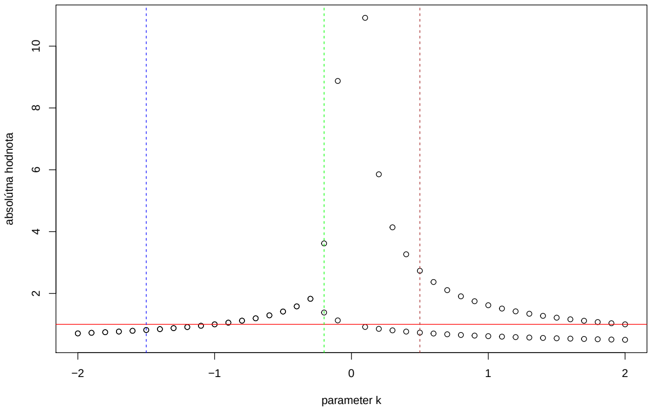
Skúsme najskôr numericky:

Pre nejaký rozsah parametra `k` nájdeme absolútnu hodnotu koreňov (pomocou `armaRoots`).

Na nasledujúcom slajde:

- Prečo pre niektoré `k` je len jedna absolútna hodnota (napríklad `k=-1.5` - modrá), kým pre iné sú dve (`k=-0.2` - zelená, `k=0.5` - hnedá)?
- Ktoré `k` vyhovovujú podmienke stacionarity a ktoré nie?

CV.2 - Stacionarita a parameter procesu



CV.2 - Stacionarita a parameter procesu

*Analytický výpočet

- Odvoďte korene a ich absolútne hodnoty.
- Zistite, kedy je splnená podmienka stacionarity.
- Skontrolujte si, či sa vaše výpočty zhodujú s numerickými na predchádzajúcom slajde

Na precvičenie programovania v R:

Spravte samostatne graf z predchádzajúceho slajdu.

CV.3 - Stacionarita a parameter procesu II.

Určte všetky hodnoty parametra `k`, pre ktoré je stacionárny proces

$$x_t = x_{t-1} + kx_{t-2} + kx_{t-3} + u_t.$$

Návod: Vyskúšajte najskôr niekoľko konkétnych hodnôt parametra `k`.