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

Zopakujte pre proces

$$x_t = 0.5 + u_t + 0.3u_{t-1} + 0.4u_{t-2} - 0.1u_{t-3}$$

.

Stacionarita a invertovateľnosť ARMA procesu

Rozhodnite, či sú nasledujúce procesy stacionárne a invertovateľné.

- $x_t = 5 + 0.2x_{t-1} - 0.3x_{t-2} + u_t - 0.15u_{t-1} - 0.1u_{t-2}$
- $x_t = 1 + 0.6x_{t-1} + 0.9x_{t-2} + u_t + 0.3u_{t-1}$
- $x_t = 1.5 + 0.6x_{t-1} + u_t - 0.3u_{t-1} - 0.2u_{t-2}$

Vypočítajte strednú hodnotu stacionárnych procesov.

ACF a PACF

Autokorelačná funkcia

Opakovanie: Už vieme: ACF napríklad pre autoregresný proces

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

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

Analogicky: ACF pre moving average 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 týchto autokorelačných funkcií?

Parciálna autokorelačná funkcia

Pri výpočte PACF doplníme do funkcie `ARMAacf` parameter `pacf=TRUE`.

PACF pre autoregresný 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.

O ktorých hodnotách vopred vieme, že budú nulové?

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

Identifikácia a odhadovanie ARMA modelov

Postup

- Zobrazíme ACF a PACF dát, jedným príkazom 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 stacionaritu a invertovateľnosť a rezíduá (ACF rezíduí a Ljung-Boxovu štatistiku)
- Ak existujú blízke korene z AR a MA časti, treba skúsiť `ARMA(p-1,q-1)` model

Cvičenie 1

Z minulého týždna:

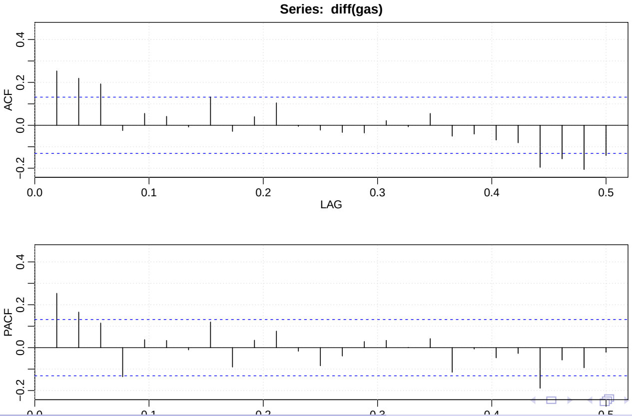
```
data(gas)
gas <- window(gas, start=c(2006,1))
```

Modelujeme diferencie premennej gas.

- Minulý týždeň sme našli AR(4) model. Ako to súvisí s ACF a PACF?
- Dal by sa nájsť aj dobrý MA model? ARMA model?

Cvičenie 1

```
acf2(diff(gas))
```



Cvičenie 2

Počet oviec v tisícoch v Anglicku a Walese, ročné dáta, 1867-1939

Na stránke cvičení v súbore `ovce.txt`

V dátach je klesajúci trend, preto budeme pracovať s diferenciami.

Cvičenie 3

Ceny kakaa z prednášky

Na stránke cvičení 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 Woldovych reprezentácií

Modely pre `diff(gas)`

Máme AR(4) a MA(3) model pre dáta `diff(gas)`. Porovnajme ich Woldove reprezentácie.

Postup:

```
k <- as.factor(rep(1:10, 2))
psi <- NA  # zatial
model <- rep(c("ar4","ma3"), each=10)
wold <- data.frame(k, psi, model)
```

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

Potom doplnte hodnoty premennej `psi`, ktorá bude v každom riadku obsahovať koeficient `psi.k` pre príslušný model.

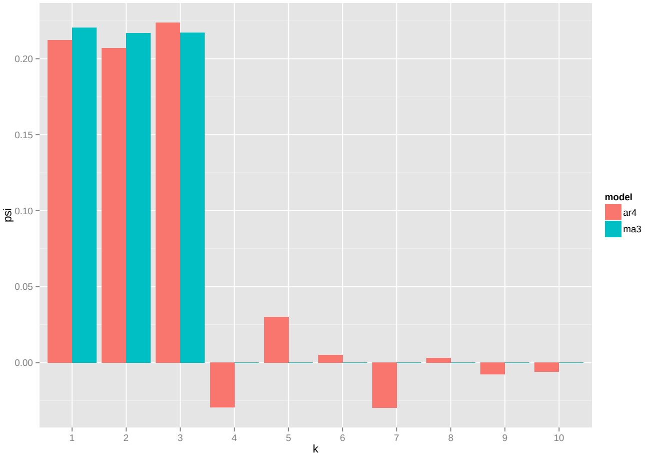
Modely pre diff(gas)

Porovnanie spravíme pomocou knižnice ggplot2:

```
library(ggplot2)

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

Modely pre `diff(gas)`



Modelovanie ceny kakaa

Takisto pre diferencie logaritmov cien kakaa máme k dispozícii aj AR aj MA proces.

Spravte podobné porovnanie Woldovych reprezentácií.

Cvičenia

AR(1) proces pozorovaný s chybou

Vygenerujeme AR(1) proces y:

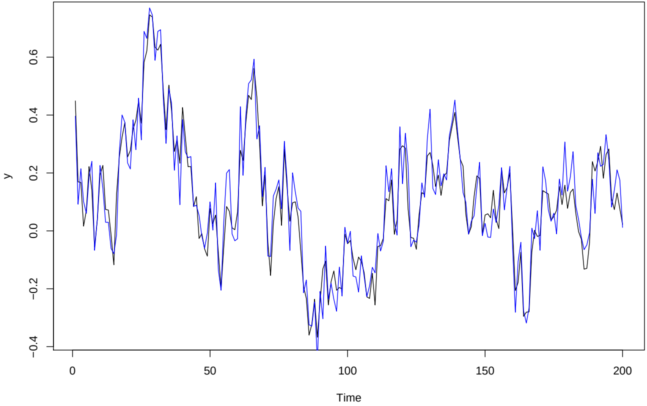
```
set.seed(12345)
y <- arima.sim(n=200, list(ar = c(0.9)), sd=0.1)
```

Pororovať ho ale budeme s chybou eps, vidíme teda proces x:

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

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

AR(1) proces pozorovaný s chybou



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 dokážeme v pondelok analyticky (nielen pre vygenerované dáta).

Stacionarita AR(2) procesu

*vC*o je na obálke knihy Woltera Endersa *Applied Econometric Time Series*? `http://www.amazon.com/Applied-Econometric-Time-Series-Edition/dp/0471230650`

"Experimentálne zistíme" podmienku stacionarity AR(2) procesu a jej súvislosť s obrázkom.

Stacionarita AR(2) procesu

Náhodne vygenerujeme parametre alpha1 a alpha2:

```
n <- 2000  # najskor menej, po odladeni napr. 2000
set.seed(123)

# runif ~ rovnomerne na (min, max)
alpha1 <- runif(n, min = -2.5, max =2.5)
alpha2 <- runif(n, min = -2.5, max = 2.5)
```

a budeme zisťovať, či je pre jednotlivé kombinácie parametrov príslušný proces stacionárny alebo nie.

Stacionarita AR(2) procesu

Vytvoríme *data frame*:

```
stationary <- rep(NA,n)
df <- data.frame(alpha1,alpha2,stationary)
head(df)
```

```
##      alpha1      alpha2 stationary
## 1 -1.0621124 -1.70163010         NA
## 2  1.4415257 -1.77742074         NA
## 3 -0.4551154 -1.75409804         NA
## 4  1.9150870  0.07217130         NA
## 5  2.2023364 -0.03586347         NA
## 6 -2.2722175  0.58171384         NA
```

Stacionarita AR(2) procesu

Doplňte pomocou funkcie `armaRoots`:

```
for (i in 1:n) {
 # doplnte cyklus
 # TRUE, FALSE podla toho, ci je proces stacionarny
 df$stationary[i] <-
}
```

Stacionarita AR(2) procesu

Potom môžeme kresliť graf pomocou funkcie `qplot` z knižnice `ggplot2`:

```
qplot(alpha1, alpha2, colour=stationary, data=df)
```

Napríklad pri `n`=2000 dostaneme výstižný obrázok.