

VREMENSKI NIZOVI 2: ARMA PROCESI

STATISTIČKI PRAKTIKUM 2

Proizvodnja piva u Australiji

U dokumentu `beer.csv` nalaze se podaci o proizvodnji piva u Australiji od siječnja 1956. do kolovoza 1995. Učitajte i grafički prikazite podatke. Iz grafa i prirode podataka pretpostavite koji bi mogao biti period sezonalne komponente te spremite podatke u varijablu `pivo` tipa `ts`.

Desezonalizacija niza

Očito je da će proizvodnja piva ovisiti o mjesecima, jer će u nekim mjesecima potrošnja biti veća, a u drugim manja. Procijenite sezonalnu komponentu, spremite ju u varijablu `pivo.s`, a desezonalizirane podatke u varijablu `pivo.ds`.

Desezonaliziran niz

Grafički prikažite desezonalizirani niz. Razmotrite moguće modele za procjenu trenda. Polinom kojeg stupnja možete sugerirati kao pogodan model?

Utvrdjivanje trenda

Procijenite parametre modela i prikažite graf modela uz desezonalizirane podatke.

Što s ostacima?

Nacrtajte graf ostataka i provjerite koreliranost grešaka.

Greške modeliramo kao ARMA proces

Uočavamo da su greške korelirane i ne možemo ih okarakterizirati kao bijeli šum. Zato greške modeliramo kao specijalnu klasu stacionarnih procesa - ARMA(p, q) *autoregressive moving-average* procesom.

Slučajni proces $\{X_t : t \in \mathbb{Z}\}$ zove se ARMA(p, q) proces ako je (X_t) stacionaran i za svaki t vrijedi

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q},$$

gdje je $(Z_t) \sim WN(0, \sigma^2)$.

Kažemo da je (X_t) ARMA(p, q) proces sa srednjom vrijednosti μ ako je $(X_t - \mu)$ ARMA(p, q) proces.

Ako je $\phi_1 = \phi_2 = \dots = \phi_p = 0$ proces zovemo MA(q), a ako je $\theta_1 = \theta_2 = \dots = \theta_q = 0$ AR(p).

Kako identificiramo pojedine procese?

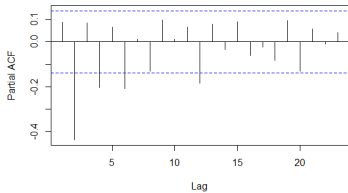
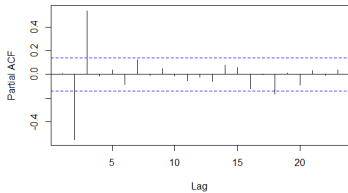
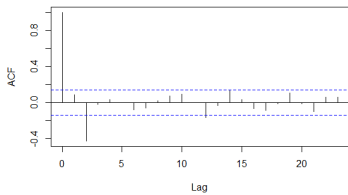
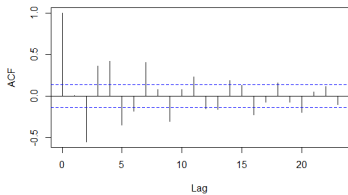
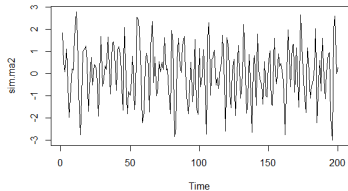
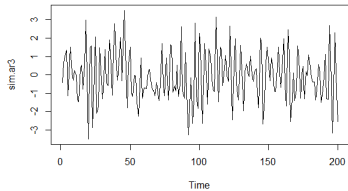
Odgovarajući stupanj (*lag*) AR ili MA modela možemo procijeniti iz grafa acf i pacf.

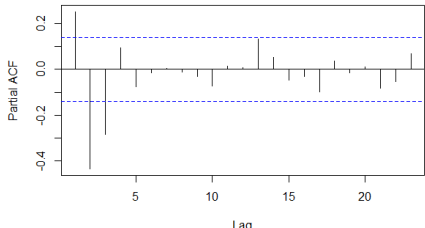
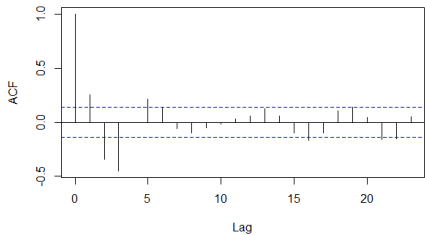
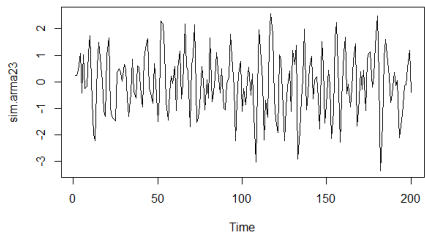
	acf	pacf
AR(p)	skok u lagu p, nakon toga geometrijski opada	nagli pad nakon laga p
MA(q)	nagli pad nakon laga q	skok u lagu q, nakon toga geometrijski opada

Simulirajmo AR(3), MA(2) i ARMA(2,3) procese.

```
> arima.sim
function (model, n, rand.gen = rnorm, innov = ...)

> sim.ar3 = arima.sim(list(ar=c(0.4,-0.5,0.6)), n=200)
> sim.ma2 = arima.sim(list(ma=c(0.4,-0.6)), n=200)
> sim.arma23 = arima.sim(list(ar=c(0.4,-0.6),
      ma=c(-0.1,0.3,-0.2)), n=200)
```





Procjena parametara

Nakon što identificiramo model (AR/MA/ARMA) možemo procijeniti parametre. Kako vrlo često ne znamo broj i tip parametara, najbolji model odabiremo pomoću AIC kriterija.

Ako naslućujemo (npr. iz grafa PACF) da se radi o AR procesu možemo koristiti proceduru ar koja će nam pomoću AIC kriterija i Yule-Walker procedure dati broj parametara (sama procjeni stupanj) i njihovu procjenu.

Povratne vrijednosti:

- ▶ ar - procijenjeni koeficijenti
- ▶ var.pred - $\hat{\sigma}^2$

Korištenjem procedure `ar` procijenite stupanj i parametre AR modela za simulirane AR(3) podatke.

Općenito, parametre ARMA(p, q) modela uz poznate stupnjeve p i q (dakle, sada ih sami moramo zadati) možemo procijeniti pomoću procedure

```
> arima(x, order = c(0, 0, 0), include.mean = TRUE, ...)
```

Gdje parametar `include.mean` označava želimo li da procedura procijeni srednju vrijednost (slobodan član u modelu) ili smo sami uklonili trend.

Procijenite parametre simuliranih MA(2) i ARMA(2,3) modela.

"Goodness-of-fit"

- ▶ provjera reziduala - trebaju dolaziti iz klase bijelog šuma (odnosno, ARMA model je opisao korelaciju), smanjila im se varijanca (tsdiag)
- ▶ provjera koeficijenata modela - značajno su različiti od 0
- ▶ provjera stacionarnosti - nepostojanje jediničnog korijena (nestacionarnost može uzrokovati nestabilne i zbrkane procjene u budućnosti)

Diferenciranje

Kažemo da je proces $ARIMA(p, d, q)$ proces, ako ga trebamo diferencirati d puta da bismo dobili $ARMA(p, q)$ proces (postoji d jediničnih korijena).

Proces će biti stacionaran $\Leftrightarrow d = 0$ (dakle, diferenciranjem rješavamo nestacionarnost).

Više modela može biti dobro - odabir ovisi o pretpostavkama s kojima radimo (u praksi uglavnom preferiramo manji red diferenciranja).

Provjera stacionarnosti

Karakteristični polinom $AR(p)$ procesa:

$$(1 - \phi_1 z - \dots - \phi_p z^p).$$

Proces će biti stacionaran ako nultočke njegovog karakterističnog polinoma leže izvan jedinične kugle.

Nestacionarnost ćemo uočiti promatrajući sumu koeficijenata u AR (odnosno MA) dijelu - ako se dogodi da je suma približno 1, možemo smanjiti red AR (MA) dijela i povećati (smanjiti) red diferenciranja.

Zadatak

Napišite funkciju `arma_AIC(x,n)` koja AIC kriterijem odabire najbolji $\text{ARMA}(p, q)$ model za podatke x , pri čemu je $p + q \leq n$. Funkcija ispisuje dobivene stupnjeve p i q te pripadnu vrijednost AIC-a.

Procjena parametara ARMA modela

Funkcijom `arma_AIC` procijenit ćemo stupnjeve ARMA modela za simulirane ARMA(2,3) podatke te za rezidualne iz primjera `pivo`.

```
> arma_AIC(sim.arma23,10)
      3      1
> arma_AIC(lm_pivo$res,10)
      5      4
```

Dobili smo da je za modeliranje ostataka proizvodnje piva najbolji ARMA(5,4) model. Procijenite parametre modela i provjerite jesu li dobiveni reziduali iz klase bijelog šuma.

Predviđanje budućih vrijednosti

Predvidite idućih 12 vrijednosti prodaje pive (tj. za iduće godišnje razdoblje).

Procijenimo i pouzdane intervale (koristimo asimptotsku formulu za 95%-pouzdana interval za vrijednost $\text{ARMA}(p, q)$ procesa).

Zadatak

U datoteci `airline.txt` nalaze se podaci o mjesečnom broju putnika u međunarodnom zračnom prometu u razdoblju od siječnja 1949. do prosinca 1960. godine (*Box, Jenkins, Reinsel*).

- ▶ Podatke o mjesečnom broju putnika prikažite grafički. Što možete reći o homogenosti podataka na temelju ovog grafa?
- ▶ Logaritmirajte podatke i spremite ih u format `ts`.
- ▶ Analizirajte komponente vremenskog niza.
- ▶ Procijenite mjesečni broj putnika za sljedeća 4 mjeseca.