VREMENSKI NIZOVI 1

STATISTIČKI PRAKTIKUM 2

6

Vremenski nizovi

Slučajni proces je familija slučajnih varijabli $\{X_t, t \in T\}$ na nekom vjerojatnosnom prostoru, gdje je T je skup indeksa koji određuje vremensku komponentu podataka (npr. $T = [0, t]$, $T = [0, \infty)$ ili $T = \{t_i : i \in I\}$).

Pri određenom mjerenju mi opažamo realizaciju (dijela) vremenskog niza $\{X_t(\omega), t \in T_0\} = \{x_t, t \in T_0\}$, gdje je $T_0 \subset T$ konačan.

Kako bismo mogli analizirati taj vremenski niz i predvidjeti njegove buduće vrijednosti, potreban nam je dobar model.

Klasična dekompozicija

Realizacija može sugerirati sljedeću reprezentaciju procesa

$$X_t = m_t + s_t + Y_t, \quad t \in T,$$

gdje je

- $t \mapsto m_t$ je polako promjenjiva (neslučajna) funkcija (*trend*)
- $t \mapsto s_t$ je (neslučajna) funkcija s poznatim periodom d (*sezonalna komponenta*)

$$s_{t+d} = s_t, \quad \forall t \in T, \qquad \sum_{t=1}^{d} s_t = 0$$

- $\{Y_t, t \in T\}$ je slučajni stacionarni niz (*buka/šum*)

Stacionaran niz

Vremenski niz $\{Y_t \ : \ t \in \mathbb{Z}\}$ je stacionaran ako je:

1. $\mathbb{E}|Y_t|^2 < \infty$ za sve $t \in \mathbb{Z}$;
2. $\mathbb{E}Y_t = \mu$ za sve $t \in \mathbb{Z}$;
3. $$\begin{aligned}\gamma_Y(r,s) &:= \mathbb{E}[(Y_r - \mu)(Y_s - \mu)] \\ &= \mathbb{E}[(Y_{r+t} - \mu)(Y_{s+t} - \mu)] \\ &=: \gamma_Y(r+t, s+t)\end{aligned}$$
   za sve $r, s, t \in \mathbb{Z}$.

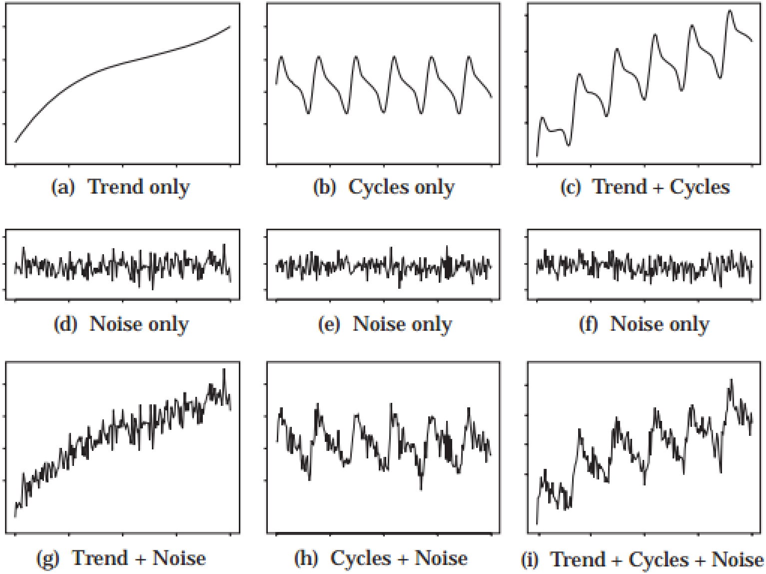
$\gamma_Y(t) := \gamma_Y(t+s, s)$ - autokovarijacijska funkcija
$\rho_Y(t) := \gamma_Y(t)/\gamma_Y(0)$ - autokorelacijska funkcija

Analiza vremenskog niza

Kod analize vremenskih nizova prvo krećemo s crtanjem grafa. Ako postoje očiti diskontinuiteti poput naglih promjena na nekoliko mjesta, vremenski niz treba razbiti na više samostalnih segmenata. Također treba provjeriti pozadinu podataka da ne bi ispalo da je podatak krivo zabilježen.

Cilj: procijeniti pojedine komponente klasične dekompozicije

1. sezonalnost s
2. trend m
3. šum Y



(a) Trend only
(b) Cycles only
(c) Trend + Cycles
(d) Noise only
(e) Noise only
(f) Noise only
(g) Trend + Noise
(h) Cycles + Noise
(i) Trend + Cycles + Noise

Primjer

Neka su $(\varepsilon_i)_{i=1}^{100} \overset{\text{njd}}{\sim} N(0,1)$, $f(t) = (\frac{t}{10})^2 - t$ i neka je

$$Z_0 = 0,\ Z_1 = \varepsilon_1;\ Z_t = \varepsilon_t - \frac{1}{2}Z_{t-1} + \frac{1}{4}Z_{t-2}, \quad t \geq 2.$$

Generirajte i grafički prikažite jednu realizaciju slučajnog niza X,

$$X_t = f(t) + Z_t, \quad t \in \{1, \dots, 100\}.$$

Primjer - nastavak

Sezonalna ponavljanja javljaju se prirodno i ovise o samoj pozadini problema.
Npr. ako nešto mjerimo po danima, može se dogoditi da svakih 5 (ili 7 ili 30) mjerenja uočavamo periodična odstupanja.
Ako promatramo populaciju po godišnjima dobima (period 4), u proljeće i ljeto će populacija biti stabilnija, a ranjivija u jesen i zimu.

Dodajte sljedeću sezonalnu komponentu našoj realizaciji:

$$d = 4,\ s_1 = 5,\ s_2 = 1,\ s_3 = -2,\ s_4 = -4$$

i grafički prikažite dobiveni niz.

Vremenski nizovi u R-u

Prvo moramo odrediti period d - on je poznat ili iz pozadine problema ili se može očitati s grafa.

Vremenski niz zapisujemo u strukturu podataka `ts`:

```
> xt=ts(xx,start=1,freq=4)
> xt
Qtr1          Qtr2         Qtr3         Qtr4
1    3.8151334  -1.7644974  -5.1097122  -8.2420158
2    0.8405033  -3.5992603  -8.1428177 -10.1876407
3   -3.3794320  -8.0372058 -12.0583187 -15.0705251
---
25   0.47481204 -0.05321804 -6.03541059 -3.14502481
```

I. Analiza vremenskog niza bez sezonalne komponente

$$X_t = m_t + Y_t$$

Ukoliko vremenski niz nema sezonalnu komponentu, ostaje nam procijeniti trend i šum. Česte metode za procjenu i/ili uklanjanje trenda su

- metoda linearnog filtera
- metoda najmanjih kvadrata
- diferenciranje niza.

Metoda linearnog filtera

Trend vremenskog niza X u trenutku t procjenjujemo s

$$\hat{m}_t = \sum_{i=-\infty}^{\infty} \lambda_i X_{t+i},$$

gdje težine λ_i reguliraju doprinos pojedine vrijednosti.
Jednostavna klasa linearnih filtera su pomični prosjeci s jednakim težinama: za $a \in \mathbb{N}$ uzmemo

$$\hat{m}_t = \frac{1}{2a+1} \sum_{k=-a}^{a} X_{t+k}.$$

U ovom slučaju procjena trenda u trenutku t je dana s prosjekom vrijednosti skupa $\{x_{t-a}, \dots, x_t, \dots, x_{t+a}\}$.

Primjer - nastavak

Procijenite trend vremenskog niza `x` metodom linearnog trenda za $a = 2$ i grafički ga prikažite.

Možete koristiti naredbu `mean`, ali u `R`-u postoji specijalizirana naredba `filter`.

Metoda najmanjih kvadrata

Za zadani skup podataka iz grafa odlučujemo o obliku funkcije trenda koju želimo prilagoditi podacima (najčešće je to polinom).

Procijenite trend podataka x metodom najmanjih kvadrata i grafički ga prikažite.

Diferenciranje

Za niz (X_t) diferenciranje j-tog reda definiramo rekurzivno

$$\nabla X_t = X_t - X_{t-1}, \quad \nabla^j X_t = \nabla(\nabla^{j-1} X_t).$$

Ako je trend polinom k-tog stupnja, niz trebamo diferencirati k puta da uklonimo trend.

$$m_t = \sum_{j=0}^{k} a_j t^j \Longrightarrow \nabla^k X_t = k! a_k + \nabla^k Y_t.$$

Diferencirajte niz x pomoću naredbe `diff` kako bi uklonili trend te grafički prikažite ostatak.

Diferenciranje primjenjujemo dovoljan broj puta dok ne dobijemo stacionaran niz. Pokazuje se da je u praksi dovoljno diferenciranje provesti nekoliko puta (1 ili 2).

II. Procjena sezonalne komponente

Postoji nekoliko načina na koje možemo procijeniti i/ili ukloniti sezonalnu komponentu vremenskog niza. Nakon toga nastavljamo s procedurama za niz bez sezonalne komponente.

Često za procjenu sezonalnosti koristimo

- metodu malog trenda
- metodu pomičnih zareza
- diferenciranje

Metoda malog trenda

Ako je trend malen (sporo promjenjiv), nije nerazumno pretpostaviti da je trend konstantan kroz neki period duljine d (npr. trend kroz godinu je isti).

Neka je $(x_n)_{n=1}^{100}$ niz podataka zapisan u vektoru, označimo k-ti kvartal j-te sezone sa

$$x_{j,k} := x_{k+4(j-1)}, \quad k = 1, 2, 3, 4, \ j = 1, \ldots, 25.$$

Trend u j-toj sezoni procjenjujemo s

$$\hat{m}_j = \frac{1}{4} \sum_{k=1}^{4} x_{j,k}.$$

Sezonalnu komponentu k-te sezone procjenjujemo s

$$\hat{s}_k = \frac{1}{25} \sum_{j=1}^{25} (x_{j,k} - \hat{m}_j), \quad k = 1, \ldots, 4.$$

Primjer - nastavak

Procijenite metodom malog trenda sezonalnu komponentu podataka xx, usporedite to sa sezonalnom komponentom koju smo dodali u podatke i nacrtajte podatke bez sezonalne komponente.

Metoda pomičnih zareza

Ova tehnika ne ovisi o pretpostavci da je u tijeku sezone trend m_t konstanta. Trend procjenjujemo linearnim filterom i to na sljedeći način

- Ako je period $d = 2q + 1$ neparan

$$\hat{m}_t = \frac{1}{d} \sum_{k=-q}^{q} x_{k+t}.$$

- Ako je $d = 2q$ paran sa

$$\hat{m}_t = \frac{0.5x_{t-q} + x_{t-q+1} + \ldots + x_{t+q-1} + 0.5x_{t+q}}{d}.$$

Dalje računamo prosjek w_k skupa $\{(x_{k+jd} - \hat{m}_{k+jd}) \ : \ q < k + jd \leq n - q\}$, $k = 1, \ldots, d$. Kako suma w_k-ova nije nužno 0, s_k procjenjujemo sa

$$\hat{s}_k = w_k - \frac{1}{d} \sum_{j=1}^{d} w_j.$$

Primjer - nastavak

Procijenite metodom pomičnih zareza (naredba `filter`) sezonalnu komponentu podataka xx, usporedite to sa sezonalnom komponentom koju smo dodali u podatke i nacrtajte podatke bez sezonalne komponente.

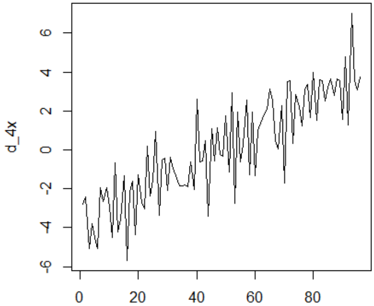
Diferenciranje

Na vremenskom nizu (X_t) sa sezonalnom komponentom perioda d možemo napraviti transformaciju

$$\nabla_d X_t = X_t - X_{t+d}.$$

Novi niz $(\nabla_d X_t)$ neće sadržavati sezonalnu komponentu. U R-u $(\nabla_d x_t)$ dobivamo naredbom `diff(x,lag=d)`.

```
> d_4x=diff(xx,lag=4)
> ts.plot(d 4x)
```



III. Procjene u R-u

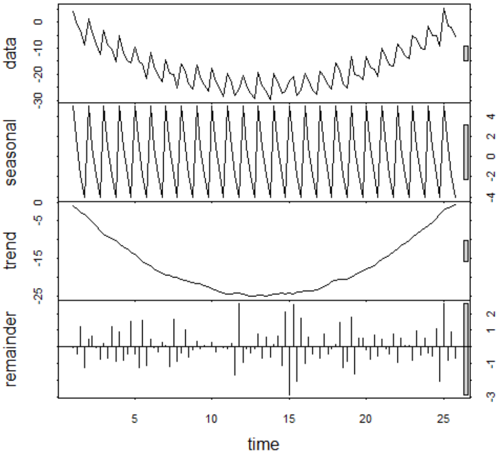
Postoje razne metode procjene komponenti vremenskog niza, jedna od njih je implementirana funkcijom `stl`. (Koristi se LOESS metoda - neparametarska lokalna polinomijalna regresija, za procjenu sezonalne komponente i trenda.)

```
> x.stl=stl(xt,"periodic")
> x.stl
Call:
stl(x = xt, s.window = "periodic")

Components
seasonal        trend    remainder
1 Q1  4.867846  -1.44873681  0.396023800
1 Q2  0.995836  -2.29294229 -0.467391078
1 Q3 -2.192161  -3.10257974  0.185028295
1 Q4 -3.671522  -3.82879551 -0.741698047
2 Q1  4.867846  -4.39355828  0.366215257
2 Q2  0.995836  -4.98682268  0.391726502
---
```

Nacrtajmo rastav koji je dobiven funkcijom `stl`.

```
> plot(x.stl)
```



Do procijenjenih vrijednosti sezonalne komponente dolazimo ovako.

```
> x.sto=x.stl$time.series
> x.s=x.sto[,1]
> x.s
Qtr1       Qtr2       Qtr3       Qtr4
1   4.867846  0.995836 -2.192161 -3.671522
2   4.867846  0.995836 -2.192161 -3.671522
3   4.867846  0.995836 -2.192161 -3.671522
----
```

IV. Analiza stacionarnog niza

Nakon što se riješimo sezonalnosti i procijenimo trend ostaje nam analiza stacionarnog niza koji je preostao.

Želimo procijeniti autokovarijacijsku funkciju γ i autokorelacijsku funkciju ρ. Uvodimo uzoračku autokovarijacijsku funkciju za niz $(x_k)_{k=1}^{n}$

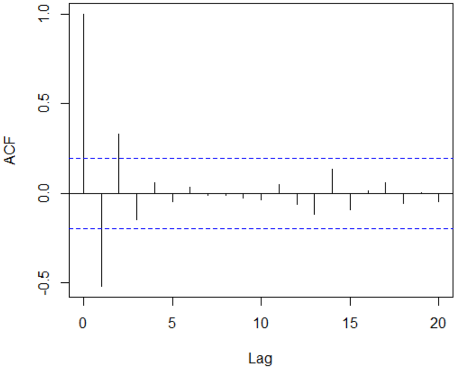
$$\hat{\gamma}(h) := \frac{1}{n}\sum_{j=1}^{n-h}(x_{j+h} - \bar{x})(x_j - \bar{x}),$$

za $0 \leq h < n$. Za negativne pomake definiramo $\hat{\gamma}(-h) := -\hat{\gamma}(h)$.

Uzoračku autokorelacijsku funkciju definiramo sa

$$\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0), \quad |h| < n.$$

Voljeli bi da je niz (Y_t) nekoreliran s očekivanjem 0 i konstantnom varijancom σ^2 - *bijeli šum* $WN(0, \sigma^2)$. Tada graf uzoračke autokorelacijske funkcije izgleda ovako.



Bilo bi dobro da bijeli šum bude normalno distribuiran.
Ako uzoračka autokorelacijska funkcija sugerira da se radi o bijelom šumu, onda provedemo testove normalnosti.

Zadatak

Neka je `xx` realizacija vremenskog niza (X_t).

(a) Procijenite sezonalnu komponentu procedurom *stl* i uklonite ju.

(b) Procijenite trend polinomom drugog stupnja i nacrtajte preostali šum.

(c) Nacrtajte residual-fit plot ostataka, autokorelacijsku funkciju i normalni vjerojatnosni graf.