

Spektrálna analýza

Beáta Stehlíková

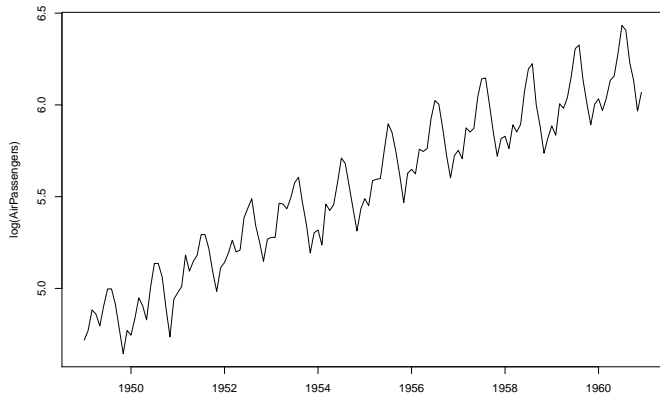
2-EFM-102 & 2-INF-191 Časové rady

Fakulta matematiky, fyziky a informatiky, UK v Bratislave

Motivácia

- ▶ Pracovali sme sa dátami o počtoch cestujúcich aerolinkami v jednotlivých mesiacoch

```
plot(log(AirPassengers))
```



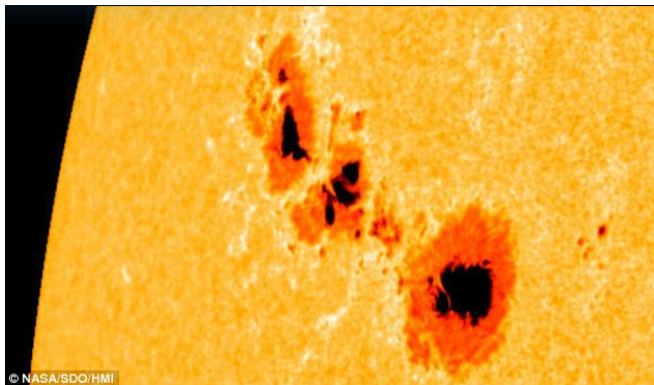
Periodické dáta na predchádzajúcich prednáškach:

- ▶ SARIMA modely
- ▶ Holt-Wintersov model

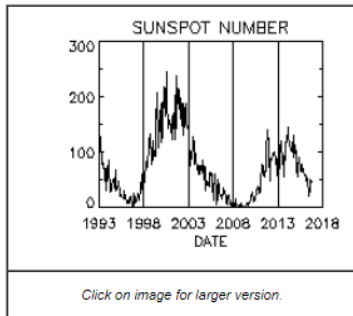
Na druhej strane:

- ▶ Nie vždy je v dátach takto *jednoznačná a pravidelná* perióda, aj keď majú periodický charakter

Slnečné škvrny



<http://www.dailymail.co.uk/sciencetech/article-2042428/Best-auroras-seen-Britainthanks-huge-solar-flares.html>



Sunspot Numbers

In 1610, shortly after viewing the sun with his new telescope, Galileo Galilei (or was it Thomas Harriot?) made the first European observations of Sunspots. Continuous daily observations were started at the Zurich Observatory in 1849 and earlier observations have been used to extend the records back to 1610. The sunspot number is calculated by first counting the number of sunspot groups and then the number of individual sunspots.

<https://solarscience.msfc.nasa.gov/SunspotCycle.shtml>

Mesačné dáta v R-ku z balíka datasets

Monthly Sunspot Data, from 1749 to "Present"

Description

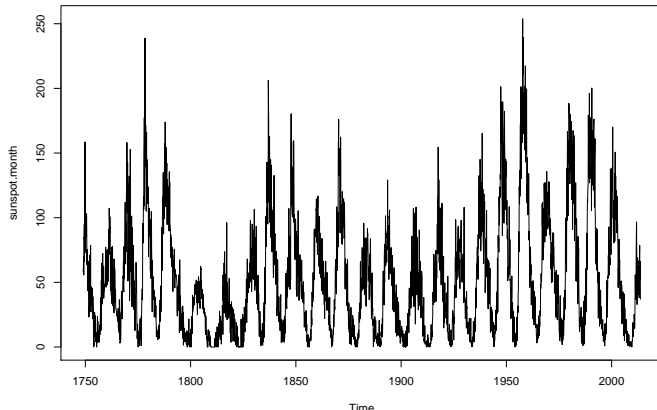
Monthly numbers of sunspots, as from the World Data Center, aka SIDC. This is the version of the data that will occasionally be updated when new counts become available.

```
str(sunspot.month)
```

```
## Time-Series [1:3177] from 1749 to 2014: 58 62.6 70 55.7
```

Otázka: Aká je perióda týchto dát (nie je úplne presná a pravidelná)?

```
plot(sunspot.month)
```



Spektrum

Spektrum

- ▶ Postupnosť $\{\gamma_j\}_{-\infty}^{\infty}$ → vytvárajúca funkcia

$$g(z) = \sum_{j=-\infty}^{\infty} \gamma_j z^j$$

- ▶ *Stacionárny proces* Y s autokovarianciami $\{\gamma_j\}_{-\infty}^{\infty}$ → spektrum

$$s_Y(\omega) = \frac{1}{2\pi} g(e^{-i\omega}) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j e^{-i\omega j},$$

kde i je imaginárna jednotka

Vlastnosti spektra

- ▶ Spektrum $s_Y(\omega)$:
 - ▶ využijeme, že $e^{-i\omega j} = \cos(\omega j) - i \sin(\omega j) \rightarrow$ dá sa napísať v tvare

$$s_Y(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j e^{-i\omega j} = \frac{1}{2\pi} \left(\gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j \cos(\omega j) \right)$$

- ▶ nadobúda teda reálne hodnoty
 - ▶ je to párna funkcia
 - ▶ je periodická s periódou 2π
- ▶ Stačí nám teda poznať hodnoty spektra na medzi 0 a π
- ▶ Dá sa dokázať, že $s_Y(\omega) \geq 0$ [Fuller, 1976]

Autokovariancie zo spektra

- ▶ Máme

$$s_Y(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j e^{-i\omega j}$$

→ pre násobíme $e^{i\omega k}$ a zintegrujeme od $-\pi$ po π

- ▶ Pre $j \neq k$ platí

$$\int_{-\pi}^{\pi} e^{-i\omega j} e^{i\omega k} d\omega = 0$$

- ▶ Dostaneme

$$\gamma_k = \int_{-\pi}^{\pi} s_Y(\omega) e^{i\omega k} d\omega$$

Disperzia zo spektra

- ▶ Pre $k = 0$ dostaneme $\gamma_0 = \int_{-\pi}^{\pi} s_Y(\omega) d\omega$ a keďže spektrum je párna funkcia,

$$\gamma_0 = 2 \int_0^{\pi} s_Y(\omega) d\omega$$

- ▶ Teda disperzia je dvojnásobkom plochy pod spektrom na intervale $(0, \pi)$.
- ▶ Z priebehu spektra teda vidíme, ktoré frekvencie najviac prispievajú k disperzii procesu - **tie, kde má spektrum vysokú funkčnú hodnotu**

Príklady výpočtu spektra

- ▶ **Príklad 1.** Biely šum u má konštantné spektrum

$$s_u(\omega) = \frac{\sigma^2}{2\pi}$$

- ▶ **Príklad 2.** MA(2) proces

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

Nenulové autokovariancie sú iba

$$\gamma_0 = (1 + 0.5^2 + 0.3^2)\sigma^2, \gamma_1 = (0.5 - 0.15)\sigma^2, \gamma_2 = -0.3\sigma^2,$$

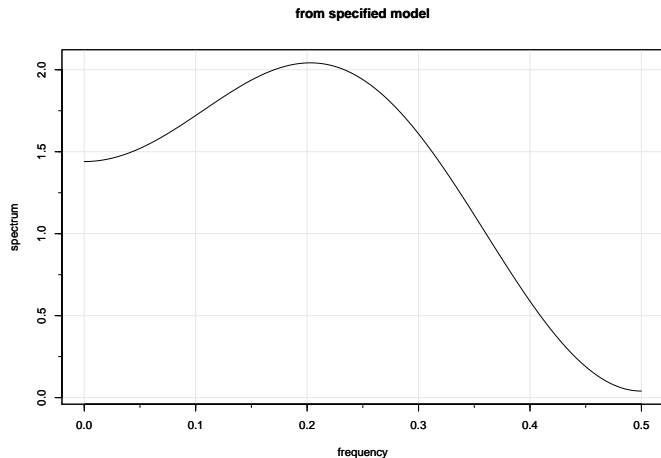
spektrum teda je

$$s_x(\omega) = \frac{\sigma^2}{2\pi} (1.34 + 0.7 \cos(\omega) - 0.6 \cos(2\omega))$$

- ▶ **Ďalšie príklady v cvičeniach** (na konci slajdov)

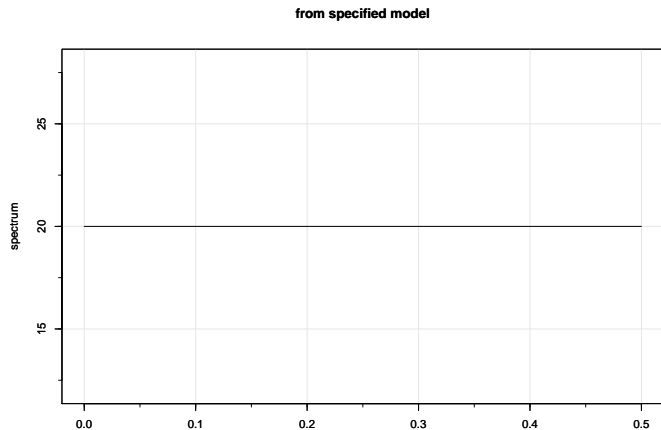
Funkcia arma.spec z balíka astsa

```
arma.spec(ma = c(0.5, -0.3))
```



- ▶ x-ová os je frekvencia $f = \frac{\omega}{2\pi}$, y-ová os je preškálované spektrum tak, aby plocha bola rovná polovici disperzie

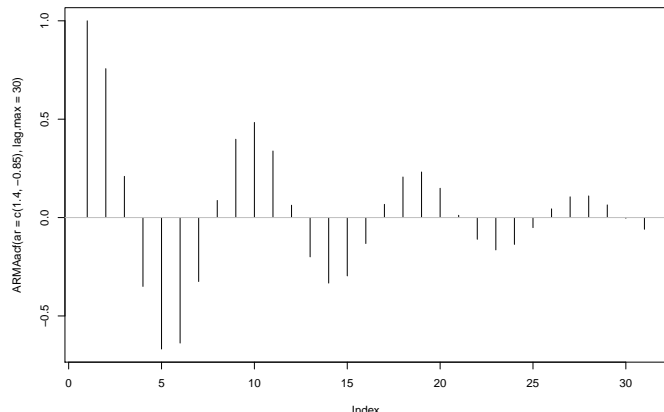
```
arma.spec(ar = c(0), var.noise = 20)
```



Príklad 3. Pripomeňme si AR(2) proces s periodickým charakterom

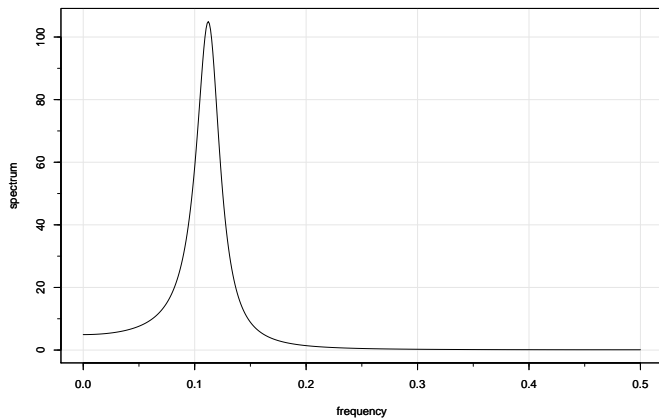
$$x_t = 1.4x_{t-1} - 0.85x_{t-2} + u_t$$

```
plot(ARMAacf(ar = c(1.4, -0.85), lag.max = 30), type="h")
```



```
spektrum_ar2 <- arma.spec(ar = c(1.4, -0.85))
```

from specified model



```
str(spektrum_ar2)
```

```
## List of 2  
## $ freq: num [1:500] 0 0.001 0.002 0.00301 0.00401 ...  
## $ spec: num [1:500, 1] 4.94 4.94 4.94 4.95 4.95 ...  
## - attr(*, "class")= chr "spec"
```

```
# pre akú frekvenciu sa nadobúda maximum spektra  
spektrum_ar2$freq[which.max(spektrum_ar2$spec)]
```

```
## [1] 0.1122244
```

```
# akej perióde to zodpovedá  
1/spektrum_ar2$freq[which.max(spektrum_ar2$spec)]
```

```
## [1] 8.910714
```

Odhadovanie spektra

Výberový periodogram a jeho zhladenie

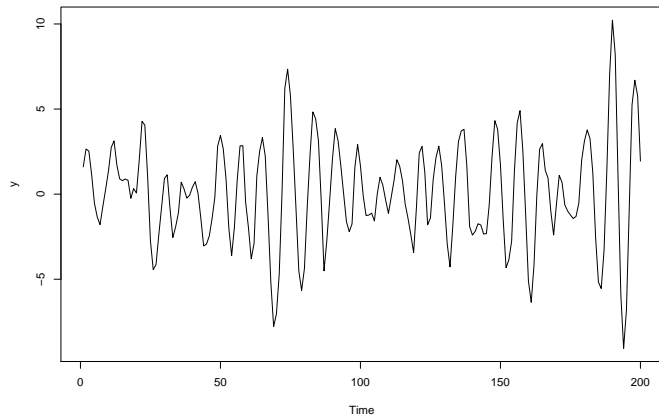
Výberový periodogram

- ▶ Máme dáta y_1, \dots, y_T a chceme z nich odhadnúť spektrum časového radu
- ▶ Prvá myšlienka:
 - ▶ nahradíme autokovariancie v definícii spektra výberovými autokovarianciami
 - ▶ odhadnúť vieme len konečne veľa autokovariancií, takže z nekonečnej sumy zostane konečná
- ▶ Takto dostaneme **výberový periodogram**

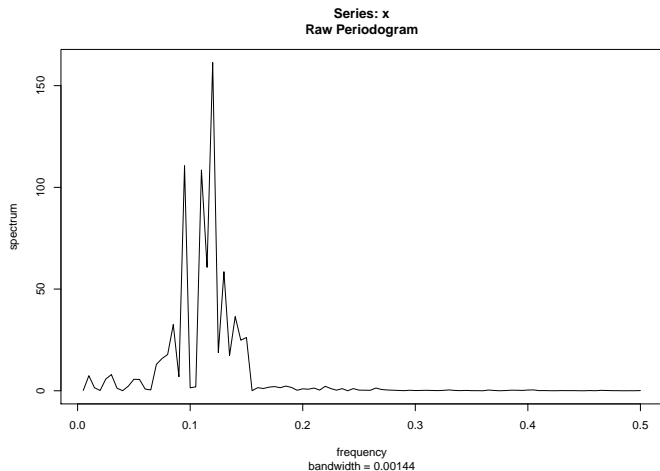
Výpočet v R-ku

Príklad 1. Vygenerujme dáta z procesu s periodickým charakterom $x_t = 1.4x_{t-1} - 0.85x_{t-2} + u_t$ (vieme, že perióda je cca 9)

```
set.seed(123)
y <- arima.sim(model = list(ar = c(1.4, -0.85)), n = 200)
plot(y)
```




```
spectrum(y, log = "no") # na x-ovej osi znovu frekvencia
```



Nevýhody

- ▶ Odhady majú veľkú disperziu
- ▶ Presnosť sa nezlepšuje so zvyšujúcim sa počtom pozorovaní (lebo odhadujeme stále viac autokovariancií)
- ▶ Preto potrebujeme iný odhad spektra

Podrobnejšie o nevýhodách:

- ▶ [Fuller, 1976]: pre veľký rozsah výberu má podiel $\frac{2\hat{s}_y(\omega)}{s_Y(\omega)}$ približne $\chi^2(2)$ rozdelenie a tieto podiely sú pre rôzne ω približne nezávislé
- ▶ $\mathbb{E}[\chi^2(2)] = 2$, teda

$$\mathbb{E}(\hat{s}_y(\omega)) \approx s_Y(\omega),$$

čo je dobrá vlastnosť

- ▶ Pravdepodobnosť, že $\chi^2(2)$ je z intervalu (0.05, 7.4) je 0.95 - preto

$$(0.025\hat{s}_y(\omega), 3.75\hat{s}_y(\omega))$$

je 95 percentný interval spoľahlivosti pre odhad spektra - tento interval je **príliš široký**

Vylepšenie

Idea: spektrum je spojité \rightarrow ak sú blízke hodnoty ω , resp. f , tak sú blízke aj hodnoty spektra \rightarrow zoberieme **vážený priemer okolitých hodnôt**

$$\hat{\sigma}_Y(\omega) = \sum_{m=-h}^h \kappa(\omega_{j+m}, \omega_j) \hat{\sigma}_Y(\omega_{j+m}),$$

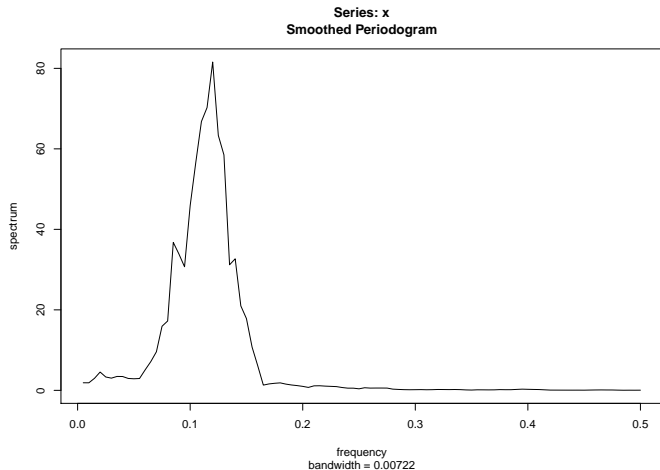
kde

- ▶ konštanta h udáva počet okolitých hodnôt, ktoré berieme do úvahy pri zhladzovaní (tzv. *bandwidth*)
- ▶ funkcia κ určuje pre tieto okolité hodnoty váhy (v súčte dávajú 1)

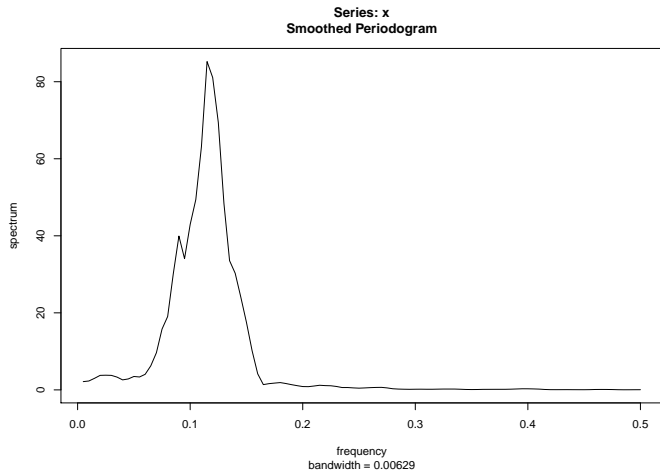
Výpočet v R-ku

Zhladenie periodogramu v R-ku

```
spectrum(y, kernel("daniell"), log = "no")
```



```
spectrum(y, kernel("modified.daniell"), log = "no")
```



Nájďme maximum a zodpovedajúcu periódu

```
sp_sim <- spectrum(y, kernel("modified.daniell"),  
                  log = "no", plot = FALSE)  
str(sp_sim)
```

```
## List of 16  
## $ freq      : num [1:100] 0.005 0.01 0.015 0.02 0.025 0  
## $ spec      : num [1:100] 2.14 2.31 3 3.77 3.81 ...  
## $ coh       : NULL  
## $ phase     : NULL  
## $ kernel    :List of 2  
## ..$ coef: num [1:3] 0.25 0.25 0.125  
## ..$ m      : int 2  
## ..- attr(*, "name")= chr "mDaniell(2)"  
## ..- attr(*, "class")= chr "tskernel"  
## $ df        : num 8.19  
## $ bandwidth : num 0.00622
```



```
# pre akú frekvenciu sa nadobúda maximum  
# odhadnutého spektra  
sp_sim$freq[which.max(sp_sim$spec)]
```

```
## [1] 0.115
```

```
# akej perióde to zodpovedá  
1/sp_sim$freq[which.max(sp_sim$spec)]
```

```
## [1] 8.695652
```

Výpočet v R-ku: dáta s časovou štruktúrou + trend

Trend v dátach

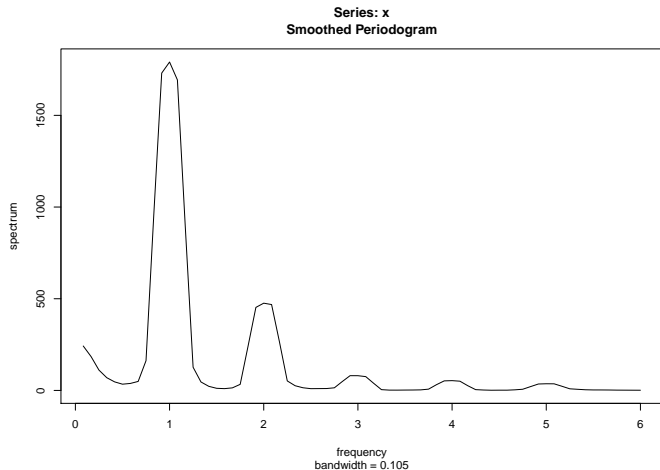
- ▶ Spektrum sa počíta pre stacionárny časový rad
- ▶ Preto ak je v dátach trend, pred ďalším výpočtom ho R-ko odstráni

Časová štruktúra:

- ▶ Doteraz boli pozorovania len “očíslované” (prvé, druhé, ...)
- ▶ Pri práci s časovými radmi im niekedy zadávame frekvenciu, napríklad pomocou `ts(data, frequency = 12, start = ...)`
- ▶ Ak majú dáta frekvenciu, na x-ová os odhadnutého spektra je preškálovaná na interval $(0, \text{frequency}(\text{data})/2)$

Príklad 2: dáta s trendom a časovou štruktúrou

```
spectrum(AirPassengers, kernel("modified.daniell"), log = 'n')
```



```
sp_air <- spectrum(AirPassengers,  
                  kernel("modified.daniell"),  
                  log = "no", plot = FALSE)
```

```
# pre akú frekvenciu sa nadobúda maximum odhadnutého spektra  
sp_air$freq[which.max(sp_air$spec)]
```

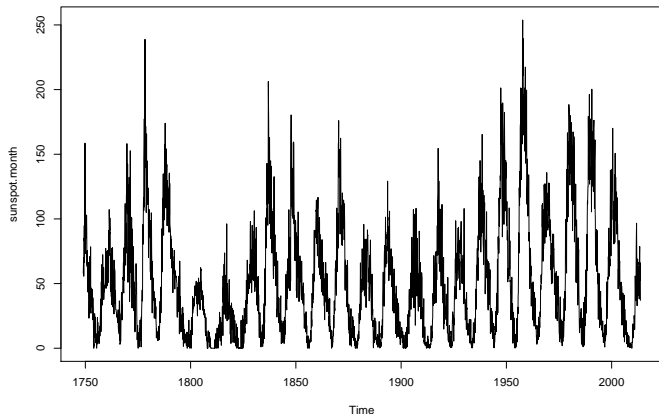
```
## [1] 1
```

```
# akej perióde to zodpovedá (v rokoch)  
1/sp_air$freq[which.max(sp_air$spec)]
```

```
## [1] 1
```

Príklad 3: splnečné škvrny z úvodu

```
plot(sunspot.month) # mesacne data, frekvencia = 12
```



```
str(spectrum(sunspot.month,  
            kernel("modified.daniell"),  
            log = "no", plot = FALSE))
```

```
## List of 16  
## $ freq      : num [1:1600] 0.00375 0.0075 0.01125 0.015  
## $ spec      : num [1:1600] 7975 8335 7546 6871 4948 ...  
## $ coh       : NULL  
## $ phase     : NULL  
## $ kernel    :List of 2  
## ..$ coef: num [1:3] 0.25 0.25 0.125  
## ..$ m      : int 2  
## ..- attr(*, "name")= chr "mDaniell(2)"  
## ..- attr(*, "class")= chr "tskernel"  
## $ df        : num 8.13  
## $ bandwidth: num 0.00472  
## $ n.used    : int 3200
```

```
sp_sun <- spectrum(sunspot.month,  
                   kernel("modified.daniell"),  
                   log = "no", plot = FALSE)
```

```
# pre akú frekvenciu sa nadobúda maximum odhadnutého spektra  
sp_sun$freq[which.max(sp_sun$spec)]
```

```
## [1] 0.09375
```

```
# akej perióde to zodpovedá  
# (v rokoch)  
1/sp_sun$freq[which.max(sp_sun$spec)]
```

```
## [1] 10.66667
```

Porovnajme s tým, čo je o slnečných škvrnách známe:

The "sunspot number" is then given by the sum of the number of individual sunspots and ten times the number of groups. Since most sunspot groups have, on average, about ten spots, this formula for counting sunspots gives reliable numbers even when the observing conditions are less than ideal and small spots are hard to see. Monthly averages (updated monthly) of the sunspot numbers ([181 kb JPEG image](#)), ([307 kb pdf-file](#)), ([62 kb text file](#)) show that the number of sunspots visible on the sun waxes and wanes **with an approximate 11-year cycle.**

Cvičenie

Pre slnečné škvrny sú v balíku `datasets` dostupné aj ročné dáta `sunspot.year`. Odhadnite spektrum tohto časového radu a určte pomocou neho periódu týchto dát. (Očakávame podobný výsledok ako pri mesačných dátach.)

Teoretické cvičenia na výpočet spektra

Ak nie je povedané inak, spektrum treba upraviť do tvaru, ktorý

- ▶ neobsahuje komplexné čísla
- ▶ neobsahuje nekonečné sumy

Príklad 1: AR(1) podľa definície

- ▶ O stacionárnom AR(1) procese (teda je $|\alpha| < 1$)

$$x_t = \delta + \alpha x_{t-1} + u_t$$

vieme, že

$$\mathbb{D}(x_t) = \frac{\sigma^2}{1 - \alpha^2}, \rho(k) = \alpha^k$$

- ▶ Autokovariancie sú

$$\gamma_k = \frac{\sigma^2}{1 - \alpha^2} \alpha^k$$

- ▶ Dosadením autokovariancií do definície vypočítajte spektrum tohto AR(1) procesu.

Užitočné tvrdenie o spektre transformovaného procesu

Tvrdenie:

- ▶ Nech x_t je stacionárny proces so spektrom $s_x(\omega)$, definujme

$$y_t = \sum_{s=-\infty}^{\infty} a_s x_{t-s}$$

- ▶ Potom jeho spektrum je

$$s_y(\omega) = A(e^{i\omega})A(e^{-i\omega})s_x(\omega),$$

kde

$$A(z) = \sum_{s=-\infty}^{\infty} a_s z^s$$

Poznámka: Všimnime si, že pri tejto definícii A sa proces y dá zapísať ako

$$y_t = A(L)x_t$$

Idea dôkazu:

- ▶ Stačí dokázať, že

$$\text{cov}(y_t, y_{t+k}) = \int_{-\pi}^{\pi} A(e^{i\omega})A(e^{-i\omega})s_x(\omega)e^{i\omega k}d\omega$$

- ▶ Touto vlastnosťou je ale charakterizované spektrum:

$$\text{cov}(y_t, y_{t+k}) = \int_{-\pi}^{\pi} s_y(\omega)e^{i\omega k}d\omega$$

- ▶ Teda

$$s_y(\omega) = A(e^{i\omega})A(e^{-i\omega})s_x(\omega)$$

Príklad 2: AR(1) proces iným postupom

- ▶ Bez ujmy na všeobecnosti nech $\delta = 0$ (rovnaké autokovariancie), proces zapíšeme v tvare

$$x_t = \alpha x_{t-1} + u_t \Rightarrow u_t = (1 - \alpha L)x_t$$

- ▶ Definujme $A(z) = 1 - \alpha z$, potom

$$u_t = A(L)x_t \Rightarrow s_u(\omega) = A(e^{i\omega})A(e^{-i\omega})s_x(\omega)$$

- ▶ Spektrum bieleho šumu už poznáme: $s_u(\omega) = \frac{\sigma^2}{2\pi}$
- ▶ Platí:

$$\begin{aligned} A(e^{i\omega})A(e^{-i\omega}) &= (1 - \alpha e^{i\omega})(1 - \alpha e^{-i\omega}) \\ &= 1 - \alpha(e^{i\omega} + e^{-i\omega}) + \alpha^2 = 1 - \alpha^2 - 2\alpha \cos(\omega) \end{aligned}$$

- ▶ A teda

$$s_x(\omega) = \frac{s_u(\omega)}{A(e^{i\omega})A(e^{-i\omega})} = \frac{\sigma^2}{2\pi} \frac{1}{1 - \alpha^2 - 2\alpha \cos(\omega)}$$

Príklad 3: Vplyv priemerovania

- ▶ Nech x_t je stacionárny proces.
- ▶ Definujme

$$y_t = \frac{1}{4}(x_t + x_{t-1} + x_{t-2} + x_{t-3}).$$

- ▶ Pre ktoré hodnoty ω je spektrum nulové? Akej frekvencii to zodpovedá?

Cvičenia na samostatné počítanie

Cvičenie 1.

Nájdite spektrum stacionárneho sezónneho procesu

$$x_t = \delta + \alpha x_{t-12} + u_t$$

Cvičenie 2.

Nech x_t je stacionárny proces. Definujme

$$y_t = \frac{1}{3}(x_{t-1} + x_t + x_{t+1}).$$

Pre ktoré hodnoty ω je spektrum procesu y nulové? Akej frekvencii to zodpovedá?

Cvičenie 3.

Nájdite spektrum stacionárneho ARMA(1,1) procesu

$$x_t = \delta + \alpha x_{t-1} + u_t - \beta u_{t-1}$$

Návod: Na určenie A nemusíme mať dané koeficienty a_s z predpisu $A(z) = \sum_{s=-\infty}^{\infty} a_s z^s$, môžeme mať napríklad zadané napríklad

$$A(z) = (1 - 0.5z)^{-1}.$$

S týmto tiež vieme pracovať, napríklad

$$A(e^{i\omega}) = (1 - 0.5e^{i\omega})^{-1} = \frac{1}{1 - 0.5e^{i\omega}}$$