Cvičení 15.4.2013

Periodogram Pro časovou řadu $\{y_t\}$ definujeme periodogram jako

$$I(\omega) = \frac{1}{4\pi}\left(a^2(\omega) + b^2(\omega)\right),$$

kde

$$a(\omega) = \sqrt{\frac{2}{n}}\sum_{t=1}^{n} y_t \cos(\omega t), \quad b(\omega) = \sqrt{\frac{2}{n}}\sum_{t=1}^{n} y_t \sin(\omega t).$$

1. Uvažujte simulovanou časovou řadu $\{y_t, t = 1, \ldots, 100\}$, která se řídí modelem

$$y_t = 20 + 2 \cdot \cos\left(\frac{2\pi}{5}t + \frac{\pi}{4}\right) + \varepsilon_t, \tag{1}$$

kde ε_t je bílý šum, $\varepsilon_t \sim \mathsf{N}(0, 2)$.

   (a) Nechte si nasimulovat takovou řadu a nakreslete si její graf.
   (b) Spočtěte „ručně“ periodogram dané řady na síti $\omega = \frac{2\pi j}{n}$ pro $j = 1, \ldots, [(n-1)/2]$. Nechte si vykreslit jeho graf. Co lze z grafu pozorovat?
   (c) Přidejte si do (1) další periodický člen s jinou frekvencí, např. $3\cos\left(\frac{2\pi}{12}t\right)$. Opět si prohlédněte graf a periodogram řady.
   (d) Spočítejte si periodogram pomocí funkce spec.pgram, a to s následujícími volbami:
   `s=spec.pgram(y,span=NULL,taper=0,log="no",demean=FALSE,detrend=FALSE,fast=FALSE)`
   Zjistěte, jak se liší periodogram z R od periodogramu, který jste počítali ručně.
   (e) Můžeme se podívat také na tzv. kumulovaný periodogram, který se často používá při reziduální diagnostice:
   `cpgram(y)`

2. Podívejte se na periodogram časové řady beer z minula:

   (a) Nejprve použijte funkci spec.pgram přímo na řadu beer. Co vidíme?
   (b) Změňte detrend=FALSE na detrend=TRUE. Pomohla tato změna?
   (c) Spočítejte periodogram pro detrendovanou řadu, tj. na řadu beer, od níž odečtete trend získaný pomocí metody klouzavých průměrů (decompose). Co vidíme nyní? Jaké periodicity lze pozorovat?

Fisherův test testuje

$$H_0 : y_t = \mu + \varepsilon_t,$$

kde ε_t jsou iid $\mathsf{N}(0, \sigma^2)$ proti alternativě

$$H_1 : y_t = \mu + \sum_{i=1}^{k}[\alpha_i \cos(\omega_i t) + \beta_i \sin(\omega_i t)] + \varepsilon_t.$$

Uvažujme $\omega_j = \frac{2\pi j}{n}$ pro $j = 1, \dots, m = [(n-1)/2]$ a označíme

$$Y_j = \frac{I(\omega_j)}{\sum_{i=1}^m I(\omega_i)}.$$

Testová statistika Fisherova testu je volena jako $W = \max_{j=1,\dots,m} Y_j$ a má rozdělení

$$\mathsf{P}(W > w) = 1 - \sum_{i=0}^{\lfloor 1/w \rfloor} (-1)^i \binom{m}{i} (1 - iw)^{m-1}, \quad \text{pro } 0 < w < 1,$$

což lze upravit jako

$$\mathsf{P}(W > w) = m(1-w)^{m-1} - \binom{m}{2}(1-2w)^{m-1} + \binom{m}{3}(1-3w)^{m-1} - \cdots,$$

kde bereme $\lfloor 1/w \rfloor$ sčítanců. Pro $m < 50$ dostaneme velmi přesnou aproximaci použitím $\mathsf{P}(W > w) \doteq m(1-w)^{m-1}$.

1. Proveďte Fisherův test na simulované periodické řadě: spočítejte hodnotu testové statistiky, příslušnou p-hodnotu a zjištěnou významnou frekvenci.

2. Pro simulovanou periodickou časovou řadu proveďte Fisherův test opakovaně, dokud nenaleznete všechny signifikantní frekvence.

3. Vyzkoušejte funkčnost Fisherova testu na simulovaném normálním bílém šumu.

ODHAD MODELU

1. Pro simulovanou časovou řadu jste detekovali významné frekvenci. Odhadněte tedy pomocí lineární regrese výsledný model. Výsledky porovnejte se zadanými koeficienty.

2. Rezidua z modelu posuďte pomocí kumulovaného periodogramu.

3. Soubor mink.csv obsahuje časovou řadu ročních prodejů norkových kožichů společnosti Hudson's Bay Company v letech 1848 až 1908.
   (a) Načtěte si a znázorněte si danou časovou řadu. Lze v ní očekávat nějakou periodicitu?
   (b) Vykreslete si periodogram dané řady. Kterou periodicitu detekujeme jako podezřelou?
   (c) Proveďte na řadu Fisherův test a nalezněte všechny signifikantní frekvence.
   (d) Odhadněte vhodný model.
   (e) Nakonec zkontrolujte rezidua pomocí kumulativního periodogramu.

CO NA CVIČENÍ DĚLAT NEBUDEME: testy náhodnosti (bude jen v domácím úkolu).