Neparametriskās statistikas pielietojumi spektrālajā analīzē

Agris Vaselāns

Latvijas Universitāte

27. oktobris, 2010

Spektrs (spektrālā blīvuma funkcija)

Teorēma

Ja $\gamma(h)$ ir stacionāra procesa $\{x_t\}$ autokovariāciju funkcija, turklāt

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty,$$

tad

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i\omega h} f(\omega) d\omega,\ h = 0, \pm 1, \pm 2, \ldots$$

un procesa $\{x_t\}$ spektrālā blīvuma funkcija ir

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i\omega h}, -1/2 \leqslant \omega \leqslant 1/2.$$

Spektrālās blīvuma funkcijas īpašības

Spektrālā blīvuma funkcija ir analoģiska sadalījuma (varbūtību) blīvuma funkcijai:

- $f(\omega) \geqslant 0, \forall\omega$
- $f(\omega) = f(-\omega)$
- $f(\omega + 1) = f(\omega)$ (periods 1)

Autokovariāciju funkcija izsaka informāciju par procesu $\{x_t\}$ izmantojot lagus, bet spektrālā blīvuma funkcija to izsaka ciklos (cikli novērotā procesa laika atstatumā (cikli/gadā; cikli/mēnesī utt.))

Frekvenču izvēle

Šajā darbā pārsvarā izmantosim frekvences ω, kas norādīs ciklus novērojumu laika vienībā (cikli/punktā).

- statistikā analizējot datus bieži tie ir laikrindu vai ekonomiska rakstura;
- R

Dažkārt ar ω apzīmē frekvenci π radiānos, proti $\lambda = 2\pi\omega$.

- spektrālā analīze spēcīgi attīstījusies fizikā, kur procesu aprakstam tieši izmanto frekvences π radiānos;
- Matlab

Frekvenču izvēle

Šajā darbā pārsvarā izmantosim frekvences ω, kas norādīs ciklus novērojumu laika vienībā (cikli/punktā).

- statistikā analizējot datus bieži tie ir laikrindu vai ekonomiska rakstura;
- R

Dažkārt ar ω apzīmē frekvenci π radiānos, proti $\lambda = 2\pi\omega$.

- spektrālā analīze spēcīgi attīstījusies fizikā, kur procesu aprakstam tieši izmanto frekvences π radiānos;
- Matlab

ARMA teorētiskais spektrs

Definīcija

Process (laikrinda) $\{x_t; t = 0, \pm 1, \pm 2, \ldots\}$ ir ARMA(p, q), ja tas ir stacionārs un

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q},$$

kur $\theta_q \neq 0$, $\phi_p \neq 0$ un $\sigma_\varepsilon^2 > 0$.

Apgalvojums

ARMA teorētiskais spektrs ir pierakstāms formā:

$$f(\omega) = \sigma_\varepsilon^2 \frac{|1 + \sum_{k=1}^{q} \theta_k e^{-2\pi i \omega k}|}{|1 - \sum_{m=1}^{p} \phi_m e^{-2\pi i \omega m}|}.$$

Periodogramma un DFT

Definīcija

Ja doti dati $x_1, x_2, \ldots, x_n$, tad par diskrēto Furjē transformāciju (DFT) sauc

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} x_t e^{-2\pi i \omega_j t},$$

kur $j = 0, 1, \ldots, n-1$ un frekvences $\omega_j = j/n$ tiek sauktas par Furjē jeb fundamentālajām frekvencēm.

Definīcija

Ja doti dati $x_1, x_2, \ldots, x_n$, tad par periodogrammu sauc

$$I(\omega_j) = |d(\omega_j)|^2$$

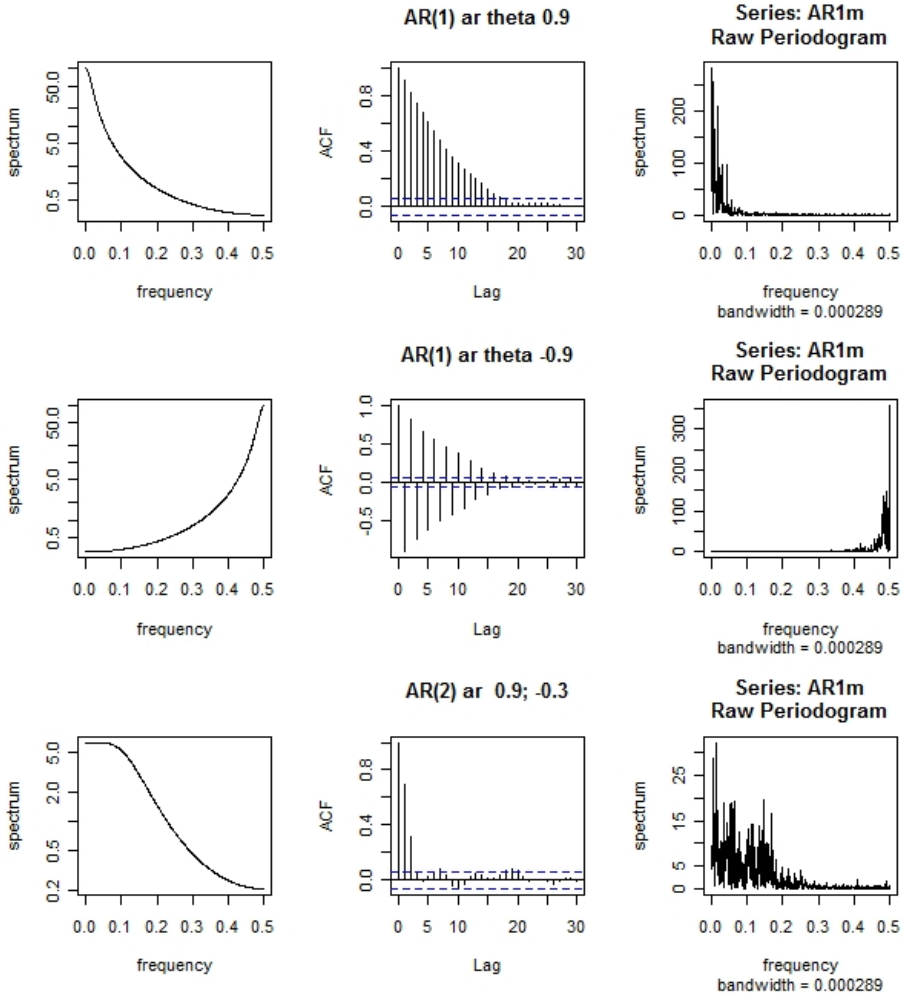
kur $j = 0, 1, \ldots, n-1$.

Periodogramma un DFT

Šeit jāpiezīmē, ka $I(0) = n\overline{x}^2$, kur $\overline{x}$ ir izlases vidējā vērtība. Bet, ja $j \neq 0$, tad $I(\omega_j) = \sum_{h=-(n-1)}^{n-1} \hat{\gamma}(h) e^{2\pi i \omega_j h}$, kur $\hat{\gamma}(h)$ ir izlases autokovariāciju funkcija:

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \overline{x})(x_t - \overline{x}).$$

AR(p) procesu spektri



Gludinātā (smoothed - averaged) periodogramma

Frekvenču josla $\mathcal{B}$ no $L << n$ sekojošām fundamentālām frekvencēm, kas centrētas ap $\omega_j = j/n$, kas tuva interesējošai frekvencei ω:

$$\mathcal{B} = \{\omega : \omega_j - m/n \leqslant \omega \leqslant \omega_j + m/n\},$$

kur $L = 2m + 1$. Lielumu $\mathcal{B}_\omega = L/n$ sauc par joslas platumu.

Definīcija

Ja $\mathcal{B}$ ir frekvenču josla no $L << n$ sekojošām fundamentālām frekvencēm, kas centrētas ap $\omega_j = j/n$, tad par vidējoto (averaged) periodogrammu sauc

$$\bar{f}(\omega) = \frac{1}{L} \sum_{k=-m}^{m} I(\omega_j + k/n).$$

Gludinātā (svērtā jeb kodolu) periodogramma un kodolu izvēle

Definīcija

Lai iegūtu precīzāku spektra novērtējumu, periodogrammu var gludināt izmantojot dažādus svarus, definēsim svērto periodogrammu:

$$\hat{f}(\omega) = \sum_{k=-m}^{m} h_k I(\omega_j + k/n),$$

kur svari h_k apmierina nosacījumus $\forall k : h_{-k} = h_k$ un $\sum_{k=-m}^{m} h_k = 1$.

Svaru h_k izvēlei tiek ieteikti Daniela, Fejer, Dirihlē un modificētais Daniela kodols.

Parametra m un joslas platuma izvēles problēma

Šajā gadījumā parametrs L jāaizvieto ar $L_h = (\sum_{k=-m}^{m} h_k^2)^{-1}$, tad joslas platums $\mathcal{B}$ svērtajai periodogrammai būs $\mathcal{B} = L_h/n$. Tātad svarīga ir parametra L vai m (window closing) izvēle.

Ieteikumi

- (līdzīgi kā statistikā sākotnēji histogrammas joslas platumam) tiek ieteikts izvēlēties $m = \sqrt{n}$.
- vēl piedāvāti: $m = 2\sqrt{n}$, $m = 0.5\sqrt{n}$
- As Hannan (1973, p. 311) says, "Experience is the real teacher and cannot be got from a book."

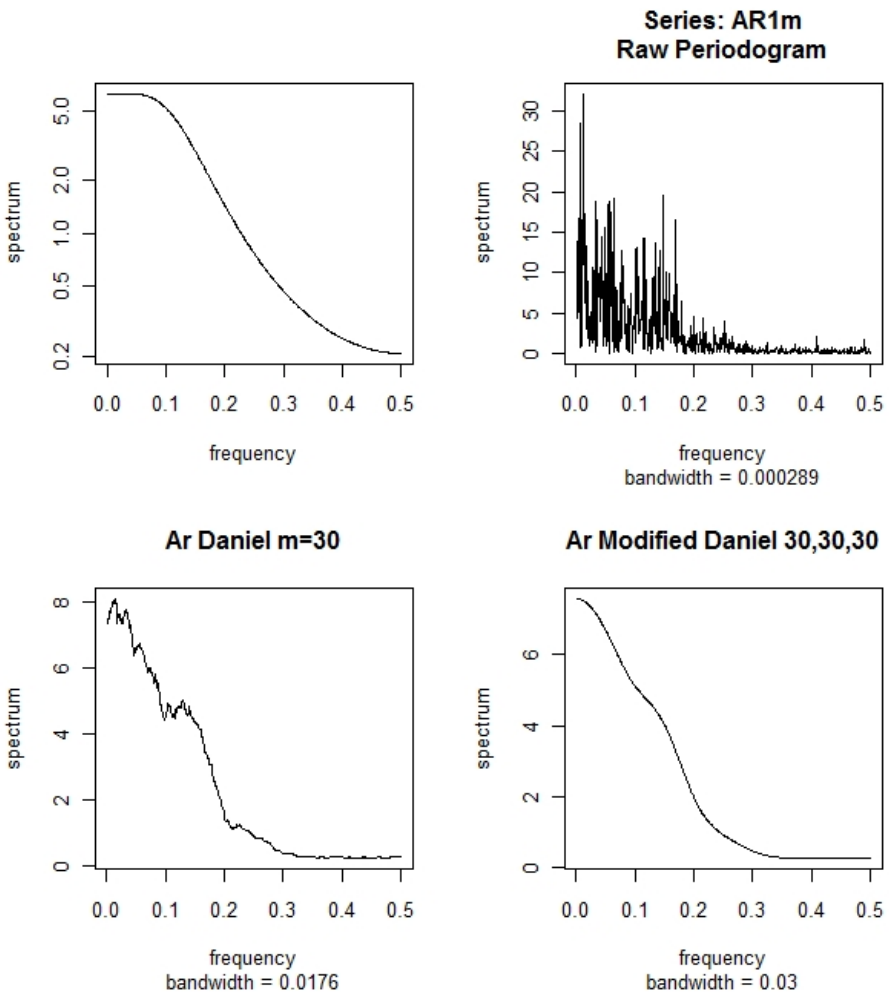
Parametra m un joslas platuma izvēles problēma

Šajā gadījumā parametrs L jāaizvieto ar $L_h = (\sum_{k=-m}^{m} h_k^2)^{-1}$, tad joslas platums $\mathcal{B}$ svērtajai periodogrammai būs $\mathcal{B} = L_h/n$. Tātad svarīga ir parametra L vai m (window closing) izvēle.

Ieteikumi

- (līdzīgi kā statistikā sākotnēji histogrammas joslas platumam) tiek ieteikts izvēlēties $m = \sqrt{n}$.
- vēl piedāvāti: $m = 2\sqrt{n}$, $m = 0.5\sqrt{n}$
- As Hannan (1973, p. 311) says, "Experience is the real teacher and cannot be got from a book."

Gludinātais AR(2)



Punktveida ticamības intervāli

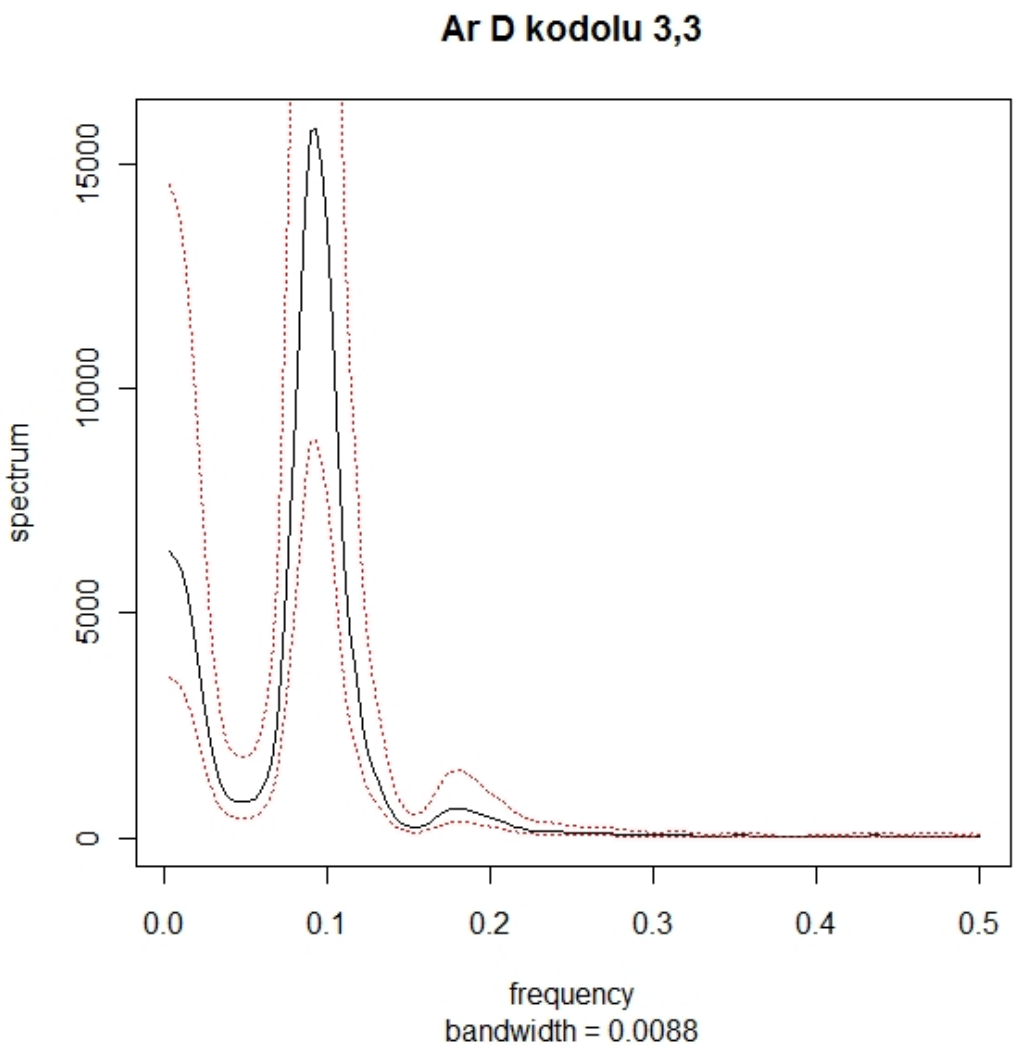
Ja izpildās iepriekšminētie nosacījumi par h_k, turklāt $m \to \infty$, kad $n \to \infty$ un $m/n \to 0$, tad var pierādīt, ka:

$$\frac{2L_h \hat{f}(\omega)}{f(\omega)} \sim \chi^2_{2L_h}$$

Tātad ticamības intervāls:

$$\frac{2L_h \hat{f}(\omega)}{\chi^2_{2L_h}(1-\alpha/2)} \leq f(\omega) \leq \frac{2L_h \hat{f}(\omega)}{\chi^2_{2L_h}(\alpha/2)}$$

Punktveida ticamības intervāli



Tapering

Tā kā spektra novērtējumā periodogrammas definēšanai izmatojām diskrēto Furjē transformāciju, tad spektra novērtējumā diezgan būtiskas novirzes dod novērotā procesa sākuma un beigu vērtības. Lai mazinātu šo datu ietekmi, lieto t.s. "taperošanu ".
Aizvietojot novērotos datus x_t ar $y_t = h_t x_t$, $t = 1, 2, \ldots, n$ un tad pielietojot DFT:

$$d_y(\omega_j) = n^{-1/2} \sum_{t=1}^{n} h_t x_t e^{-2\pi i \omega_j t},$$

un definējot šiem datiem periodogrammu $I_y(\omega_j) = |d(\omega_j)|^2$.

Tapering

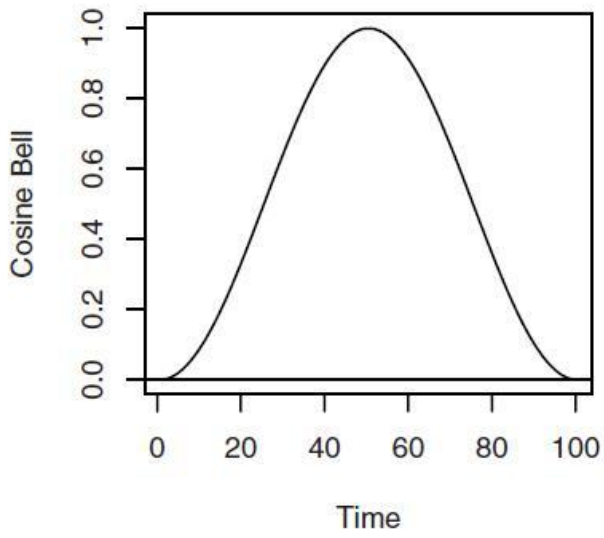
Viens no visbiežāk pielietotajiem ir cosine bell vai split cosine bell:

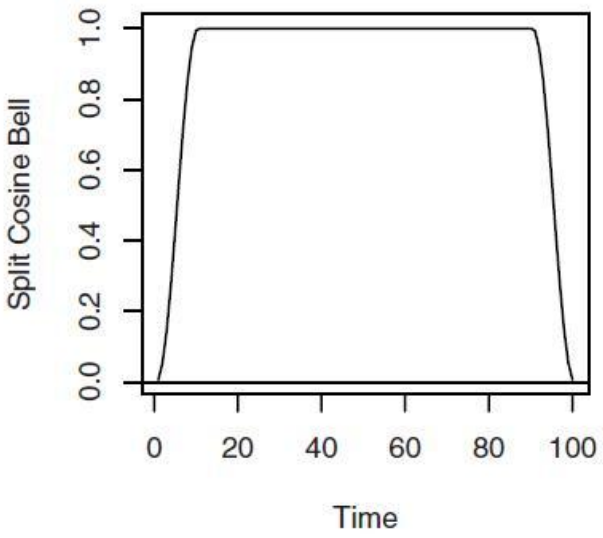
$$h_t = \frac{1}{2}\left\{1 - \cos\left[\frac{2\pi(t-0.5)}{n}\right]\right\}$$

$$h_t = \begin{cases} \frac{1}{2}\left\{1 - \cos\left[\frac{\pi(t-1/2)}{m}\right]\right\} & \text{for } 1 \le t \le m \\ 1 & \text{for } m+1 \le t \le n-m \\ \frac{1}{2}\left\{1 - \cos\left[\frac{\pi(n-t+1/2)}{m}\right]\right\} & \text{for } n-m+1 \le t \le n \end{cases}$$

100p% cosine bell taper, kur p = 2m/n.

Cosine Bell un 10% Taper Split Cosine Bell ar n = 100





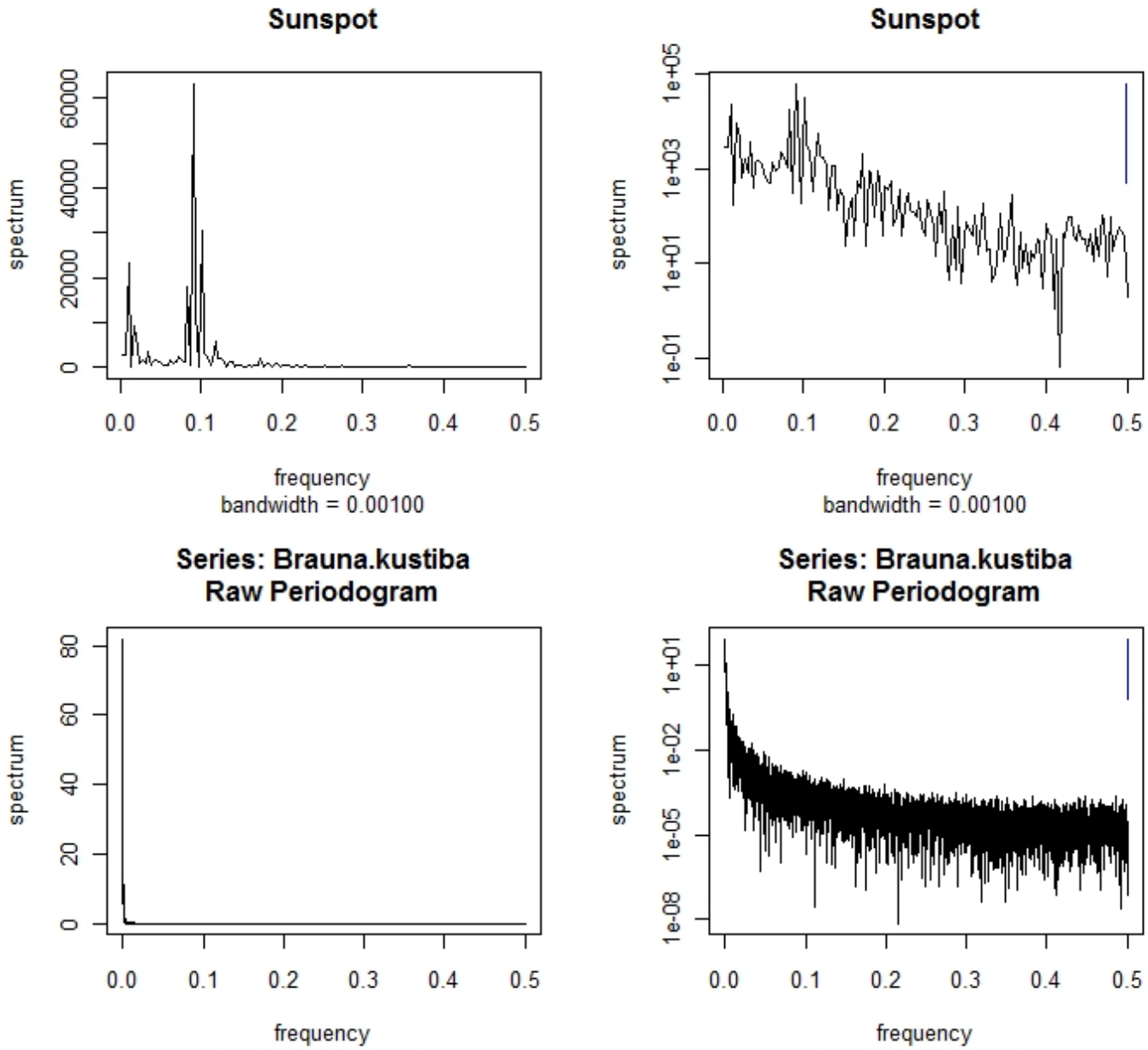
Log vai lineārā skala

Log skala

- lai aplūkotu lielas un mazas amplitūdas signālus vienlaicīgi, proti, log skalā katrai frekvences amplitūdai ir vienāds "nozīmīgums";
- Lietojot logaritmisko skalu mazi signāli ir vieglāk pamanāmi;
- ja ir kāda liela frekvence, kas "traucē"pamanīt citas, kas arī ir būtiskas;
- biežāk izmantojama fizikālu eksperimentu analizēšanai.

Savukārt lineārā skala labi parāda frekvenču amplitūdu atšķirības un biežāk izmantojama ekonomisku utml. datu analīzē.

Log vai lineārā skala



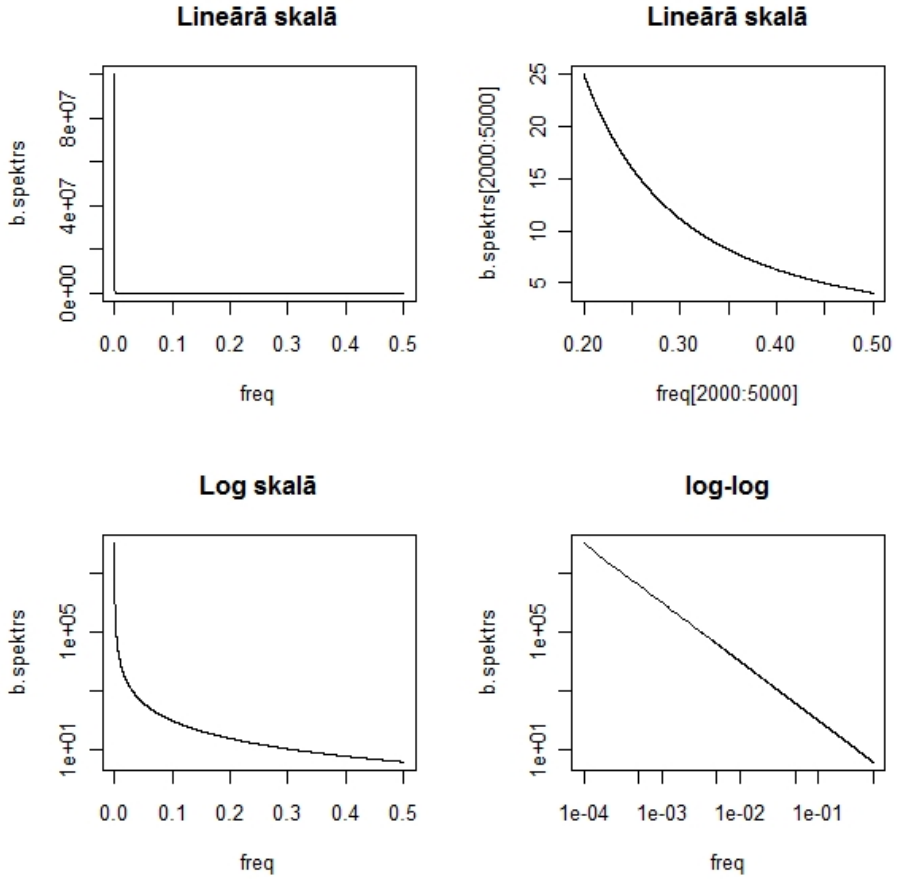
Brauna kustības teorētiskais spektrs

Viens no veidiem kā raksturot dažāda tipa (white, pink, Brownian) trokšņus ir aplūkot to spektrus. Teorētiskais trokšņu spektrs ir $1/\omega^{\beta}$, , beta ≥ 0.

- ja $\beta = 0$ tad process ir baltais troksnis;
- $\beta = 2$ - Brauna troksnis jeb kustība (arī sarkanais troksinis);
- $0 < \beta < 2$, - rozā troksnis.

Skaidrs, ka attēlojot spektru log-log skalā, taisnes virziena koeficients norādīs β.

Brauna kustības teorētiskais spektrs

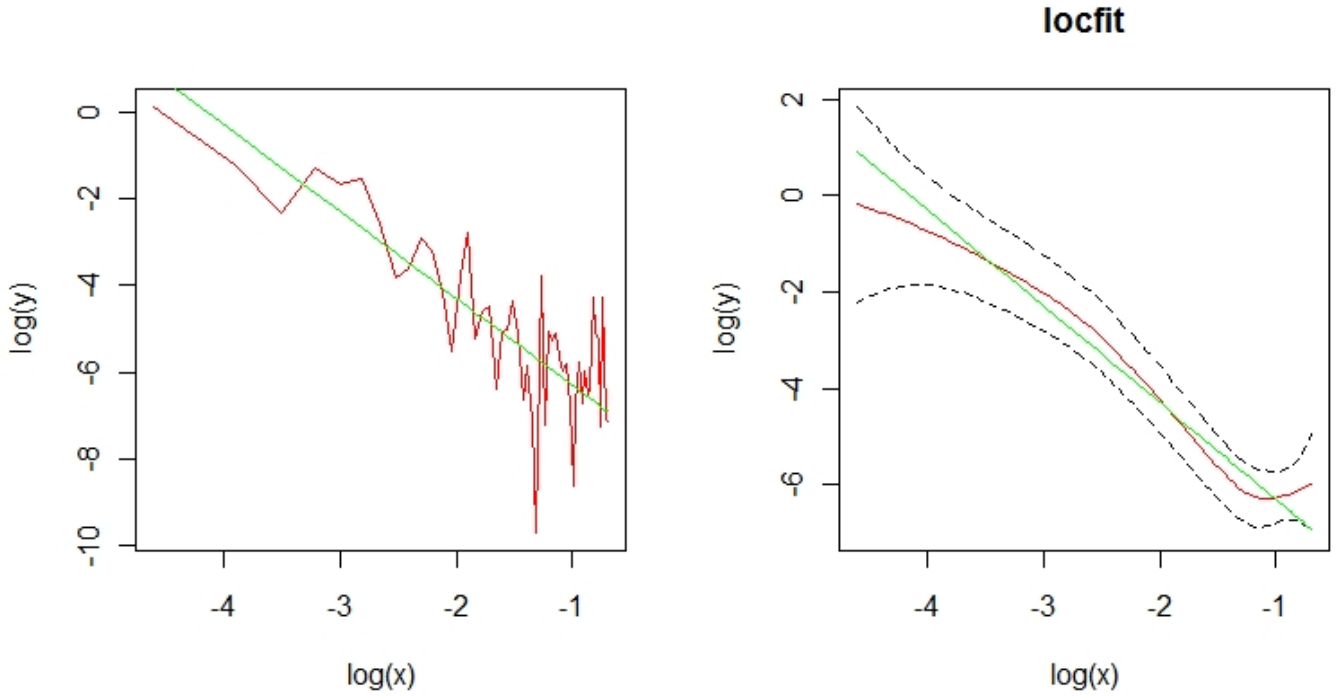


att.: Brauna kustības teorētiskais spektrs dažādās skalās

Brauna kustības spektra novērtējums ar lokālās regresijas palīdzību un ticamības joslas; pārbaude virziena koeficientam

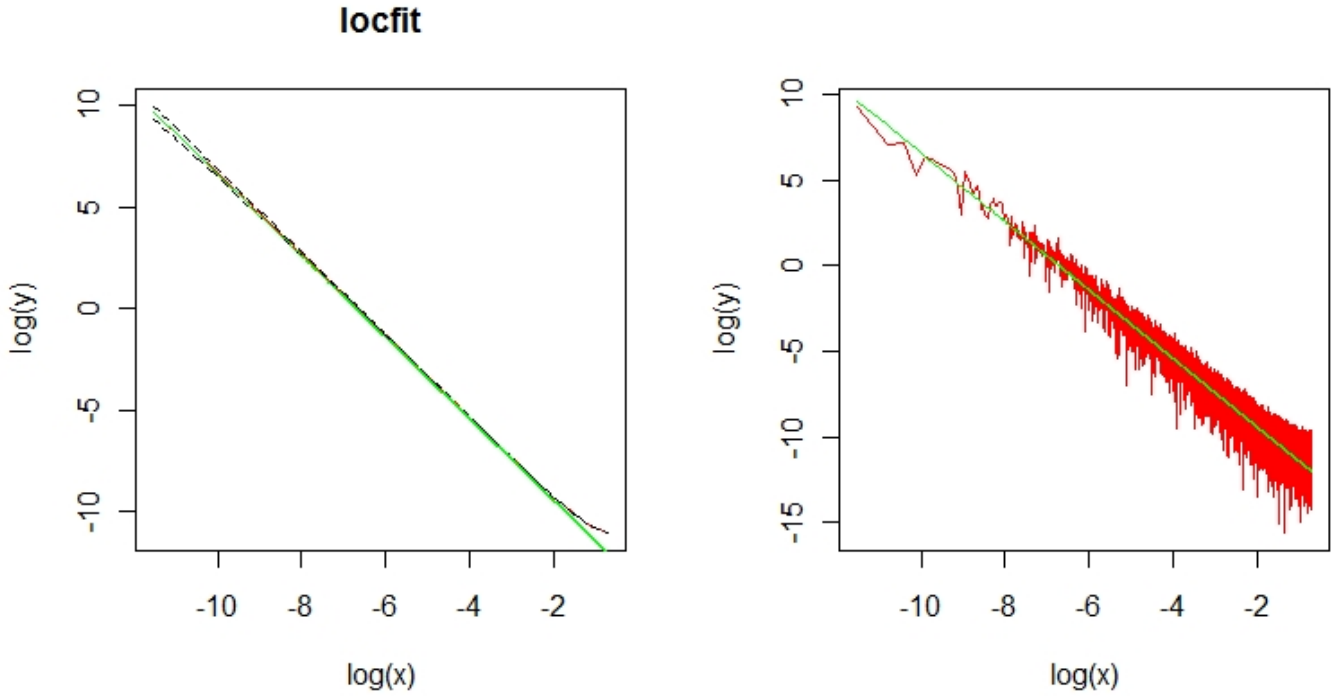
Tātad viens no veidiem, kā pārbaudīt, vai process nav Brauna kustība, ir konstruēt periodogrammas lokālās regresijas novērtējumu un tad uzzīmējot vienlaicīgās ticamības joslas teorētiskajam spektram. Tātad veikt grafisko pārbaudi.

Brauna kustības spektra novērtējums ar lokālās regresijas palīdzību un ticamības joslas; pārbaude virziena koeficientam



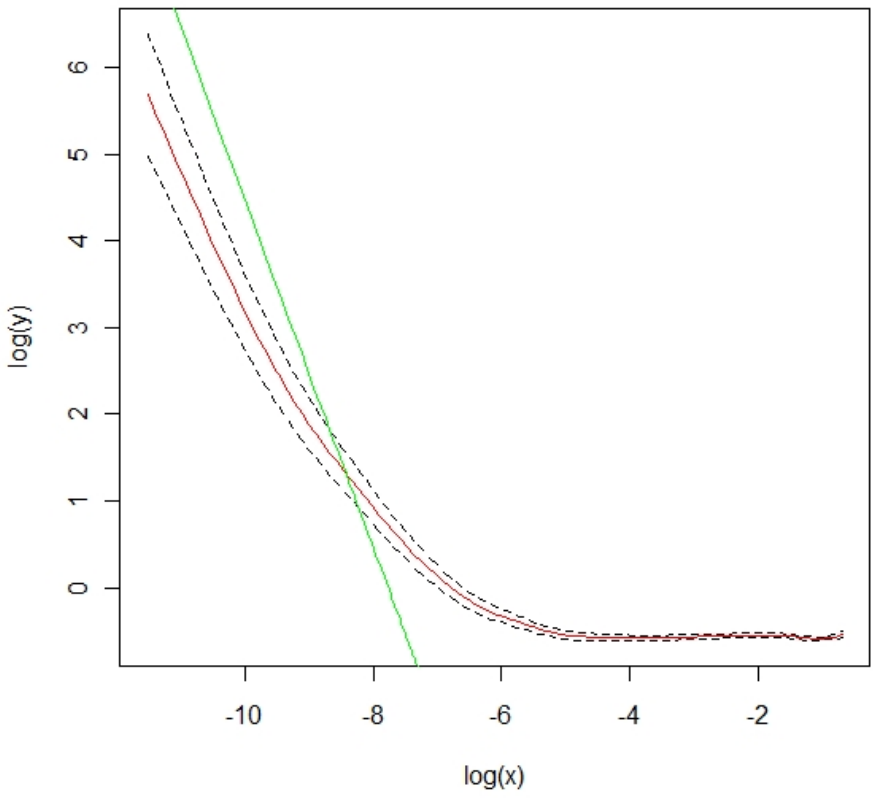
att.: Brauna kustības pārbaude, N = 100

Brauna kustības spektra novērtējums ar lokālās regresijas palīdzību un ticamības joslas; pārbaude virziena koeficientam



att.: Brauna kustības pārbaude, N = 10000

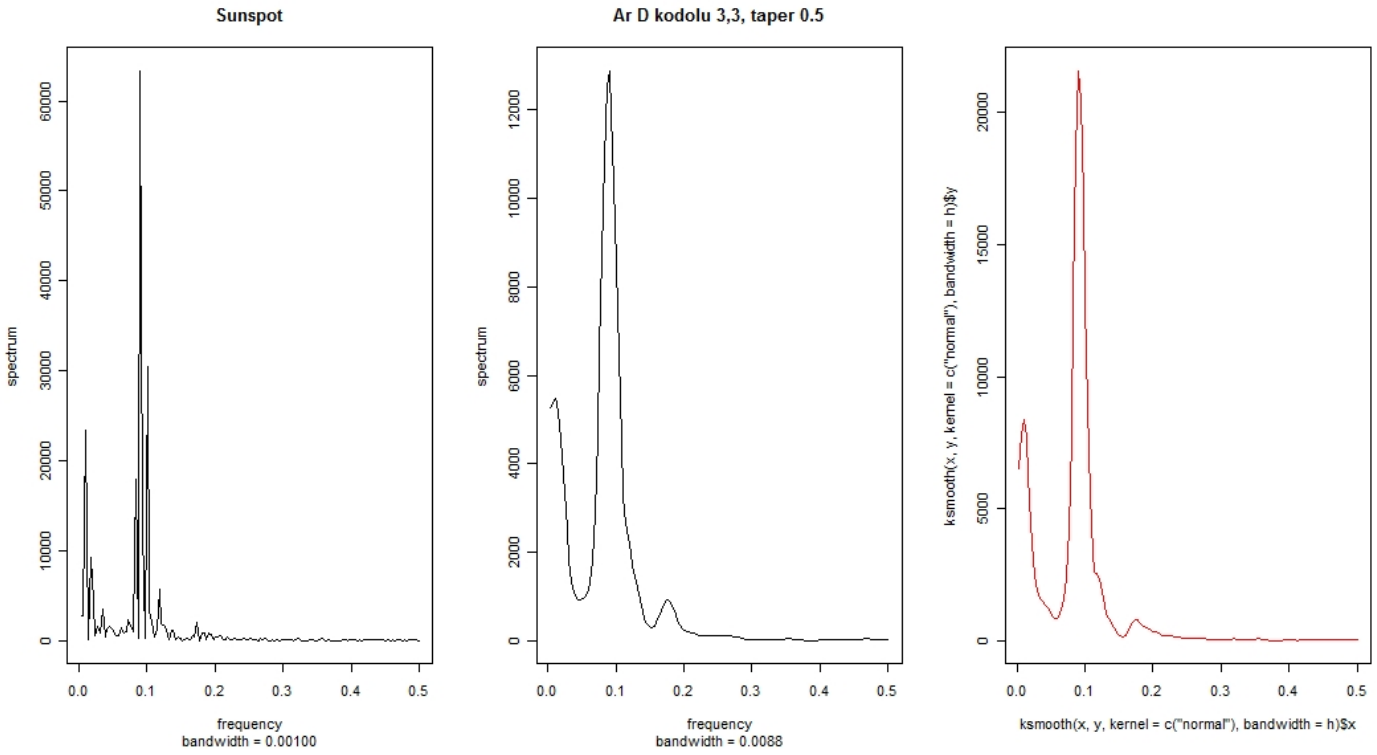
Brauna kustības spektra novērtējums ar lokālās regresijas palīdzību un ticamības joslas; pārbaude virziena koeficientam



att.: Brauna kustības pārbaude, N = 10000, (BM() – rnorm)

Kodolu gludinātā periodogramma

Viena no idejām ir izmantot jau statistikā zināmo regresijas kodolu gludināšanu .



Literatūra

1. Shumway/Stoffer: Time Series Analysis and Its Applications, 2nd ed.
2. Cryer/Chan: Time Series Analysis, Second Edition
3. Cowpertwait/ Metcalfe Introductory Time Series with R
4. publikācijas ..