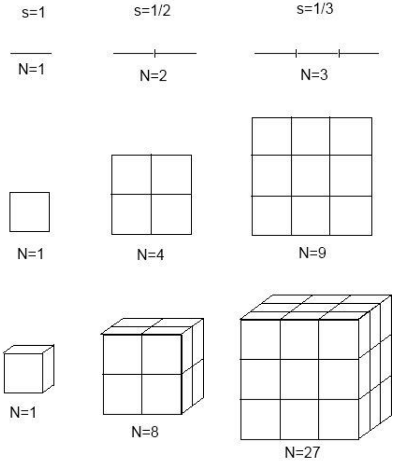
Laiko eilučių fraktališkumas: Hursto laipsnio rodiklis ir jo įvertinimo metodai

Aleksejus Kononovičius

2012-10-11

Įprasta dimensija ir “dėžučių” skaičiavimo metodas



[http://www.wahl.org/fe/HTML_version/link/FE4W/c4.htm]

Ką gi turime?

$$N = \frac{1}{s^D},$$

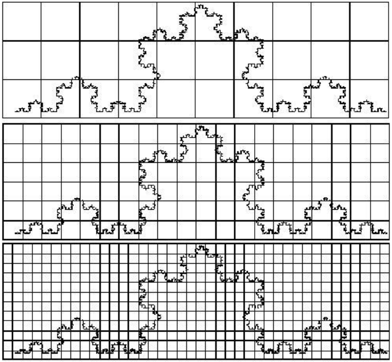
$$\ln N = D \ln s^{-1},$$

$$D = \frac{\ln N}{\ln s^{-1}}.$$

Bendriau:

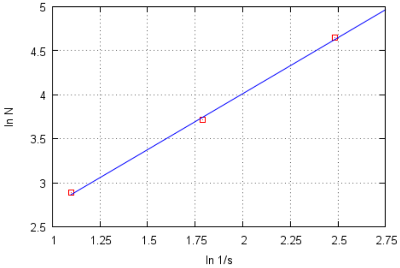
$$D = \lim_{s \to 0} \frac{\ln N}{\ln s^{-1}}.$$

Sudėtingesnis atvejis: Kocho snaigė

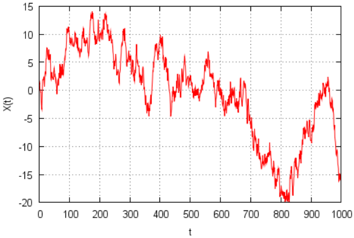

[http://www.wahl.org/fe/HTML_version/link/FE4W/c4.htm]

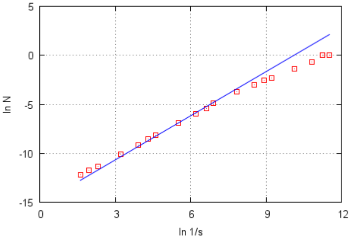
Ką gi turime?

Mastelyje 1/3 - 18 langelių, kuriuos kerta kreivė, 1/6 - 41, 1/12 - 105. Nusipiešę matome, kad $D \approx 1.27$.



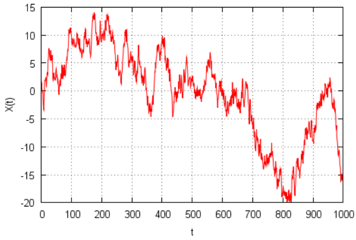
Brauno proceso laiko eilutės dimensija

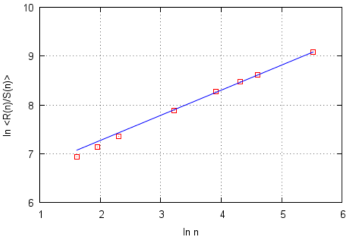




Šiek tiek pasistengus rezultatą galima gauti artimą teoriniam: $D = 1.517 \pm 0.061$.

Brauno proceso laiko eilutės Hursto laipsnio rodiklis





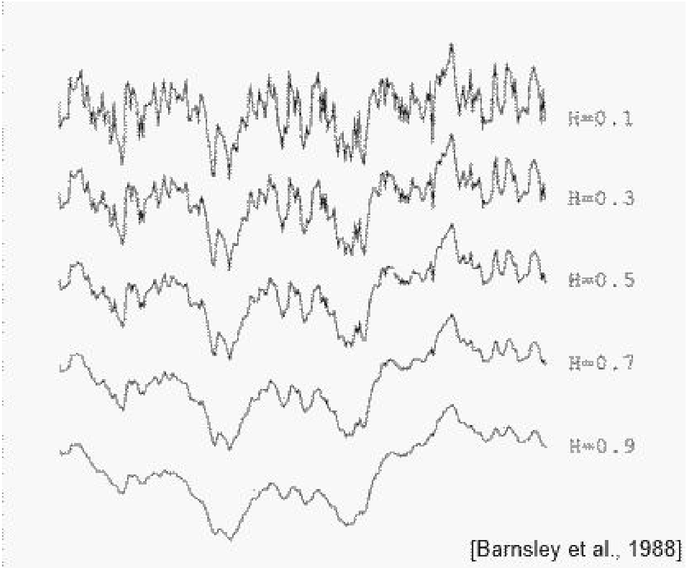
Šiek tiek pasistengus rezultatą gauname: $H = 0.514 \pm 0.006$. Čia H turi geometrinę

$$D = 2 - H = 1.486,$$

ir statistinę prasmes,

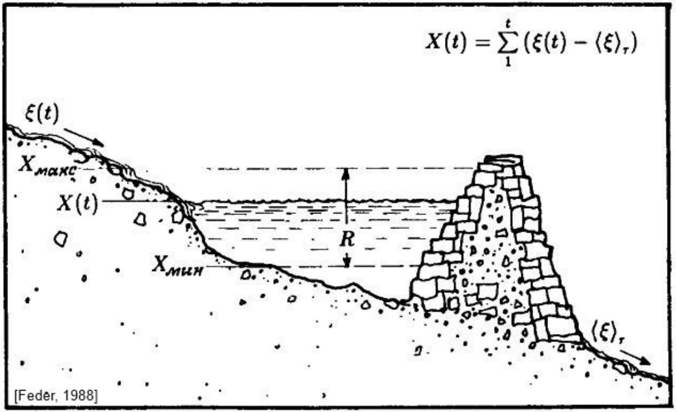
$$X(at) \stackrel{d}{=} a^H X(t).$$

Geometrinis Hursto laipsnio rodiklio interpretavimas



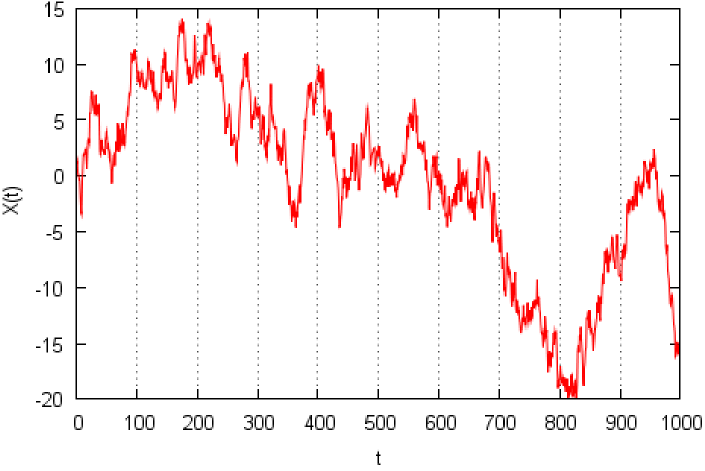
R/S, masteliuoto aukščio (rescaled range), metodas

Pagrindinė metodo idėja



Palyginęs įvairių upių ir ežerų vandens lygio duomenis Hurstas gavo, kad $\langle R(t)/S(t) \rangle \sim t^H$, kur $H = 0.726 \pm 0.082$.

Padalinkime $X(t)$ signalą į n dydžio segmentus. Tokiu atveju segmentų yra $N' = \text{floor}(N/n)$.



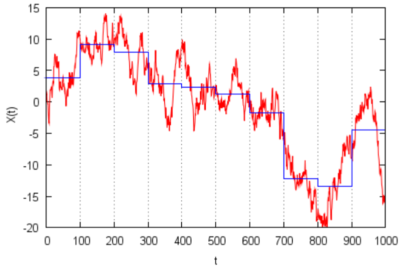
R/S, masteliuoto aukščio (rescaled range), metodas

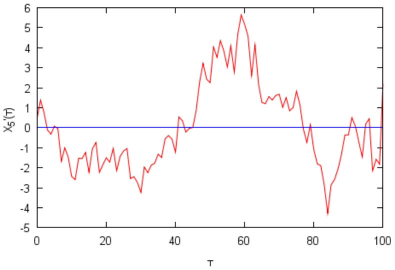
2 etapas. Vidurkių skaičiavimas

Kiekvienam signalo $X(t)$ segmentui apskaičiuojame vidurkius, $\langle x \rangle_i$. Atimame $\langle x \rangle_i$ iš visų to segmento duomenų (tą darome su visais segmentais):

$$X'_i(\tau) = X[(i-1)n + \tau] - \langle x \rangle_i, \quad \forall i \in [1; N'],$$

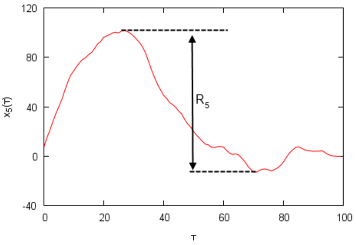
čia $X'_i(\tau)$ yra i-tojo segmento laiko eilutė, o τ yra "vidinis" segmento laikas.





Kiekvienam segmentui $X_i'(\tau)$ apskaičiuojame profilį:

$$x_i(\tau) = \sum_{s=0}^{\tau} X_i'(s), \quad \forall i.$$



Nustatome kiekvieno segmento **profilio** aukštį:

$$R_i = \max(w_i) - \min(w_i), \quad \forall i.$$

Nustatome kiekvieno segmento standartinį nuokrypį:

$$S_i = \sqrt{\frac{1}{n}\sum_{s=0}^{n-1}\left[X_i'(s)\right]^2}, \quad \forall i.$$

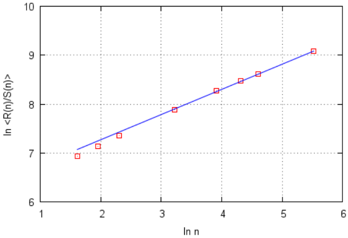
R/S, masteliuoto aukščio (rescaled range), metodas

4 etapas. Vidurkinimas per segmentus; 5 etapas. Keičiame n.

Santykį R/S vidurkiname per visus segmentus:

$$\left\langle \frac{R(n)}{S(n)} \right\rangle = \frac{1}{N'} \sum_{i=1}^{N'} \frac{R_i}{S_i} \sim n^H.$$

Čia R ir S yra pažymėtos kaip funkcijos nuo n, nes toliau kartojame tuos pačius žingsnius su skirtingais n. Pavaizdavę $\left\langle \frac{R(n)}{S(n)} \right\rangle$ kaip priklausomybę nuo n log-log skalėje gausime H.



Šiuolaikiniai metodai

Metodai kuriuos nagrinėsiu:

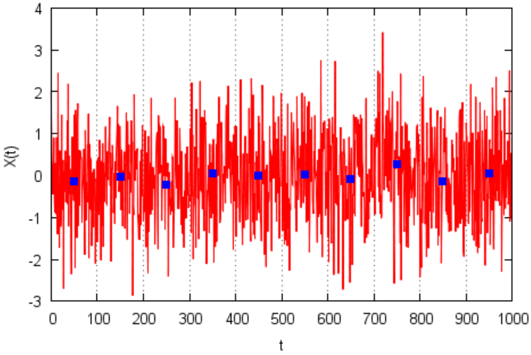
- Sukauptos variacijos (aggregated variance),
- Sukauptos eilutės modulio (modulus of aggregated series),
- Dispersinės analizės,
- Higuchi,
- Betendencių fliuktuacijų analizės (detrended fluctuation analysis).

Kurių neliesiu:

- Periodogramų metodai,
- Whittle metodas,
- Bangelių analize (wavelet analysis) paremti metodai.

Mažiname laiko eilutės ilgį iki $N' = \text{floor}\,(N/n)$ taškų vidurkindami:

$$X^{(n)}(k) = \frac{1}{n} \sum_{i=(k-1)n}^{kn-1} X(i), \quad \forall k \in [1; N'].$$



Kuri jau priklauso nuo metodo:

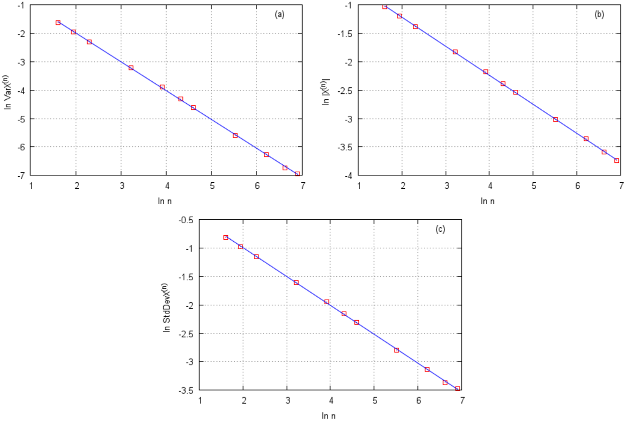
$$\mathrm{Var} X^{(n)} = \frac{1}{N'} \sum_{k=1}^{N'} \left[X^{(n)}(k) - \langle X^{(n)}(k) \rangle \right]^2 \sim n^{S_{var}},$$

$$\left| X^{(n)} \right| = \frac{1}{N'} \sum_{k=1}^{N'} \left| X^{(n)}(k) \right| \sim n^{S_{mod}},$$

$$\mathrm{StdDev} X^{(n)} = \frac{\sqrt{\mathrm{Var} X^{(n)}}}{\sqrt{\mathrm{Var} X^{(n_0)}}} \sim n^{S_{disp}}.$$

Sukauptų laiko eilučių metodai

3 etapas. Keičiame mastelį, n

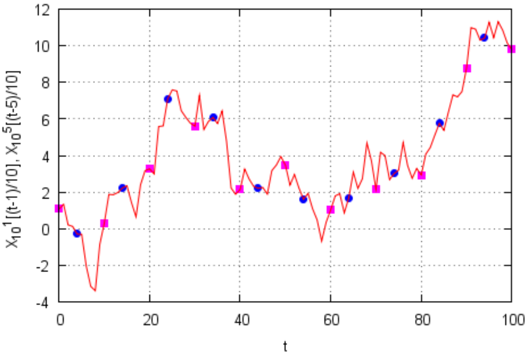


Gauti polinkiai: $S_{var} = -1.002$, $S_{mod} = -0.507$, $S_{disp} = -0.501$. Gautos H vertės: $H_{var} = 1 + \frac{S_{var}}{2} = 0.499$, $H_{mod} = 1 + S_{mod} = 0.493$, $H_{disp} = 1 + S_{disp} = 0.499$. R/S metodu: $H_{R/S} = 0.514$.

Metodo esmė taip pat yra skyros mažinimas, bet ne per vidurkinimą:

$$X_n^m(i) : X(m), X(m+n), X(m+2n), \dots$$

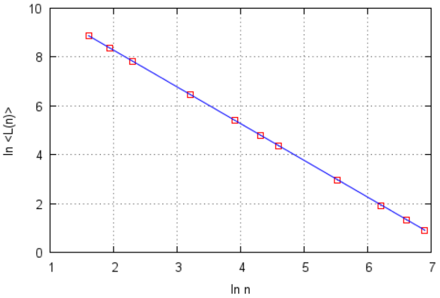
Šių eilučių ilgis yra $1 + N'$, kur $N' = \text{floor}\left[\frac{N-m}{n}\right]$.



Eilutės, kurios mastelis n ir pradžia m, ilgis:

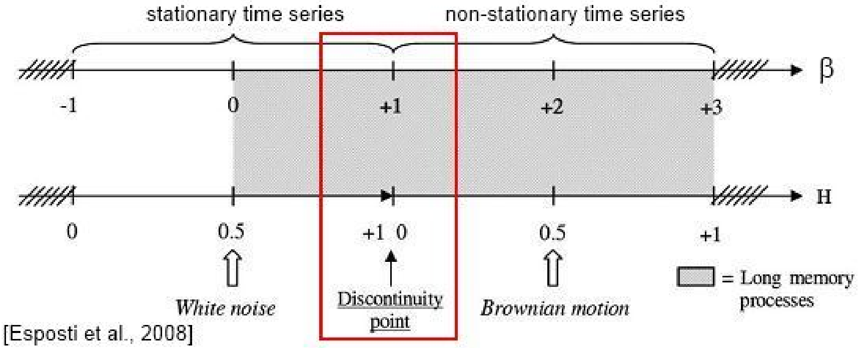
$$L_m(n) = \frac{N-1}{n^2} \cdot \frac{1}{N'} \sum_{i=1}^{N'} \left| X_n^m(i) - X_n^m(i-1) \right| .$$

Tikimės, kad $\langle L(n) \rangle \sim n^{S_{hig}}$, kur $S_{hig} = -D$.



$S_{hig} = -1.501 \quad \Rightarrow \quad H_{hig} = 2 + S_{hig} = 0.499.$

Kuom šie metodai mums netinka?

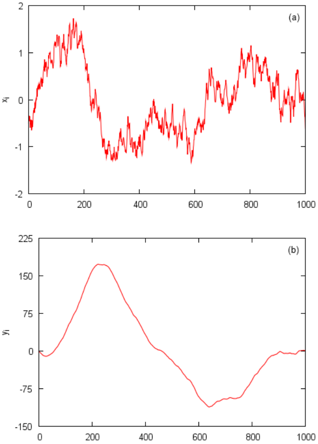


Nes mums rūpi $S(f) \sim 1/f^{\beta}$ su $\beta \approx 1$! O esant arti šio taško arba tektų naudoti du skirtingus metodus arba integruoti ar diferencijuoti.

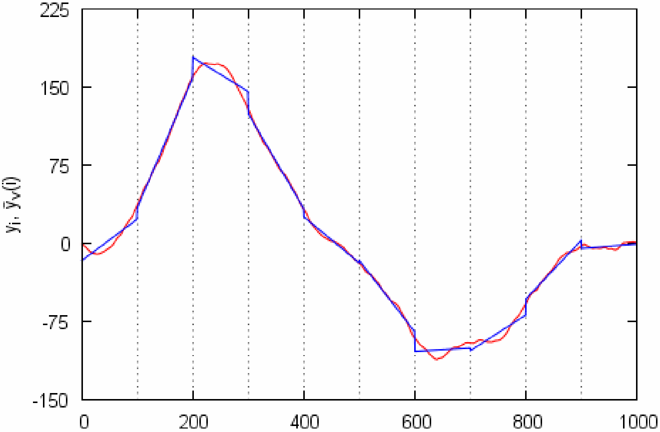
Skaičiuojame laiko eilutės, x_i, profilį (time series profile):

$$y_i = \sum_{k=1}^{i} \left[x_k - \langle x \rangle \right].$$

Dėl tokio apibrėžimo profilis dažnai turi naudingą savybę - bent viena (paskutinė) profilio taško reikšmė yra lygi nuliui.
Pradinė profilio vertė dažnai yra artima nuliui.



Padalijame profilį į n dydžio segmentus ($N' = \mathrm{floor}(N/n)$), kuriuose nustatome profilio tendencijas, $\bar{y}_\nu(i)$.



ν-tojo, $\nu \in [1; N']$, segmento kvadratinė fluktuacinė funkcija:

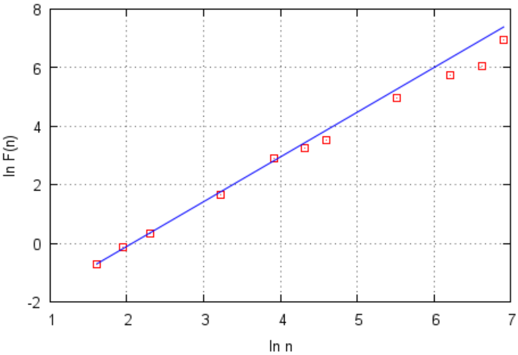
$$F_\nu^2(n) = \frac{1}{n} \sum_{i=1}^{n} \left[y_{(\nu-1)n+i} - \bar{y}_\nu(i) \right]^2 .$$

O jei turime ir segmentų einančių iš profilio pabaigos, $\nu \in [N'+1; 2N']$:

$$F_\nu^2(n) = \frac{1}{n} \sum_{i=1}^{n} \left[y_{N-(\nu-N')n+i} - \bar{y}_\nu(i) \right]^2 .$$

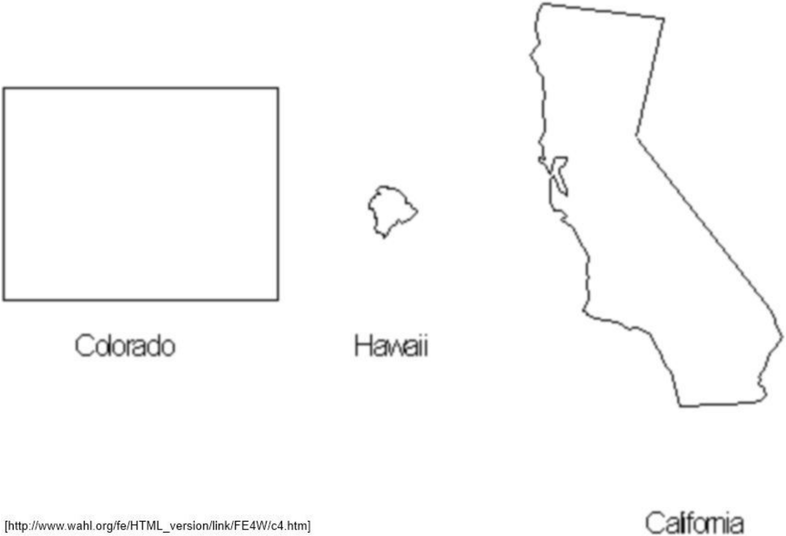
Fluktuacijų vidurkis:

$$F(n) = \sqrt{\langle F_\nu^2(n) \rangle} \sim n^h .$$



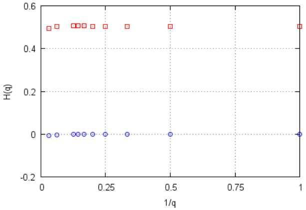
Gauname polinkį: $h = 1.52 \pm 0.04$. Jis yra susijęs su neigiamu spektro polinkiu: $\beta = 2h - 1$. Atitinkamai: $H = h - 1$ (nestacionaru, $h > 1$) arba $H = h$ (stacionaru, $h < 1$).

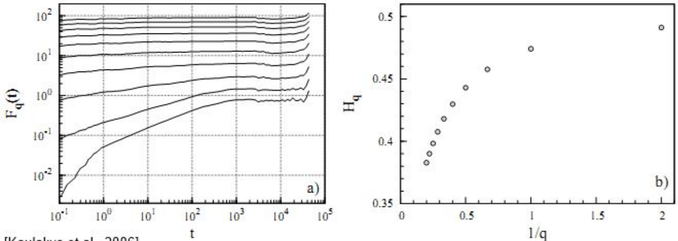
Tačiau vis dar kažko trūksta...



Apibendrintas aukščio-aukščio koreliacijos metodas

$$F_q(\tau) = \langle |y(t+\tau) - y(t)|^q \rangle^{1/q} \sim \tau^{H(q)}, \quad q > 0.$$





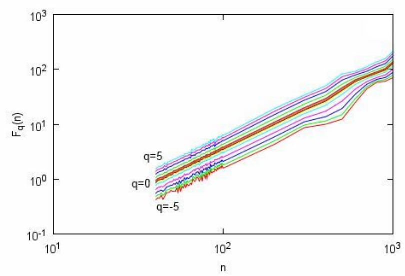
[Kaulakys et al., 2006]

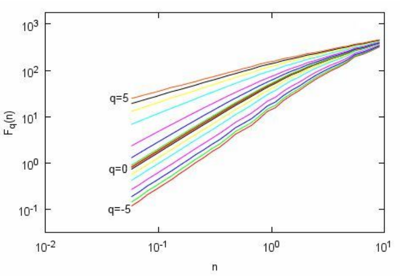
Multifraktalinis DFA metodas

Metodas identiškas Betendencių fliuktuacijų analizės (DFA) metodui, bet fliuktuacijų vidurkio išraiška yra sudėtingesnė:

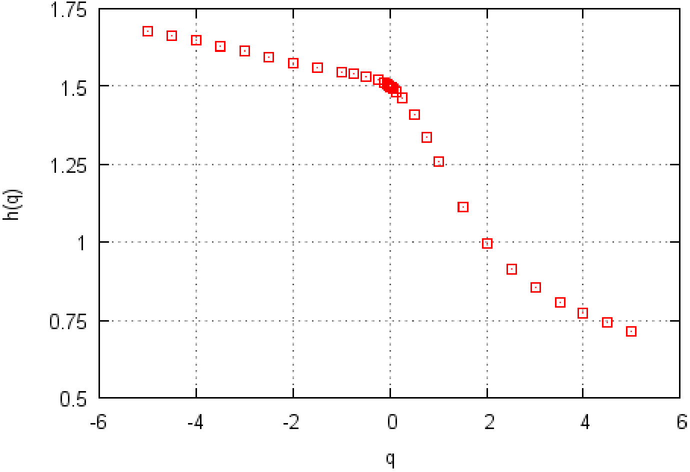
$$F_q(n) = \left\{ \left\langle \left[F_\nu^2(n) \right]^{\frac{q}{2}} \right\rangle \right\}^{\frac{1}{q}} \sim n^{h(q)}, \quad q \neq 0,$$

$$F_0(n) = \exp\left\{ 0.5 \left\langle \ln\left[F_\nu^2(n) \right] \right\rangle \right\} \sim n^{h(0)}.$$

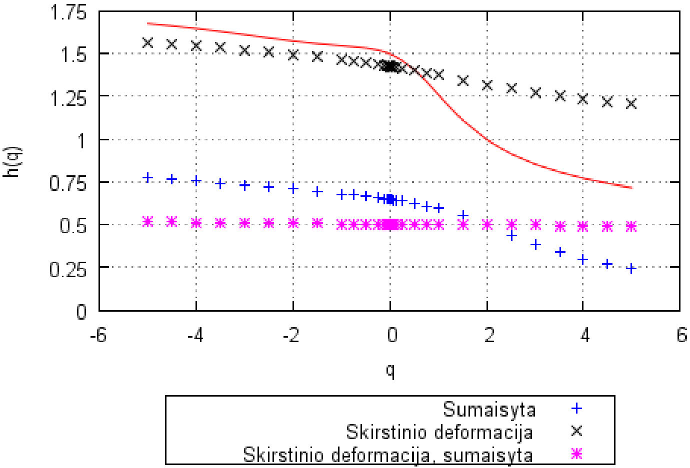




Multifraktalinis DFA metodas



Multifraktališkumo kilmė



Ačiū už dėmesį