

Bloku garuma izvēle apakšizlašu metodēs

Ieva Dasmane

Latvijas Universitāte

15.12.2011

- Ilglaicīgās atmiņas procesi, Hursta parametrs, veivletu konstrukcija
- Ticamības intervālu konstruēšana ar apakšizlašu metodi
- Optimālā bloka garums, simulācijas programmā R

Hursta parametrs H

- hidrologs Hursts (1951) - Nīlas upes problēma
- biofizikā, datorzinātnē, finansēs
- fraktāļu un haosa teorijā, ilglaicīgos procesos, spektrālajā analīzē

H ir procesu sev līdzības mērs un norāda uz ilglaicīgo atkarību
 $H \in (1/2, 1)$.

Sev līdzīgs process

Definīcija

Nepārtrauka laika stohastisks process $Y = \{Y(t) : 0 \leq t < \infty\}$ ir sev - līdzīgs process ar Hursta parametru $0 < H < 1$, ja procesiem $\{Y(at) : 0 \leq t < \infty\}$ un $\{a^H Y(t) : 0 \leq t < \infty\}$ ir vienādi galīgdimensionālie sadalījumi visiem $a > 0$.



Ilglaicīgās atmiņas process

Diskrēta laika stacionārs process ($X_j; j > 1$) ir ilglaicīgās atmiņas (long range dependent jeb (LRD)) process, ja tā korelācijas koeficienti

$$\rho(k) = \sigma^{-2} \{E(X_j X_{j+k}) - (EX_j)^2\}$$

ir formā

$$\rho(k) \sim c_r k^{-2(1-H)}, \text{ kad } k \rightarrow \infty \text{ un } 1/2 < H < 1,$$

kur k ir autokorelācijas lags, H ir Hursta parametrs un c_r ir absolūtā konstante.

Frakcionālā Brauna kustība

Tādu Gausa procesu $B_H = \{B_H(t) : 0 \leq t \leq \infty\}$ ar $0 < H < 1$ sauc par frakcionālo Brauna kustību (FBM), ja izpildās šādas īpašības:

- $B_H(t)$ ir stacionāri pieaugumi ;
- $B_H(0) = 0$ un $E(B_H(t)) = 0$ visiem $t \geq 0$;
- $E(B_H)^2(t) = t^{2H}$ visiem $t \geq 0$;
- $B_H(t)$ ir Gausa sadalījums visiem $t > 0$.

Kovariāciju funkcija

$$\gamma(s, t) = E(B_H(s)B_H(t)) = \frac{1}{2}\{t^{2H} + s^{2H} - (t - s)^{2H}\} \quad (1)$$

visiem $0 < s < t$.

Frakcionālais Gausa troksnis

Frakcionālās Brauna kustības pieaugumu procesu

$X = \{X_k : k = 0, 1, \dots\}$ sauc par frakcionālo Gausa troksni, ja

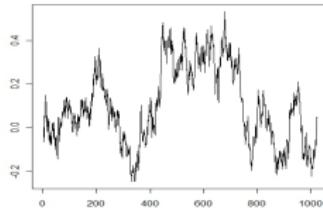
$$X_k = B_H(k+1) - B_H(k).$$

Kovariāciju funkcija

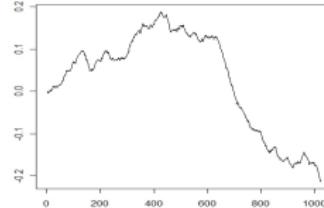
$$\gamma(k) = \frac{1}{2} [|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}] \quad (2)$$

visiem $k \in \mathbb{Z}$.

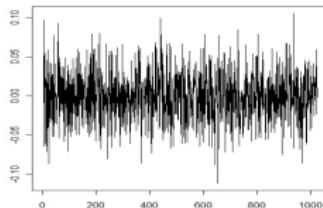
- Ja $H = 1/2$, tad $\gamma(k) = 0$;
- Ja $H < 1/2$, tad $\gamma(k) < 0$;
- Ja $H > 1/2$, tad $\gamma(k) > 0$.



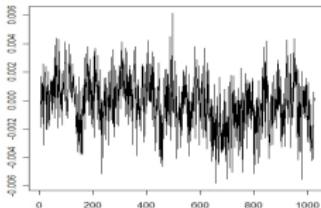
(a)



(b)



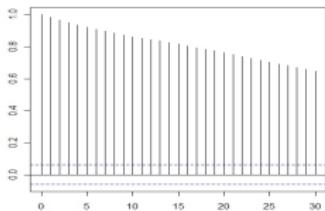
(c)



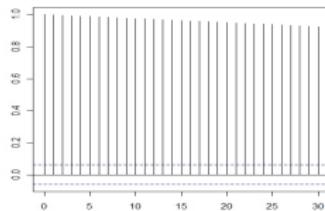
(d)

att.: Izlases apjoms $N = 1024$. (a) - standarta Brauna kustība ($H = 0.5$), (b) - frakcionālā Brauna kustība ($H = 0.9$), (c) un (d) - procesu pieaugumi.

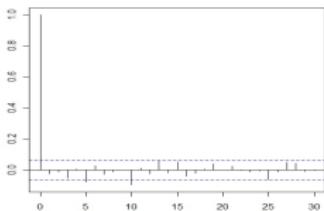
Autokovariāciju funkcijas



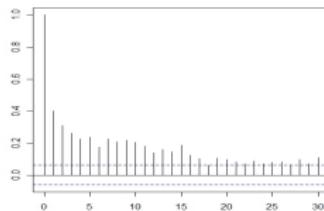
(a)



(b)



(c)



(d)

att.: (a) - standarta Brauna kustībai, (b) - frakcionālajai Brauna kustībai ($H = 0.9$), (c) un (d) - abu procesu pieaugumiem.

- J. Fourier frekvenču analīze (1807)
- A.Haar tēze (1909)
- būtība: tiek konstruēta funkcija, kas tiek mērogota jeb skalēta un translēta; tad tā tiek pielietota signāla procesa novērtēšanai
- pluss : mazāk jūtīga pret trokšņu ietekmi, jo tā mēra vidējās signāla svārstības dažādos līmenos
- pielieto : signālu un attēlu apstrādē, medicīnā, finansēs, ģeoloģijā un mūzikā

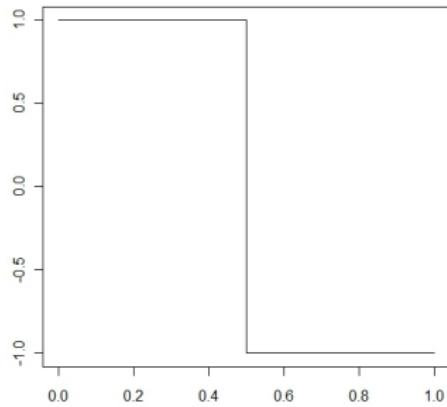


Veivleti ir funkciju saime, kuru lieto, lai aproksimētu citas funkcijas. Veidot i no tēva (skalas) ϕ un mātes ψ veivleta.

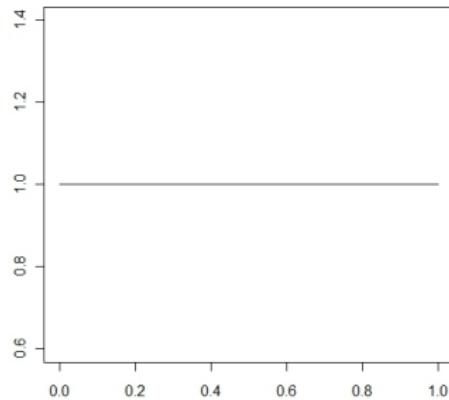
Haar tēva ϕ un mātes ψ veivlets

$$\phi(x) = \begin{cases} 1, & \text{ja } x \in [0, 1], \\ 0, & \text{citur.} \end{cases}$$

$$\psi(x) = \begin{cases} 1, & \text{ja } x \in [0, \frac{1}{2}), \\ -1, & \text{ja } x \in [\frac{1}{2}, 1), \\ 0, & \text{citur.} \end{cases}$$



(a)



(b)

att.: (a) - mātes veivlets ψ , (b) - tēva veivelts ϕ

Veivleti

Mērogošana un translēšana

$$\phi_{j,k}(t) = 2^{-j/2} \phi(2^{-j}t - k), \quad j, k \in \mathbb{Z}$$

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k), \quad j, k \in \mathbb{Z}$$

Skalas un veivletu koefīcienti

$$c_{j,k} = \int f(x) \phi_{j,k}(x) dx = \langle f, \phi_{j,k} \rangle$$

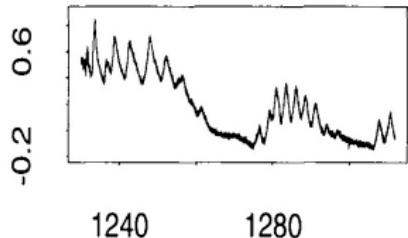
$$W_{j,k} = \int f(x) \psi_{j,k}(x) dx = \langle f, \psi_{j,k} \rangle$$

Funkcijas $f(x)$ tuvinājums $f_J(x)$

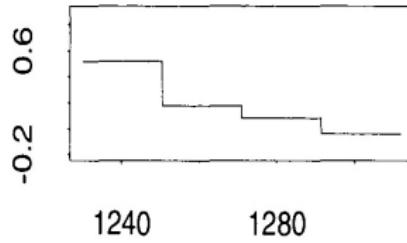
$$f_J(x) = \sum_{k=0}^{2^J-1} c_{j,k} \phi_{j,k}(x) + \sum_{k=0}^{2^J-1} W_{j,k} \psi_{j,k}(x). \quad (3)$$



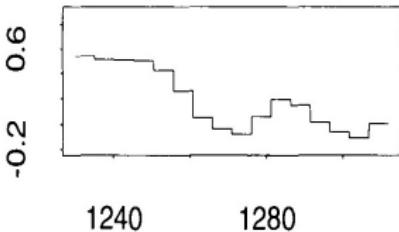
Piemērs



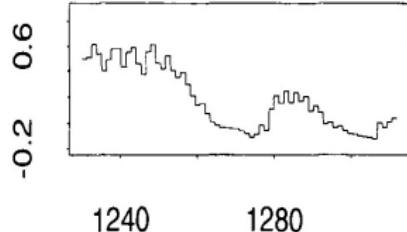
(a)



(b)



(c)



(d)

att.: (a) - "Plethysmography" dati apjomā $N = 4096$ un to Haar tēva veivletu tuvinājumi līmeņos $J = 2$ (b), $J = 4$ (c) un $J = 6$ (d). X- ass ir laiks sekundēs un Y - ass ir milivolti

Izlašu pārveidošanas metodes

- butstrapa metodes (Efron 1979)
- apakšizlašu metodes (Politis un Romano, 1992, 1994)

Piemēri statistikā

- (1) Chernozhukov un Fernandez-Val (2005). Apakšizlašu metodes ietekme uz kvantiļu regresiju procesiem
- (2) Gonzalo un Wolf (2005). Apakšizlašu metodes ietekme uz sliekšņotiem autoregresīviem modeļiem
- (3) Linton, Maasoumi un Whang (2005). Izstrādāja apakšizlašu testēšanas procedūru stohastiskajai dominancei
- (4) Hong un Scaillet (2006). Piedāvā ātro apakšizlases metodi nelineāriem dinamiskiem modeļiem
- (5) Lee un Pun (2006). Izpētīja apakšizlases metodi nestandarta M-novērtējumiem ar traucējošiem parametriem

Ticamības intervāli ar apakšizlasēm

- $\hat{\theta}_n$ - parametra θ novērtējums no izlases apjomā n
- definē

$$T_n = \sqrt{\tau_n}(\hat{\theta}_n - \theta),$$

kur τ_n ir normalizējošā konstante

- sadala izlasi N pārklājušos blokos $B_i = (X_i, \dots, X_{i+b-1})$ garumā b, kur $i = 1, \dots, N$ un $N = n - b + 1$
- definē T_n apakšizlašu kopiju, kas bāzēta uz blokiem B_i :

$$\hat{T}_{b,i} = \sqrt{\tau_b}(\hat{\theta}_{b,i} - \hat{\theta}_n)$$

Ticamības intervāli ar apakšizlasēm

- empīriskā sadalījuma funkcija apakšizlasēm

$$\hat{Q}_n(x) = \frac{1}{N} \sum_{i=1}^N I_{(\hat{T}_{b,i} \leq x)}, \quad x \in \mathbb{R}$$

Ticamības intervāls novērtētajam parametram ar apakšizlasēm

$$\left(\hat{\theta}_n + \frac{1}{\sqrt{\tau_n}} \hat{Q}_n^{-1}\left(\frac{\gamma}{2}\right), \hat{\theta}_n + \frac{1}{\sqrt{\tau_n}} \hat{Q}_n^{-1}\left(1 - \frac{\gamma}{2}\right) \right),$$

kur $0 < \gamma < 1$ un \hat{Q}_n^{-1} ir kvantile.

FGN parametra H novērtējums

tabula: Pārklājuma precizitātes ticamības intervāliem ar apakšizlašu metodi. Simulēts FGN process apjomā $n = 2^{11}$.

Bloka lielums: $l = 128 = 2^7$							
	$H = 0.6$		$H = 0.8$		$\alpha = 0.95$	$\alpha = 0.99$	
	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.95$	$\alpha = 0.99$			
Veivlets	0.937	0.969	0.927	0.965			
Haar	0.972	0.992	0.971	0.992			
Daubechies 2	1	1	1	1			
Daubechies 4							
Bloka lielums: $l = 256 = 2^8$							
	$H = 0.6$		$H = 0.8$		$\alpha = 0.95$	$\alpha = 0.99$	
	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.95$	$\alpha = 0.99$			
Veivlets	0.925	0.963	0.925	0.958			
Haar	0.948	0.976	0.951	0.977			
Daubechies 2	0.992	0.999	0.996	0.999			
Daubechies 4							

FGN parametra H novērtējums

tabula: Simulēts FGN process apjomā $n = 2^{11}$.

Bloka lielums: $l = 512 = 2^9$				
Veivlets	$H = 0.6$		$H = 0.8$	
	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.95$	$\alpha = 0.99$
Haar	0.937	0.970	0.922	0.964
Daubechies 2	0.917	0.972	0.944	0.976
Daubechies 4	0.971	0.991	0.979	0.992

tabula: Simulēts FGN process apjomā $n = 2^{15}$, $H = 0.6$, Haar veivlets, $\alpha = 0.95$

Bloka lielums	P.p.	Vid.vērt.	Rangs
$l = 2^9$	0.945	0.044	0.009
$l = 2^{10}$	0.955	0.041	0.013
$l = 2^{11}$	0.964	0.039	0.019

- Bloka lieluma izvēle ietekmē apakšizlases metodes precizitāti

Metodes bloka lieluma izvēlei atkarībā no izlases lieluma

- (1) Kalibrēšanas metode, kuru ieviesa Loh (1987);
- (2) Ticamības intervālu nepastāvības minimizēšanas metode.
Balstās uz tāda bloka lieluma izvēli, kas minimizē apkašizlašu ticamības intervālu galapunktu empirisko variāciju.

Algoritms

1. No dotajiem datiem X_1, \dots, X_n generējam K apakšizlases $X_1^{*k}, \dots, X_n^{*k}$, saskaņā ar butstrapa sadalījumu P_n^* (neatkarīgiem datiem var izmantot Efron butstrapa tehniku, atkarīgiem - slīdošo bloku butstrapu), kur $k = 1, \dots, K$.

Aprēķinām $1 - \lambda$ ticamības intervālu CI_l^k , dažādiem izvēlētiem bloka lielumiem l .

2. Katram l aprēķina $\hat{g}(l) = \#\{\hat{\theta}_n \in CI_l^k\}/K$.

Saprātīgs bloka lielums tad apmierinās $\hat{g}(l) \approx 1 - \lambda$.

Ja jau šajā solī bloka lielums l ar $\hat{g}(l) = 1 - \lambda$ ir atrasts, tad vairāk kalibrēšana nav jāpielieto. Tad, izmantojot iegūto l var konstruēt ticamības intervālus.

Algoritms

1. Interesējošā parametra θ novērtējumam, aprēķinām ticamības intervāus pie vēlamā ticamība līmeņa tādiem bloka lielumiem l , kas mainās no $l = l_{\text{mazs}}$ līdz $l = l_{\text{liels}}$, un iegūstam ticamības intervālu galapunktus $I_{l,\text{mazs}}$ un $I_{l,\text{liels}}$. Bloka lielumus izvēlas $l_{\text{mazs}} = c_1 n^\eta$ un $l_{\text{liels}} = c_2 n^\eta$ tādiem $0 < c_1 < c_2$ un $0 < \eta < 1$. Romano un Wolf (2001) iesaka izvēlēties $c_1 \in [0.5, 1]$, $c_2 \in [2, 3]$ un $\eta = 0.5$.
2. Mazam veselam k aprēķina VI_l , kas ir standartnovirze galapunktu kopai $\{I_{l-k,\text{apak}}, \dots, I_{l+k,\text{apak}}\}$ plus standartnovirze $\{I_{l-k,\text{aug}}, \dots, I_{l+k,\text{aug}}\}$ (Romano un Wolf iesaka izvēlēties $k = 2$ vai $k = 3$).
3. Izvēlas l^* , kurš atbilst vismazākajam indeksam VI_l un iegūst optimālo ticamības intervālu $[I_{l^*,\text{apak}}, I_{l^*,\text{aug}}]$.

AR(1) procesa vidējās vērtības novērtējums

tabula: Simulēts AR(1) process apjomā $n = 250$, novērtētais parametrs - vidējā vērtība, l -bloka lielums, ρ - AR(1) procesa parametrs . 1000 simulācijas. Attēlotas ticamības intervālu pārklājuma precizitātes. Ticamības līmenis $\alpha = 0.95$.

ρ	bloka lielums				$k = 2$		$k = 3$	
	$l = 4$	$l = 8$	$l = 16$	$l = 32$		1		1
0.2	0.931	0.932	0.924	0.886	0.924	16	0.920	17
0.5	0.879	0.899	0.905	0.868	0.876	31	0.876	31
0.8	0.709	0.812	0.871	0.836	0.872	19	0.871	20
0.95	0.421	0.547	0.639	0.690	0.672	21	0.666	20
-0.5	0.963	0.951	0.925	0.893	0.947	7	0.928	17

Paldies par uzmanību!