

LATVIJAS UNIVERSITĀTE
FIZIKAS UN MATEMĀTIKAS FAKULTĀTE
MATEMĀTIKAS NODAĻA

**NEPARAMETRISKIE MAINAS PUNKTA TESTI ILGLAICĪGI
ATKARĪGIEM NOVĒROJUMIEM**

MAĢISTRA DARBS

Autors: **Ieva Dasmane**

Stud. apl. id06030

Darba vadītājs: doc. Dr.math. Jānis Valeinis

RĪGA 2013

Anotācija

Šī darba mērķis ir maiņas punktu analīze ilglaicīgi atkarīgiem procesiem. Tiks apskatīti Vilkoksona un vidējo vērtību starpības divu izlašu testi, kas nesen ieviesti ilglaicīgās atmiņas procesiem (skatīt Taqqu, Dehling un Roach [1]). Ar simulāciju palīdzību tiks salīdzināta testu empīriskā jauda, kā arī analizēta statistiku asimptotiskā uzvedība pie nosacījuma, ka izpildās nulles hipotēze.

Tiks analizēta empīriskās ticamības funkcijas metodes pielietojamība maiņas punktu noteikšanā ilglaicīgi atkarīgiem procesiem un salīdzināta ar divu izlašu testiem no [1], kā arī ar standarta *CUSUM* algoritma procedūru.

Atslēgas vārdi: ilglaicīgi atkarīgs process, Hursta parametrs, maiņas punkta testi, CUSUM, EL

Abstract

The aim of this work is change point analysis for the long range dependent processes. The Wilcoxon and the difference-of-means two sample tests will be discussed, based on recent results for long range dependent processes (see Taqqu, Dehling and Roach [1]). In a simulation study the power and the confidence level of both tests will be compared, when the null hypothesis holds.

The applicability of the empirical likelihood method for the change point detection in long memory processes will be analyzed and compared with the two sample tests from [1] as well as with the standart *CUSUM* procedure.

Keywords: long memory process, Hurst parameter, change point tests, CUSUM, EL

Saturs

Apzīmejumi	3
Ievads	4
1 Ilglaicīgās atmiņas procesi, Hursta parametrs un tā novērtēšanas metodes	6
1.1 Sev - līdzīgi procesi un ilglaicīgā atkarība	6
1.2 Frakcionālā Brauna kustība un frakcionālais Gausa troksnis	8
1.3 Hursta parametra novērtēšanas metodes	11
1.3.1 Kopējās variācijas metode	11
1.3.2 Periodogrammu metode	11
1.3.3 R/S metode	12
1.3.4 Veivletu metode	13
1.3.5 Ticamības intervāli parametram H ar apakšizlasēm	13
2 Maiņas punkta analīze	16
2.1 Maiņas punkta analīze	16
2.1.1 $CUSUM$ algoritms	17
2.1.2 Maiņas punkta analīze stacionāru stohastisku procesu gadījumā	18
3 Asimptotiskā $\bar{X} - \bar{Y}$ uzvedība	19
3.1 Asimptotiskā ekvivalence un lēni variējošas funkcijas	20
3.2 Vienas sadalītas izlases gadījums	20
3.2.1 Asimptotiskā teorija	20
3.2.2 Simulācijas	21
4 Vilkoksona un vidējo vērtību starpības testi	25
4.1 Vilkoksona testa asimptotiskais sadalījums	26
4.2 Vilkoksona rangu tests	28
4.2.1 Pielietojums	30
4.3 Vidējo vērtību starpību tests un tā robežsadalījums	31
5 Empīriskā ticamības funkcija ilglaicīgi atkarīga procesa vidējai vērtībai	33
5.1 Empīriskā ticamība divu izlašu gadījumā	33
5.2 Blokveida empīriskās ticamības konstrukcija vidējai vērtībai	35
5.2.1 Kitamura piedāvātā blokveida EL konstrukcija	36
5.2.2 Pieņēmumi	36
5.2.3 Blokveida EL attiecības robežsadalījums LRD novērojumiem	37

6	Praktiskā daļa	39
6.1	Parametra H novērtējumi programmā R	39
6.2	Maiņas punkta testi ilglaicīgi atkarīgiem procesiem	42
6.2.1	Normāli sadalīti dati	45
6.2.2	Smago astu dati	48
6.3	EL ilglaicīgi atkarīgiem procesiem	51
6.3.1	EL statistikas robežsadalījuma analīze	51
6.3.2	EL metodes pārklājuma precizitāte	53
6.3.3	Maiņas punkta noteikšana	53
	Nobeigums	61
	Pielikums	63
	Izmantotā literatūra un avoti	66
A	Pielikums	68
A.1	Vilkoksona un vidējo vērtību starpības testu R kods	68
A.2	p – vērtību grafiki maiņas punkta noteikšanai	73

Apzīmejumi

FBM - frakcionālā Brauna kustība

fGn - frakcionālais Gausa troksnis

iid - neatkarīgi un vienādi sadalīti dati

SRD - īslaicīgi atkarīgi procesi

LRD - ilglaicīgi atkarīgi procesi

R/S - R/S testa statistika

EL - empīriskās ticamības metode

W_n - Vilksona testa statistika

D_n - vidējo vērtību starpības testa statistika

$CUSUM$ - kumulatīvo summu metode

$\rightarrow^d$ - konverģence pēc sadalījuma

$\rightarrow^p$ - konverģence pēc varbūtības

$\rightarrow^{g.d.}$ - gandrīz droša konverģence

Ievads

Ilglaicīgās atmiņas procesi (*long range dependent* jeb *LRD*) ir mūsdienās plaši analizēti procesi (skatīt, piemēram, [2]). Pēdējā desmitgadē *LRD* procesi plaši pielietoti telekomunikāciju zinātnē. Ilglaicīgās atmiņas procesu svarīgākā raksturojošā pazīme ir tā, ka autokovariāciju funkcija dilst ļoti lēni. Ilglaicīgās atmiņas procesu apraksta Hursta parametrs H (nosaukts par godu hidrologam Hurstam, kurš 1951. gadā atklāja šo parametru, pētot Nīlas upes problēmu [3]), kur $H \in (1/2, 1)$ liecina, ka procesam pastāv ilglaicīgā atmiņa. Mūsdienās Hursta parametrs parādās daudzās nozarēs sākot no biofizikas līdz pat datorzinātnei. Tomēr jaunākās Hursta parametra novērtēšanas metodes nāk no fraktālās matemātikas.

Maiņas punkta analīze ir statistiska metode, ar kuras palīdzību iespējams noteikt, vai dotā laikrinda ir nemainīga laikā [4]. Šī metode sāka rasties 20.gs vidū [5] kvalitātes kontroles kontekstā. Maiņas punkta analīzi var iedalīt divās apakšnozarēs. Pirmā ir hipotēžu pārbaude un otrā atbild par maiņas punkta atrašanās vietas noteikšanu laikrindā. Ar izmaiņu rašanos kādā procesā mēs varam saskarties ļoti daudzās nozarēs, piemēram, ekonomikā, medicīnā, ģeoloģijā, fizikā utt. Līdz ar to maiņas punktu noteikšana un izpēte mūsdienās ir aktuāla.

Neatkarīgu novērojumu gadījumā maiņas punkta problēmas ir plaši pētītas, kā arī ieviesti gan parametriskie, gan neparametriskie testi (skatīt, piemēram, Csorgo un Horvath [6]). Ir arī vairāki rezultāti vāji atkarīgiem novērojumiem, piemēram, Ling [7], Aue, Hormann, Horvath un Reimherr [8] un Wied, Kramer un Dehling [9].

Gadījumā, kad datiem ir ilglaicīgā atkarība, ir atrodamas daudz mazāk zinātnisku pētījumu. Csorgo un Horvath pētīja tā saucamos CUSUM testus vidējās vērtības izmaiņām. 2002. gadā Kramer, Sibbertsen un Kleiber [10] novēroja, ka, ja datiem ir ilglaicīgās atkarības struktūra, tad tas var likt nepatiesi noraidīt nulles hipotēzi par to, ka izmaiņu nav.

Antoch, Huškova, Janic un Ledwina [11] pētīja rangu testus neatkarīgiem novērojumiem maiņas punkta analīzē. Šajā darbā viens no mērķiem ir izpētīt rangu testus vidējās vērtības izmaiņām ilglaicīgi atkarīgiem procesiem.

Ar empīriskās ticamības funkciju (turpmāk tiks lietots apzīmējums EL), ar kuru iepazīstināja Owen [12] un [13], var veikt statistisko datu analīzi, nepieņemot nekādus nosacījumus par datu veidu. Tā ir neparametriska metode, kas ir plaši analizēta neatkarīgiem datiem. Šajā darbā tiks aplūkota un analizēta EL statistikas uzvedība ilglaicīgās atmiņas procesiem ar pielietojumu

maiņas punkta noteikšanā.

Šī darba mērķi ir:

- analizēt Vilkoksona un vidējo vērtību starpību testu statistiku asimptotisko uzvedību ilglaicīgās atmiņas procesiem
- pielietot divu izlašu testus maiņas punktu noteikšanā, konstruējot p –vērtību līknes
- EL metodes analīze divu izlašu gadījumā atkarīgiem novērojumiem ar pielietojumu maiņas punktu noteikšanā

Pirmajā nodaļā definēti ilglaicīgās atmiņas procesi un metodes atmiņas parametra novērtēšanai. Otrajā - apskatīta maiņas punkta analīze. Trešajā nodaļā aprakstīta asimptotiskā statistikas $\bar{X} - \bar{Y}$ uzvedība, pamatojoties uz kuru, aplūkoti neparametriskie maiņas punkta testi ceturtajā nodaļā. Piektajā nodaļā aprakstīta EL metode ilglaicīgi atkarīgu datu gadījumā. Sestajā nodaļā veikts metožu praktiskais pielietojums.

1 Ilglaicīgās atmiņas procesi, Hursta parametrs un tā novērtēšanas metodes

1.1 Sev - līdzīgi procesi un ilglaicīgā atkarība

No sākuma definēsim sev līdzīgus (*self-similar*) un ilglaicīgi atkarīgus (*long range dependent*) procesus. Vēlāk apskatīsim, kā šie jēdzieni ir saistīti ar frakcionālo Brauna kustību un frakcionālo Gausa troksni. Šim mērķim izmantosim literatūru [14], [15] un [16].

Definīcija 1. $X = \{X_k : k = 0, 1, 2, \dots\}$ ir stacionārs diskrēta laika stohastisks process, ja vektoriem $(X_{k_1}, \dots, X_{k_d})$ un $(X_{k_1+n}, \dots, X_{k_d+n})$ ir tāds pats kopējais sadalījums visiem veseliem $d, n \geq 1$ un $k_1, \dots, k_d \geq 0$. Gausa procesiem tas nozīmē to pašu, kā prasīt, lai autokovariāciju funkcija $\gamma(k)$ nav atkarīga no n .

Procesam X_k izpildās īpašība, ka

$$X_{km} + \dots + X_{(k+1)m-1} = a_m X_k. \quad (1.11)$$

visiem $m \geq 1, k \geq 0$ un a_m ir reizinātājs.

Definīcija 2. Process $X^{(m)} = \{X_k^{(m)} : k = 0, 1, 2, \dots\}$ visiem $m \geq 1$, kur

$$X_k^{(m)} = \frac{1}{m}(X_{km} + \dots + X_{(k+1)m-1}).$$

Definīcija 3. [16] Diskrēta laika stohastisku procesu X sauc par *sev - līdzīgu* ar Hursta parametru $0 < H < 1$, ja procesiem X un $m^{1-H} X^{(m)}$ ir vienādi galīgdimensionālie sadalījumi visiem $m \geq 1$. Tas nozīmē, ka visiem $d \geq 1$ un $0 \leq k_1 < \dots < k_d$ vektoram

$$(X_{k_1}, \dots, X_{k_d})$$

ir tas pats sadalījums, kas vektoram

$$(m^{1-H}X_{k_1}^{(m)}, \dots, m^{1-H}X_{k_d}^{(m)}).$$

Tad no (1.11) $a_m = m^H$, ja X ir sev - līdzīgs ar Hursta parametru H .

Definīcija 4. 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 $a > 0$.

Tagad iepazīsimies ar *ilglaicīgo atkarību*. Intuitīvi diskrēta laika stohastisks process X ir ilglaicīgi atkarīgs (ar ilglaicīgo atmiņu), kad tā autokovariāciju funkcija γ dilst tik lēni, ka $\sum_{k=0}^{\infty} \gamma(k) = \infty$.

Definīcija 5. [15] Diskrēta laika stacionārs process $(X_j : j > 1)$ ar dispersiju $\sigma^2 = E(X_j^2) - (EX_j)^2$ ir *ilglaicīgi atkarīgs process (ilglaicīgās atmiņas jeb (LRD))*, ja tā korelācijas koeficienti

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

ir formā

$$\rho(k) \sim c_r k^{-(1-\alpha)}, \text{ kad } k \rightarrow \infty \text{ un } 0 < \alpha < 1, \quad (1.12)$$

kur k ir autokorelācijas lags, α ir *ilglaicīgās atmiņas parametrs* un c_r ir absolūtā konstante (skatīt Beran [17] un Taqqu [2]).

Apgalvojums 1. Pieņemsim, ka procesam $(Z(t) : t \geq 0)$ ir stacionāri pieaugumi un tas ir sev - līdzīgs ar Hursta parametru H . Tad, ja $1/2 < H < 1$, tad procesa pieaugumi $(X_j = Z_j - Z_{j-1} : j > 1)$ veido LRD procesu ar autokorelācijas koeficientiem

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

Bet, ja $H = 1/2$,

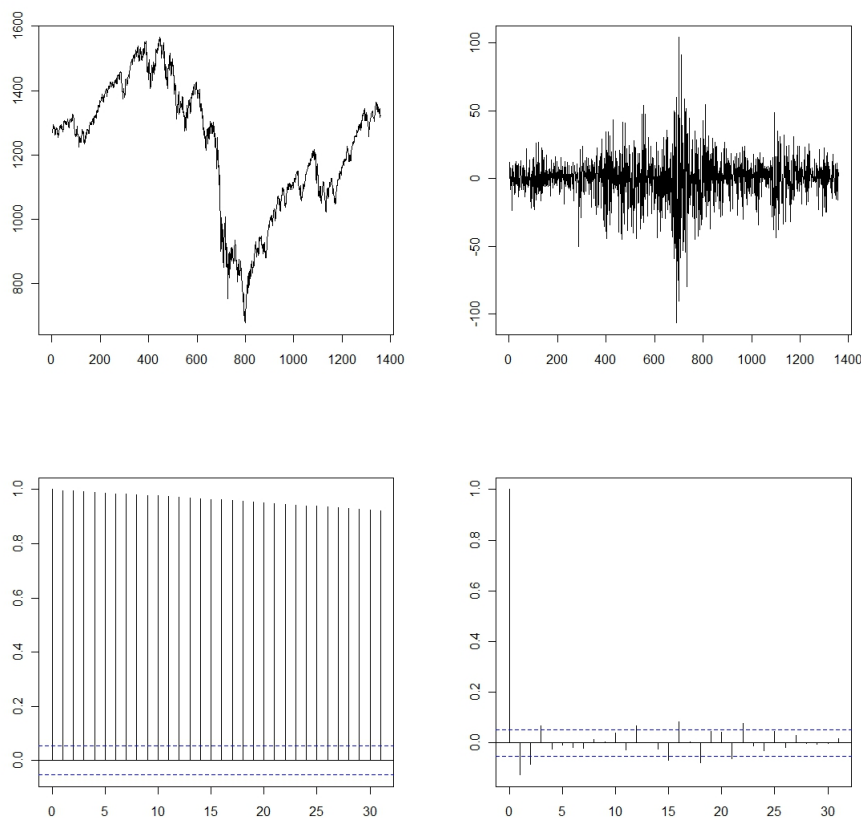
$$\rho(k) = 0, \quad \forall k \geq 0,$$

un tāpēc process ir *īsilaicīgi atkarīgs*.

Kad $1/2 < H < 1$, tad Hursta parametru H un ilglaicīgās atmiņas parametru α saista vienādība

$$H = \frac{1 + \alpha}{2}. \quad (1.13)$$

Piemērs 1. Aplūkosim procesu $SP500$ (datus var iegūt mājas lapā [18]). $SP500$ ir populārs 500 lielāko uzņēmumu aktīvu tirgus cenas indekss Amerikas Savienotajās valstīs kopš 1957. gada. Dati iegūti par laika periodu 03.01.2006 – 25.05.2011.



Att. 1.1: Dati laika intervālā: 03.01.2006 – 25.05.2011. Augšā pa kreisi - $SP500$ process, augšā pa labi - $SP500$ procesa pieaugumi, apakšā attiecīgās autokovariāciju funkcijas

Kā var redzēt 1.1 attēlā, $SP500$ process ir nestacionārs. Ievērojam, ka procesam $SP500$ autokovariāciju funkcija dilst ļoti lēni, kas liecina par ilglaicīgo atmiņu.

1.2 Frakcionālā Brauna kustība un frakcionālais Gausa trokšnis

Definīcija 6. [14] 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 $t > 0$.

H ir frakcionālās Brauna kustības *Hursta parametrs*. Frakcionālās Brauna kustības *kovariāciju funkcija* ir

$$\gamma(s, t) = E(B_H(s)B_H(t)) = \frac{1}{2}\{t^{2H} + s^{2H} - (t - s)^{2H}\} \quad (1.21)$$

visiem $0 < s < t$ (skatīt [14]). Tā kā Gausa procesiem vidējā vērtība un kovariāciju struktūra nosaka galīgdimensionālo sadalījumu, tāpēc, no (1.21) mēs secinām, ka procesam $\{B_H(at) : 0 < t < \infty\}$ un $\{a^H B_H(t) : 0 < t < \infty\}$ ir vienādi galīgdimensionālie sadalījumi. Tātad frakcionālā Brauna kustība ir *sev līdzīga* ar Hursta parametru H . Frakcionālā Brauna kustība ir vienīgais Gausa process ar stacionāriem pieaugumiem, kurš ir sev līdzīgs (skatīt [14]).

Definīcija 7. Pieaugumu procesu $X = \{X_k : k = 0, 1, \dots\}$ sauc par *frakcionālo Gausa troksni*, ja

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

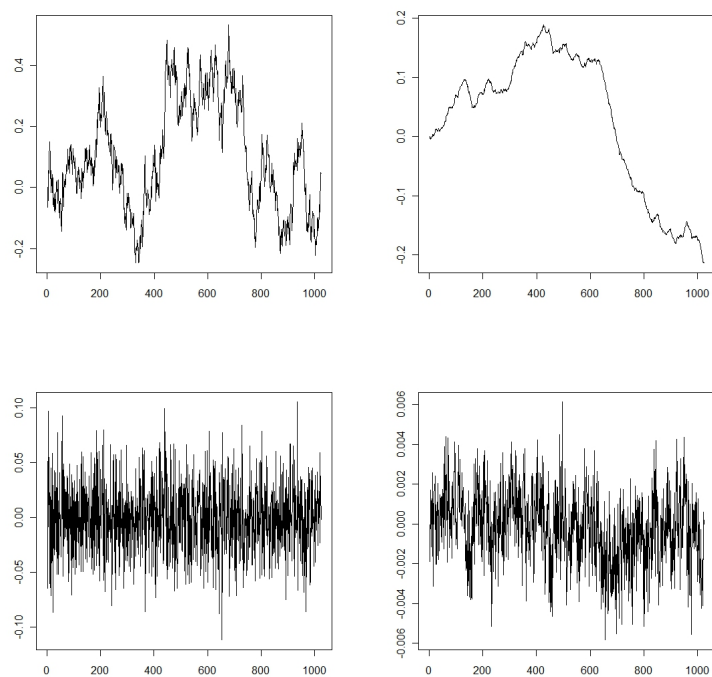
Procesa X_k *autokovariāciju funkcija* $\gamma(k)$ ir

$$\gamma(k) = \frac{1}{2}[|k - 1|^{2H} - 2|k|^{2H} + |k + 1|^{2H}] \quad (1.22)$$

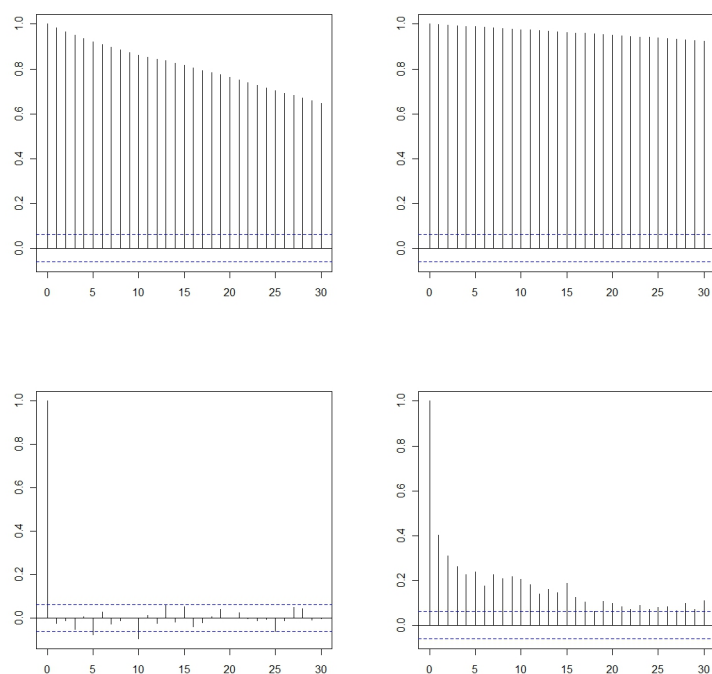
visiem $k \in \mathbb{Z}$.

- Ja $H = 1/2$, tad $\gamma(k) = 0$ (izņemot gadījumu, kad $k = 0$). Tā ir standarta Brauna kustība, kuras pieaugumu ir neatkarīgi;
- Ja $H > 1/2$, tad $\gamma(k) > 0$

1.2 attēlā ir standarta Brauna kustības un frakcionālās Brauna kustības piemēri. Var redzēt, ka pieaugumu process abiem ir stacionārs. 1.3 attēlā varam redzēt, ka standarta Brauna kustības autokovariācijas dilst ievērojami straujāk, nekā frakcionālās Brauna kustības ($H = 0.9$) autokovariācijas. Aplūkojot pieaugumu autokovariācijas, var redzēt, ka, kad $H = 1/2$, $\gamma(k) = 0$, bet, kad $H = 0.9$, autokovariācijas ir nozīmīgas.



Att. 1.2: Izlases apjoms $N = 1024$. Augšā pa kreisi - standarta Brauna kustība ($H = 0.5$), augšā pa labi - frakcionālā Brauna kustība ($H = 0.9$), apakšā abu procesu pieaugumi



Att. 1.3: Autokovariāciju funkcijas: augšā pa labi - standarta Brauna kustībai, augšā pa kreisi - frakcionālajai Brauna kustībai ($H = 0.9$), apakšā abu procesu pieaugumiem

1.3 Hursta parametra novērtēšanas metodes

Apskatīsim četras Hursta parametra novērtēšanas metodes programmā **R**, kuras piedāvātas publikācijā [14]. Lai atrastu metodi, kura Hursta parametru novērtē visprecīzāk, šīs metodes sākotnēji tiks pielietotas simulētiem datiem - frakcionālajam Gausa troksnim ar $H = 0.6$ un $H = 0.8$ apjomā $n = 2^9$ (šāds izlases apjoms tiek izvēlēts, jo veivletu metode, kuru apskatīsim, pieprasa datu apjomu divnieka pakāpē). Metodes, kas dos vislabākos rezultātus, pielietosim reāliem datu piemēriem, lai noteiktu, vai dati ir ilglaicīgi atkarīgs process ($H \geq 0.5$), lai šos procesus darba turpinājumā varētu pielietot maiņas punkta noteikšanai.

1.3.1 Kopējās variācijas metode

Kopējās variācijas metode ir balstīta uz sev - līdzības īpašību. Kopējais process $X^{(m)}$ izskatās šādi:

$$X_k^{(m)} = \frac{1}{m}(X_{km} + \dots + X_{(k+1)m-1}),$$

kur $k = 0, 1, \dots$ un m apzīmē to, cik elementu ietilpst summā $X_k^{(m)}$. Sev līdzības dēļ procesam $X^{(m)}$ ir tas pats galīgdimensionālais sadalījums, kas procesam $m^{H-1}X$ lieliem m . Turklāt $Var(X_k^{(m)}) = m^{2H-2}Var(X_k)$. Procesa $X_k^{(m)}$ dispersija ir vienāda visiem k un tās novērtējums ir

$$\widehat{Var}(X_k^{(m)}) = \frac{1}{M} \sum_{i=0}^{M-1} (X_i^{(m)} - \bar{X}^{(m)})^2, \quad (1.31)$$

kur $\bar{X}^{(m)}$ ir:

$$\bar{X}^{(m)} = \frac{1}{M} \sum_{i=0}^{M-1} X_i^{(m)},$$

un M ir veselā daļa no n/m (n ir izlases apjoms).

Apgalvojums 2. [14] Parametra H novērtējums tiek iegūts attēlojot $\widehat{Var}(X_k^{(m)})$ pret m log-log skalā. Kad variācijas novērtējumi sakrīt ar īstajām vērtībām, tad visi punkti izvietojas uz taisnas līnijas ar slīpumu $2H - 2$.

1.3.2 Periodogrammu metode

Periodogrammu metode novērtē Hursta parametru, pielāgojot taisnu līniju spektrālajā domeinā. Tā balstās uz novērojumu, ka Gausa trokšņa spektrālā blīvuma funkcija uzvedas kā $c_f |\lambda|^{1-2H}$, kad $\lambda \rightarrow 0$ (6.37)

Definīcija 8. Spektrālo blīvumu var novērtēt ar *periodogrammu*

$$I(\lambda) = \sum_{j=-(N-1)}^{N-1} \hat{\gamma}(j) e^{ij\lambda},$$

kur $\hat{\gamma}(j)$ ir izlases autokovariāciju funkcija.

Periodogrammas definīcija ir ekvivalenta ar

$$I(\lambda) = \frac{1}{N} \left| \sum_{k=0}^{N-1} (X_k - \bar{X}) e^{ij\lambda} \right|^2,$$

(skatīt [14]). Periodogramma un spektrālais blīvums - abi ir simetriski nulles reģionā. Tāpēc var parādīt, ka periodogramma ir asimptotiski nenovirzīts novērtējums spektrālajam blīvumam, tas ir,

$$\lim_{N \rightarrow \infty} E[I(\lambda)] = f(\lambda).$$

Apgalvojums 3. [14] Parametra H novērtējums tiek iegūts, grafiski attēlojot $I(\lambda_k)$ pret izlases frekvencēm $\lambda(k)$ log-log skalā un pielāgojot taisnu līniju caur datiem, kurai teorētiskais slīpums log-log grafikā ir $1 - 2H$.

1.3.3 R/S metode

1951. gadā hidrologs Hursts aprakstīja R/S statistiku, vadoties pēc *Nīlas upes* problēmas (skatīt [17]). Tiek pieņemts, ka apskatāmais laika intervāls ir diskrets un nav nekādu ūdenstilpnes krātuves zudumu (iztvaikošana, noplūde utt.). Mērķis ir sasniegt, ka ūdens izplūde ir vienmērīga un ka laikā $t + k$ ūdens krātuve ir tik pat pilna kā laikā t , un ka ūdenstilpne nekad nepārplūst.

Ar X_i apzīmēsim ūdens ieplūdi laikā i , un $Y_j = \sum_{i=1}^j X_i$ ir kopējā ūdens ieplūde līdz laika brīdim j . Tad ideālais ūdens tilpums tiek aprakstīts kā

$$R(t, k) = \max_{0 \leq i \leq k} \left[Y_{t+i} - Y_t - \frac{i}{k} (Y_{t+k} - Y_t) \right] - \min_{0 \leq i \leq k} \left[Y_{t+i} - Y_t - \frac{i}{k} (Y_{t+k} - Y_t) \right].$$

Standartizēts $R(t, k)$ ir $S(t, k)$, kas izskatās šādi:

$$S(t, k) = \sqrt{k^{-1} \sum_{i=t+1}^{t+k} (X_i - \bar{X}_{t,k})^2},$$

kur $\bar{X}_{t,k} = k^{-1} \sum_{i=t+1}^{t+k} X_i$. Attiecību

$$R/S = \frac{R(t, k)}{S(t, k)} \quad (1.32)$$

sauc par R/S statistiku. Hursts grafiski attēloja $\log(R/S)$ pret dažādām k vērtībām un novēroja, ka lieliem k , $\log(R/S)$ atrodas apkārtņē taisnai līnijai ar slīpumu $\frac{1}{2}$. Hursts novēroja, ka Nīlas upes datiem un citiem hidroloģiskiem, ģeofizikālīgiem un klimatoloģiskiem datiem, R/S uzvedas kā konstante reiz k^H visiem $H > 1/2$. To sauc par *Hursta efektu*. Lieliem k ,

$$\log E[R/S] \approx a + H \log k, \text{ kur } H > \frac{1}{2}.$$

Teorēma 4. [17] Pieņem, ka X_t ir tāds, ka X_t^2 ir ergodisks un $t^{-H} \sum_{s=1}^t X_s$ konverģē vājā nozīmē uz frakcionālo Brauna kustību, kad $t \rightarrow \infty$. Tad, visiem $k \rightarrow \infty$,

$$k^{-H} R/S \rightarrow_d \xi,$$

kur ξ ir nedeģenerēts gadījuma lielums.

Tas nozīmē, ka grafika $\log R/S$ pret $\log k$ punkti būs vienmērīgi izkaisīti ap taisni ar slīpumu H . Ilglaicīgās atmiņas gadījumā ap $H > 1/2$, pretējā gadījumā ap $H = 1/2$.

1.3.4 Veivletu metode

Veivletu novērtējums ir cieši saistīts ar periodogrammas novērtējumu. Dati tiek transformēti un iegūts novērtējums no transformētajiem datiem (periodogrammas metodē novērtējums ir veikts frekvenču domēnā). Tiek aprēķināti veivletu koeficienti (skatīt pielikumā 6.38).

Pielietojot veivletu transformāciju izlasei, varam iegūt veivletu koeficientu dispersiju, ko aprēķina

$$v_j = 2^j \sum_{k=0}^{2^{-j}-1} (W_x(j, k))^2. \quad (1.33)$$

Veivletu novērtējums parametram H ir asimptotiski nenovirzīts (praksē pat mazām datu izlasēm ir ļoti zemas novirzes).

Apgalvojums 5. [14] Veivletu novērtējums Hursta parametram tiek atrasts, attēlojot veivletu koeficientu dispersijas logaritmu $\log(v_j)$ pret skalas līmeni j . Iegūst j punktus, kuriem tiek piemērota taisne pēc mazākās kvadrātu metodes. Taisnes slīpums ir $2H - 1$.

1.3.5 Ticamības intervāli parametram H ar apakšizlasēm

Konstruēsim ticamības intervālus visām četrām iepriekš apskatītajām Hursta parametra novērtēšanas metodēm, lai noteiktu, kura metode ir vislabākā. Šim mērķim izmantosim publikācijā [15] piedāvāto apakšizlasu metodi.

Hursta parametra novērtējumu $\hat{H}_n$ centrēto versiju, kas balstīta uz ilglaicīga Gausa stacionāra procesa $(X_j; j \geq 1)$ n novērojumiem, apzīmēsim ar

$$T_n = \sqrt{\nu_{j_1(n)}(n)}(\hat{H}_n - H),$$

kur $\nu_{j_1(n)}(n)$ apzīmē veivletu koeficientu skaitu līmenī $j_1 = j_1(n)$. Ar

$$Q_n(x) = P(T_n \leq x) \quad (1.34)$$

apzīmēsim sadalījuma funkciju. Tiek pieņemts, ka

$$T_n \rightarrow^d N(0, \sigma^2(H)).$$

T_n sadalījums $Q_n(x)$ galīgiem n nav zināms. Tāpēc novērtēsim to ar sadalījumu $\hat{Q}_n$.

Pieņemsim, ka $B_i = (X_i, \dots, X_{i+l-1})$, $i = 1, \dots, N$ ir $N = n - l - 1$ pārklājoši bloki garumā l , kur $l = l_n$ ($1 \leq l \leq n$). Pieņemsim, ka $\hat{H}_{l,i}$ ir parametra H novērtējums, izmantojot datus blokā B_i . Novērojam, ka bloki B_i pārklājas un tiem ir blakus stāvoši sākuma punkti.

Apakšizlases kopija T_n , kas bāzeta uz B_i , ir uzdota kā

$$\hat{T}_{l,i} = \sqrt{\nu_{j_1(l)}(l)}(\hat{H}_{l,i} - \hat{H}_n). \quad (1.35)$$

Tad sadalījuma funkcija (1.34), kas bāzēta uz apakšizlases B_i , ir vienkārši *empīriskā sadalījuma funkcija* (EDF)

$$\hat{Q}_n(x) = \frac{1}{N} \sum_{i=1}^N I_{(\hat{T}_{l,i} \leq x)}, \quad x \in \mathbb{R}. \quad (1.36)$$

Apgalvojums 6. Ja Pieņēmums 1 (skatīt pielikumu) ir spēkā, tad

$$\sup_{x \in \mathbb{R}} |\hat{Q}_n(x) - Q_n(x)| \xrightarrow{p} 0 \quad \text{kad } n \rightarrow \infty.$$

Pierādījumu skatīt [15].

Apgalvojums 7. Pieņemam, ka Pieņēmums 1 ir spēkā un kopas $\hat{Q}_n^{-1}(u) = \inf\{x : \hat{Q}_n(x) \geq u\}$ un $Q_n^{-1}(u) = \inf\{x : Q_n(x) \geq u\}$. Tad

$$\hat{Q}_n^{-1}(u) - Q_n^{-1}(u) \rightarrow 0, \quad \text{kad } n \rightarrow \infty, \quad \forall u \in (0, 1).$$

Pieņēmums 1 un Apgalvojums 7 pasaka, ka sadalījuma funkcija Q_n var tikt tuvināta ar empīrisko sadalījuma funkciju $\hat{Q}_n$, kas iegūta ar apakšizlašu metodi. Kvantile Q_n^{-1} var tikt tuvināta ar izlases kvantili $\hat{Q}_n^{-1}$. Tad katram $0 < \gamma < 1$ intervāls

$$\left(\hat{H}_n + \frac{1}{\sqrt{\nu_{j1}(n)}} \hat{Q}_n^{-1} \left(\frac{\gamma}{2} \right), \hat{H}_n + \frac{1}{\sqrt{\nu_{j1}(n)}} \hat{Q}_n^{-1} \left(1 - \frac{\gamma}{2} \right) \right) \quad (1.37)$$

ir Hursta parametra ticamības intervāls ar apakšizlašu metodi, kur $1 - \gamma$ ir asimptotiskā pārklājuma precizitāte.

2 Maiņas punkta analīze

2.1 Maiņas punkta analīze

Iepazīsimies ar maiņas punkta analīzes vispārīgo raksturojumu (izmantota literatūra [19]). Maiņas punkta analīžu metodēs ir nepieciešams pārliecināties par to, vai modelis ir laikā nemainīgs. Ja novērots, ka modelī ir notikušas izmaiņas, ir vērts apskatīt galvenos jautājumus, kurus cenšas risināt ar maiņas punkta analīzes palīdzību:

- Kad modelī ir notikušas izmaiņas?
- Vai modelī ir bijusi tikai viena izmaiņa vai vairākas?
- Kāds ir kopējais maiņas punktu skaits? utt.

Definīcija 9. Laika momentu, kurā statistiskajā modelī ir parādījušās izmaiņas, sauc par *maiņas punktu*.

Piemērs 2. Modelī ir iespējamās dažāda veida statistiskās izmaiņas, piemēram, izmaiņas procesa vidējā vērtībā, dispersijā, kā varam redzēt 2.1.attēlā.

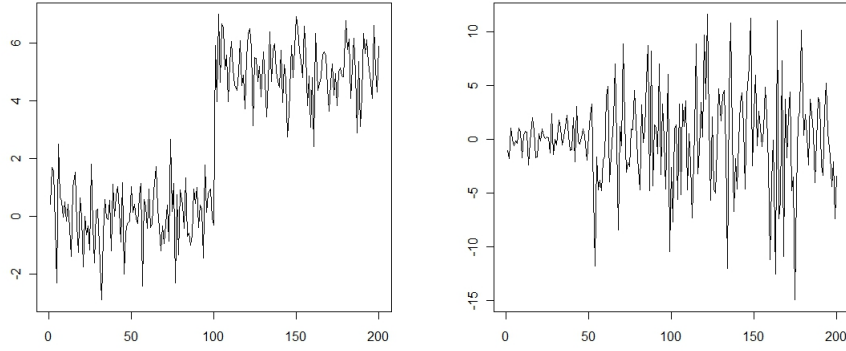
Tā kā šajā darbā pievērsīsimies procesiem, kuriem mainās vidējā vērtība, tad apskatīsim vispārējo maiņas punkta problēmas nostādni vidējās vērtības hipotēžu pārbaudes gadījumā, kad sākuma vidējā vērtība un dispersija nav zināmas. Tad nulles hipotēze H un alternatīva A pie-rakstāmas

$$\begin{aligned} H : \quad Y_i &= \mu + \varepsilon_i, & i &= 1, \dots, n, \\ A : \quad \exists m \in \{1, \dots, n-1\} \text{ tāds, ka} & & & (2.11) \end{aligned}$$

$$Y_i = \mu + \varepsilon_i, \quad i = 1, \dots, m, \quad (2.12)$$

$$Y_i = \mu + \delta + \varepsilon_i, \quad i = m+1, \dots, n, \delta \neq 0, \quad (2.13)$$

kur $\mu, \delta \neq 0$ ir parametri, $\varepsilon_1, \dots, \varepsilon_n$ ir gadījuma lielumi ar sadalījuma funkciju F , vidējo vērtību nulle un dispersiju σ^2 .



Att. 2.1: Pa kreisi: laikrinda, kurai mainās vidējā vērtība; pa labi- laikrinda, kurai mainās dispersija

Definīcija 10. Tādu m , kam izpildās (2.11), (2.12) un (2.13), sauc par maiņas punktu.

2.1.1 *CUSUM* algoritms

Viena no vienkāršākajām metodēm, kā noteikt, vai laikrindā pastāv maiņas punkts, ir *CUSUM* (kumulatīvo summu) metode, kuru 1955. gadā izstrādāja E.S.Page. *CUSUM* algoritms (izmantota literatūra [20]) neatkarīgiem datiem $X_1, \dots, X_n$ ir šāds :

1. Aprēķina vidējo vērtību laikrindā $\bar{X} = \sum_{i=1}^n X_i / n$.
2. Definē sākotnējo kumulatīvo summu $S_0 = 0$.
3. Aprēķina kumulatīvās summas $S_i = S_{i-1} + (X_i - \bar{X})$, kad $i = 1, \dots, n$.
4. Aprēķina $S_{diff} = S_{max} - S_{min}$, kur $S_{max} = \max_{i=1, \dots, n} S_i$ un $S_{min} = \min_{i=1, \dots, n} S_i$.

Lai iegūtu *CUSUM* testa nozīmības līmeni, algoritmā nepieciešams ieviest sekojošu butstrapa procedūru:

5. No sākotnējās izlases iegūst butstrapoto izlasi bez atkārtojumiem $X_1^*, \dots, X_n^*$.
6. Analogiski iepriekšējiem soļiem definē $S_0^* = 0$ un aprēķina butstrapotās $S_1^*, \dots, S_n^*$, kā arī S_{diff}^*, S_{max}^* un S_{min}^* .
7. Aprēķina $k = \sum_{j=1}^N I_{\{S_{diff} > S_{diff}^*\}}$, kur N – butstrapa reižu skaits un indikatorfunkcija

definēta kā

$$I_{\{S_{diff} > S_{diff}^*\}} = \begin{cases} 1, & S_{diff} > S_{diff}^* \\ 0, & S_{diff} \leq S_{diff}^* \end{cases}.$$

8. Iegūst $1 - \alpha$ ticamības līmeni, kas vienāds ar $100 \frac{k}{N}$.

Kad noskaidrots, ka procesam eksistē maiņas punkts m , tā atrašanās vietu var noteikt ar *CUSUM* novērtējumu: $|S_m| = \max_{i=1, \dots, n} |S_i|$, vai arī vidējās kvadrātiskās kļūdas (*MSE*) novērtējumu.

2.1.2 Maiņas punkta analīze stacionāru stohastisku procesu gadījumā

Iepriekš tika aprakstīts modelis, kurā būtisks pieņēmums bija datu neatkarība. Ja šāds pieņēmums neizpildās, nepieciešams maiņas punkta problēmu risināt citādi. Aplūkosim šādu modeli, kas piemērots, lai aplūkotu vidējās vērtības izmaiņas laikrindās (izmantotā literatūra no A. Vasselāna diplomdarba, skatīt mājas lapā [21]). Novērojumus apzīmēsim ar Y_i , kurus iegūstam no

$$Y_i = \begin{cases} \mu_1 + V_i, & 1 \leq i \leq \tilde{k}, \\ \mu_2 + V_i, & \tilde{k} \leq i \leq T, \end{cases} \quad (2.14)$$

kur V_i ir stacionārs process ar vidējo vērtību nulle, $i = 1, \dots, T$ un k, μ_1, μ_2 nav zināmi. Tad problemātika izsakāma ar hipotēžu pārbaudi

$$H_0 : \tilde{k} < T, \mu_1 \neq \mu_2 \text{ pret } A : \tilde{k} = T. \quad (2.15)$$

Tad *CUSUM* statistika

$$C_T = \max_{1 \leq k \leq T} \left| \frac{1}{\sqrt{T}} \sum_{j=1}^k (Y_j - \bar{Y}_T) \right|. \quad (2.16)$$

Apgalvojums 8.

$$\frac{C_T}{\tau} \rightarrow^d \sup_{0 \leq t \leq 1} |B(t)|, \quad (2.17)$$

kur $B(t)$ ir standarta Brauna tilta process un $\tau^2 = 2\pi f(0)$, kur f ir spektrālā blīvuma funkcija procesam V_i . Kritiskās vērtības iegūst, izmantojot butstrapa metodes.

Maiņas punkta noteikšanai ar *CUSUM* metodi, šajā darbā tiks izmantota programmas **R** paketē *strucchange* iebūvētā funkcija *efp*.

3 Asimptotiskā $\bar{X}$ - $\bar{Y}$ uzvedība

Pirms pētīsim maiņas punkta testus, apskatīsim divu izlašu Gausa testu, kuru tradicionāli izmanto (skatīt [22]), lai noteiktu divu normāli sadalītu izlašu lokācijas parametra starpību. Šis tests ir kā īpašgadījums no vidējo vērtību starpību testa, ko aplūkosim vēlāk.

Pieņemsim, ka tiek novēroti neatkarīgi un vienādi sadalīti dati (*iid*) un pēc m -tā novērojuma ir bijusi izmaiņa vidējā vērtībā

$$\begin{aligned} X_1, X_2, \dots, X_m &\sim N(\mu_1, \sigma^2), \\ X_{m+1}, X_{m+2}, \dots, X_{m+n} &\sim N(\mu_2, \sigma^2). \end{aligned} \quad (3.01)$$

Mērķis ir noskaidrot, vai izmaiņa patiešām ir bijusi: Veiksim hipotēžu pārbaudi

$$H : \mu_1 = \mu_2 \quad (3.02)$$

pret alternatīvu

$$A : \mu_1 \neq \mu_2. \quad (3.03)$$

Šo problēmu apzīmēsim ar (H, A) . Loģiska ideja, kā noteikt, vai maiņas punkts ir bijis, ir salīdzināt abu izlašu vidējās vērtības. Šim nolūkam pārsauksim otro izlasi par Y -izlasi ($Y_k := Y_{m+k}$). Tad

$$\begin{aligned} X_1, X_2, \dots, X_m &\sim N(\mu_1, \sigma^2), \\ Y_1, Y_2, \dots, Y_n &\sim N(\mu_2, \sigma^2). \end{aligned}$$

Apgalvojums 9. [22] *Gadījumā, kad dati ir neatkarīgi un normāli sadalīti (*iid*) ar nezināmu dispersiju σ^2 problēmai (H, A) izmanto t -testu, kurš balstās uz vidējo vērtību starpībām*

$$T = \sqrt{\frac{mn}{m+n}} \frac{\bar{X} - \bar{Y}}{s_{X,Y}}$$

ar dispersiju

$$s_{X,Y}^2 = \frac{(m-1)s_x^2 + (n-1)s_y^2}{m+n-2},$$

kur s_x^2 un s_y^2 ir X un Y izlašu empīriskās dispersijas. Ja izpildās H no (3.02), tad $T \sim^d t_{m+n-2}$. Hipotēze H tiek noraidīta pie nozīmības līmeņa α , ja $|T| > t_{\alpha/2, n+m-2}$, kur $t_{\alpha, q}$ ir Stjudenta t -sadatlījuma augšējā α -kvantile ar brīvības pakāpi q .

Tagad aplūkosim LRD gadījumu ar stacionāra Gausa procesa novērojumiem $(X_i)_{i \geq 1}$ ar vidējo vērtību 0, dispersiju 1 un autokovariāciju funkciju

$$\gamma_k = Cov[X_i, X_{i+k}] \sim k^{-D} L(k). \quad (3.04)$$

Tiek sagaidīts, ka statistika $\bar{X} - \bar{Y}$ parādīs citādu uzvedību šajā gadījumā, tā kā ilglaicīgā atmiņa pieprasa *stingrākus* normalizācijas koeficientus un bieži vien nenoved pie normāli sadalītiem robežsadatlījumiem.

3.1 Asimptotiskā ekvivalence un lēni variējošas funkcijas

Definīcija 11. [22] Divas reālas funkcijas f un g sauc par *asimptotiski ekvivalentām*, kad $x \rightarrow \infty$ (pieraksta $f \sim g$), ja

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 1. \quad (3.11)$$

Definīcija 12. [22] Reālu funkciju $L : (0, \infty) \rightarrow (0, \infty)$ sauc par *lēni variējošu*, ja visiem $a > 0$

$$\lim_{x \rightarrow \infty} \frac{L(ax)}{L(x)} = 1.$$

Jebkura funkcija ar $\lim_{x \rightarrow \infty} L(x) = b \in (0, \infty)$ un jebkuras pakāpes logaritms $L(x) = \log^c x$, $c \in \mathbb{R}$ ir lēni variējošs. Parastās pakāpes $f(x) = x^c$ nav lēni variējošas.

3.2 Vienas sadalītas izlases gadījums

Aplūkosim asimptotisko Gausa testa uzvedību divām izlasēm, kas iegūtas (3.01), tātad no vienas ilglaicīgi atkarīgas laukrindas, kas sadalīta divās izlasēs.

3.2.1 Asimptotiskā teorija

Teorēma 10. [22] Pieņemsim, ka $(X_i)_{i \geq 1}$ ir stacionārs Gausa process ar vidējo vērtību 0, dispersiju 1 un kovariācijām

$$\gamma(k) = Cov[X_i, X_{i+k}] \sim \frac{c}{\Gamma(1-D)} k^{-D} L(x),$$

kur c ir konstante, D - ilglaicīgās atmiņas parametrs, $D \in (0, 1)$ un L ir lēni variējoša funkcija. Pieņemam, ka mums ir izlase ar N novērojumiem, kas sadalīta divās daļās

$$X_1, X_2, \dots, X_m \quad \text{un} \quad X_{m+1}, X_{m+2}, \dots, X_{m+n},$$

kur $N = m + n$, $m = \lfloor \lambda N \rfloor$ un $n = \lfloor (1 - \lambda)N \rfloor$ kādai $\lambda \in (0, 1)$. Konstante λ nosaka maiņas punkta atrašanās vietu. Otrā izlasi apzīmēsim kā Y -izlasi ($Y_k := Y_{m+k}$). Tad abu izlašu vidējās vērtības starpībai izpildās

$$\sqrt{\frac{mn}{(m+n)^{2-D}L(m+n)}} \frac{\bar{X} - \bar{Y}}{\sigma_{diff}} \rightarrow^d N(0, 1), \quad (3.21)$$

kur

$$\begin{aligned} \sigma_{diff}^2 &= \frac{2c}{\Gamma(3-D)} \lambda^{1-D} (1-\lambda) + \lambda (1-\lambda)^{1-D} - 1 + \lambda^{2-D} + (1-\lambda)^{2-D} \\ &= \frac{2c}{\Gamma(3-D)} \lambda^{1-D} + (1-\lambda)^{1-D} - 1. \end{aligned}$$

3.2.2 Simulācijas

Pielietosim Teorēmu 10 galīgai izlasei. Ar programmas **R** paketi *fArma* 10000 reizes simulēsim *fGn* procesu garumā $N = 10, 50, 100, 500, 1000, 2000$ ar Hursta parametru H jeb $D = 2 - 2H$. Šajā gadījumā autokovariācijas ir

$$\begin{aligned} \gamma_n &\sim \left(1 - \frac{D}{2}\right) (1-D) k^{-D} \\ &= \frac{c}{\Gamma(1-D)} k^{-D} L(k) \quad \text{ar} \quad L(k) \equiv 1 \quad \text{un} \quad c = \frac{1}{2} \Gamma(3-D). \end{aligned}$$

Tad no Teorēmas 10

$$T := \sqrt{\frac{mn}{(m+n)^{2-D}}} \frac{\bar{X} - \bar{Y}}{\sigma_{diff}} \rightarrow^d N(0, 1),$$

kur

$$\sigma_{diff}^2 = (\lambda^{1-D} (1-\lambda) + \lambda (1-\lambda)^{1-D} - 1 + \lambda^{2-D} + (1-\lambda)^{2-D}).$$

Laikrindai, kura tiek simulēta 10000 reizes, tiek aprēķināta statistika T , un, balstoties uz šīm vērtībām, aprēķināta T dispersija. Visas dispersijas ir diezgan tuvas vērtībai 1, neatkarīgi no laikrindas garuma vai maiņas punkta atrašanās vietas. Rezultāti 3.1., 3.2. un 3.3. tabulā.

Tabula 3.1: Simulēts fGn process ar Hursta parametru $H = 0.6$ ($D = 0.8$). Izlases apjoms N , izlase sadalīta divās daļās pēc $[\lambda N]$ novērojuma. Aprēķināta statistikas T dispersija. Simulācijas veiktas 10000 reižu

N/λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	1.006	1.012	1.027	1.019	1.015	1.018	1.029	0.998	1.004
50	1.002	1.012	1.015	1.006	0.997	0.986	1.016	1.006	1.023
100	0.966	0.999	0.989	0.995	0.991	1.003	1.005	0.982	0.981
500	0.993	0.979	0.973	1.003	0.997	0.998	0.996	0.996	0.974
1000	0.999	1.010	1.016	1.010	1.014	1.025	1.015	1.010	1.021
2000	0.991	1.011	0.991	0.979	0.992	1.025	1.030	1.033	1.007

Tabula 3.2: Simulēts fGn process ar Hursta parametru $H = 0.7$ ($D = 0.6$). Izlases apjoms N , izlase sadalīta divās daļās pēc $[\lambda N]$ novērojuma. Aprēķināta statistikas T dispersija. Simulācijas veiktas 10000 reižu

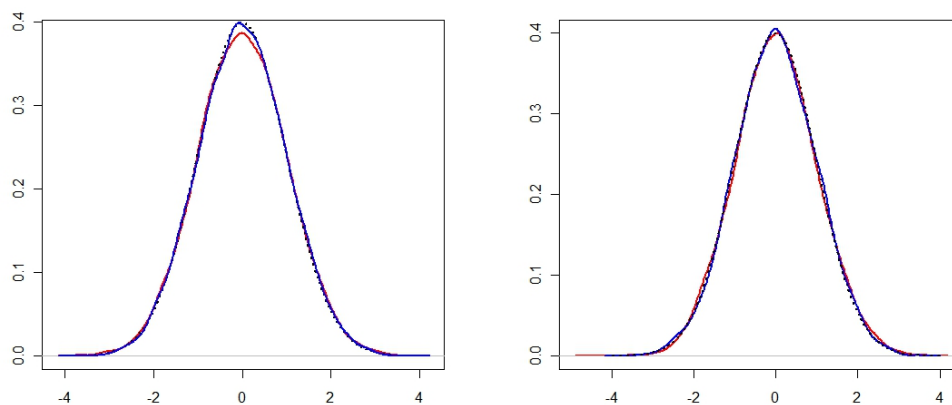
N/λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	1.007	1.013	1.028	1.021	1.018	1.022	1.031	1.002	1.005
50	1.001	1.011	1.015	1.006	0.995	0.985	1.010	1.003	1.019
100	0.967	0.997	0.989	0.994	0.991	1.003	1.005	0.985	0.984
500	0.992	0.979	0.975	1.002	0.997	0.998	0.996	0.994	0.975
1000	0.999	1.011	1.015	1.012	1.017	1.027	1.019	1.015	1.021
2000	0.991	1.007	0.989	0.979	0.993	1.023	1.029	1.034	1.009

Tabula 3.3: Simulēts fGn process ar Hursta parametru $H = 0.9$ ($D = 0.2$). Izlases apjoms N , izlase sadalīta divās daļās pēc $[\lambda N]$ novērojuma. Aprēķināta statistikas T dispersija. Simulācijas veiktas 10000 reižu

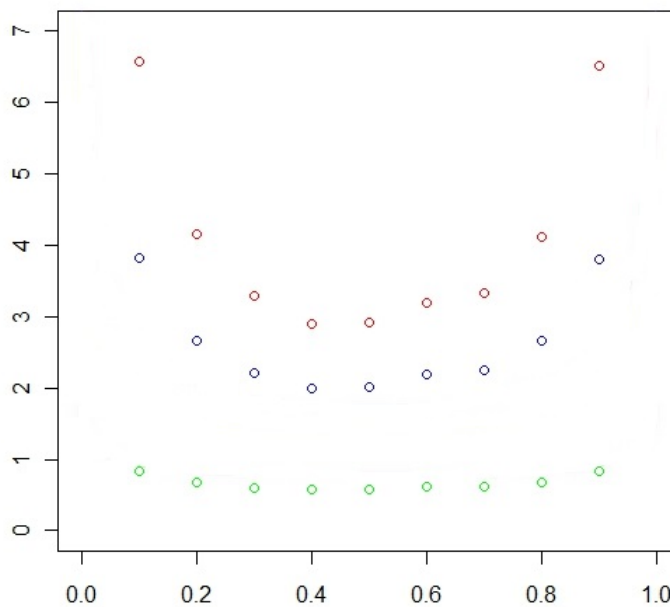
N/λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	1.009	1.017	1.029	1.026	1.025	1.028	1.034	1.009	1.009
50	0.999	1.008	1.011	1.003	0.992	0.984	0.999	0.997	1.009
100	0.969	0.994	0.989	0.993	0.992	1.003	1.005	0.989	0.989
500	0.992	0.978	0.979	1.000	0.997	0.997	0.996	0.991	0.977
1000	1.003	1.012	1.015	1.016	1.021	1.029	1.025	1.022	1.021
2000	0.991	1.002	0.987	0.984	0.997	1.021	1.028	1.037	1.016

3.1. attēlā novērtētā T blīvuma funkcija salīdzināta ar standart normālo blīvuma funkciju. Kā var redzēt blīvuma funkcijas pie dažādiem apjomiem ir līdzīgas. Gadījumā, kad $N = 2000$

blīvuma funkcija un standartnormālā blīvuma funkcija pārklājas ļoti labi.



Att. 3.1: Statistikas $\bar{X} - \bar{Y}$ blīvuma funkcija, skalēta un normalizēta. 10000 reizes simulēts fGn process ar $H = 0.7$ ($D = 0.6$) apjomā $N = 10$ (sarkanā līnija) un $N = 2000$ (zilā līnija). Ar raustīto līniju $N(0, 1)$ blīvuma funkcija. Pa kreisi - $\lambda = 0.3$, pa labi - $\lambda = 0.5$



Att. 3.2: $N^{D/2}(\bar{X} - \bar{Y})$ dispersija atkarībā no λ fGn procesam ar $H = 0.6$ (sarkanā krāsā), $H = 0.7$ (zilā krāsā) un $H = 0.9$ (zaļā krāsā). Simulācijas veiktas 10000 reižu, izlases apjoms $N = 2000$

Lai saprastu, kā dažādiem λ un H mainās statistikas $\bar{X} - \bar{Y}$ dispersija, var izteikt

$$T := \sqrt{\frac{mn}{(m+n)^{2-D}}} \frac{\bar{X} - \bar{Y}}{\sigma_{diff}} = \sqrt{\frac{\lambda(1-\lambda)}{N^{-D}}} \frac{\bar{X} - \bar{Y}}{\sigma_{diff}}.$$

3.2 attēlā pārādīta nenormalizētā statistika $N^{D/2}(\bar{X} - \bar{Y})$. Kā varēja sagaidīt, dispersija ir minimāla pie aptuveni vienādiem izlašu apjomiem ($\lambda = 0.5$), un, jo lielāka atkarības struktūra, jo mazāka ir dispersija. To var izskaidrot arī intuitīvi: laikrindā, kur atkarības struktūra nav tik stipra, Y novērojumiem ir lielāka iespēja atšķirties no X novērojumiem, bet, kad laikrindā ir ļoti stipra atkarība, abas izlases ir līdzīgas, kas noved pie mazākas dispersijas. Tas pats izpildās pie izlašu apjomiem: ja pirmā izlase ir lielāka, otrajai izlasei ir mazākas iespējas kompensēt pirmās izlases svārstības.

4 Vilkoksona un vidējo vērtību starpības testi

Šajā nodaļā apskatīsim maiņas punkta problēmas ilglaicīgi atkarīgām laikrindām $(X_i)_{i \geq 1}$ (izmantojot [1]), kas ir formā

$$X_i = \mu_i + \epsilon_i,$$

kur $(\epsilon_i)_{i \geq 1}$ ir ilglaicīgi atkarīgs stacionārs process ar vidējo vērtību nulle un kur $(\mu_i)_{i \geq 1}$ ir nezināmas vidējās vērtības. Apskatīsim gadījumu, kad $(\epsilon_i)_{i \geq 1}$ ir funkcionālis no stacionāra Gausa procesa, tas ir,

$$\epsilon_i = G(\xi_i), \quad i \geq 1.$$

Pieņemsim, ka $(\xi_i)_{i \geq 1}$ ir Gausa process ar vidējo vērtību nulle, $E(\xi_i^2) = 1$ un autokovariāciju funkciju

$$\rho(k) = k^{-D} L(k), \quad k \geq 1, \quad (4.01)$$

kur $0 < D < 1$ ir ilglaicīgās atmiņas parametrs un $L(k)$ ir lēni variējoša funkcija. Turklāt, $G : \mathbb{R} \rightarrow \mathbb{R}$ ir mērojama funkcija, kas apmierina $E(G(\xi_i)) = 0$.

Balstoties uz novērojumiem $X_1, \dots, X_n$, gribam pārbaudīt hipotēzi

$$H : \mu_1 = \dots = \mu_n, \quad (4.02)$$

ka datu vidējā vērtība nemainās pret alternatīvo hipotēzi

$$A : \mu_1 = \dots = \mu_k \neq \mu_{k+1} = \dots = \mu_n \text{ kādam } k \in \{1, \dots, n-1\}.$$

Šo problēmu apzīmēsim ar (H, A) .

Maiņas punkta testa problēmas parasti sākas no divu izlašu problēmas (skatīt [1]), kad tiek pieņemts, ka maiņas punkts ir zināms, tas ir, dotam $k \in \{1, \dots, n-1\}$, alternatīvā hipotēze ir

$$A_k : \mu_1 = \dots = \mu_k \neq \mu_{k+1} = \dots = \mu_n.$$

Testa problēmai (H, A_k) Vilkoksona divu izlašu rangs tests ir visbiežāk lietotais neparametris-
kais tests. Vilkoksona tests tiek noraidīts pie lielām un mazām testa statistikas vērtībām

$$W_{k,n} = \sum_{i=1}^k \sum_{j=k+1}^n 1_{\{X_i \leq X_j\}}.$$

$W_{k,n}$ saskaita, cik daudz reizes izlases otrās daļas vērtības pārsniedz izlases pirmās daļas vērtības. Jāatzīmē, ka šī ir Mann - Whitney U-statistikas reprezentācija Vilkoksona ranga testa statistikai (skatīt Lehmann [23]).

Apgalvojums 11. [1] *Maiņas punkta problēmas (H, A) testi, kuros netiek pieņemts, ka k ir zināms, var tikt bāzēti uz testa statistikas $W_{1,n}, \dots, W_{n-1,n}$ vektoru. Nulles hipotēze tiks noraidīta pie lielām $W_n = \max_{k=1, \dots, n-1} |W_{k,n}|$.*

Piezīme 12. *Iespējams izvēlēties arī citas funkcijas no vērtībām $W_{1,n}, \dots, W_{n-1,n}$ - nav obligāti jāizvēlas maksimālā vērtība.*

4.1 Vilkoksona testa asimptotiskais sadalījums

Lai noteiku Vilkoksona testa kritiskās vērtības, jāzina W_n sadalījums pie nulles hipotēzes, ka nav maiņas punkta. Tā kā precīzu sadalījumu ir grūti iegūt, tiek izmantots asimptotiskais sadalījums pie lieliem izlases apjomiem. Šim mērķim apskatīsim procesu

$$W_{n,[n\lambda]} = \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n 1_{\{X_i \leq X_j\}}, \quad 0 \leq \lambda \leq 1,$$

kas atkarīgs no parametra λ .

Apgalvojums 13. [1] *Veicot procesa centrēšanu un normalizēšanu, iegūstam*

$$W_n(\lambda) = \frac{1}{nd_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n \left(1_{\{X_i \leq X_j\}} - \int_{\mathbb{R}} F(x) dF(x) \right), \quad 0 \leq \lambda \leq 1. \quad (4.11)$$

Apgalvojums 14. [1]

- *Ja sadalījuma funkcija F ir nepātraukta, tad centrējošā konstante*

$$\int_{\mathbb{R}} F(x) dF(x) = \frac{1}{2}.$$

- *Ja izpildās nulles hipotēze, $W_n(\lambda)$ sadalījums nav atkarīgs no vidējās vērtības, tas ir, $\mu := \mu_1 = \dots = \mu_n$. Tādējādi, var pieņemt, ka $\mu = 0$ un no šejienes $X_i = G(\xi_i)$.*

Lai analizētu asimptotisko sadalījumu procesam $(W_n(\lambda))_{0 \leq \lambda \leq 1}$, tiks pielietots empīriskā procesa invariāciju princips virknei $(G(\xi_i))_{i \geq 1}$, kas parādīts Dehling un Taqqu publikācijā [24]. Aplūkosim Hermītu izvirzījumu (*Hermite expansion*)

$$1_{\{G(\xi_i) \leq x\}} - F(x) = \sum_{q=1}^{\infty} \frac{J_q(x)}{q!} H_q(\xi_i),$$

kur H_q ir q -tās kārtas Hermītu polinoms (piemēru skatīt pielikumā (6.32) un kur

$$J_q(x) = E H_q(\xi_i) 1_{\{G(\xi_i) \leq x\}}.$$

Definīcija 13. [1] Hermītu rangs funkciju klasei $\{1_{\{G(\xi_i) \leq x\}} - F(x), x \in \mathbb{R}\}$ ir

$$m := \min\{q \geq 1 : J_q(x) \neq 0 \text{ kādam } x \in \mathbb{R}\}, \quad (4.12)$$

Definīcija 14. [1] Normalizējošā konstante funkciju klasei $\{1_{\{G(\xi_i) \leq x\}} - F(x)\}$ ir

$$d_n^2 = \text{Var} \left(\sum_{j=1}^n H_m(\xi_j) \right). \quad (4.13)$$

Turpmāk izmantosim pieņēmumu, ka $0 < D < \frac{1}{m}$, un šajā gadījumā

$$d_n^2 \sim c_m n^{2-mD} L^m(n), \quad (4.14)$$

kur $c_m = \frac{2m!}{(1-Dm)(2-Dm)}$. Simbols $\sim$ apzīmē asimptotisko ekvivalenci (skatīt 3.11). Tad, pielietojot procesu invariāciju principu, ko aprakstīja Dehling un Taqqu, mēs iegūstam, ka

$$\left(d_n^{-1} [n\lambda] (F_{[n\lambda]}(x) - F(x)) \right)_{x \in [-\infty, \infty], \lambda \in [0, 1]} \rightarrow^d \left(\frac{J_m(x)}{m!} Z_m(\lambda) \right)_{x \in [-\infty, \infty], \lambda \in [0, 1]},$$

kad $D([-\infty, \infty] \times [0, 1])$, kur $(Z_m(\lambda))_{\lambda \in [0, 1]}$ ir m -tās kārtas Hermīta process, ko nedefinējis Taqqu [25].

Process $(Z_m(\lambda))_{\lambda \geq 0}$ ir sev - līdzīgs process ar parametru

$$H = 1 - \frac{mD}{2} \in \left(\frac{1}{2}; 1 \right),$$

tas ir, procesiem $Z_m(c\lambda)$ un $c^H Z_m(\lambda)$ ir viens un tas pats galīgdimensionālais sadalījums visām konstantēm $c > 0$. Kad $m = 1$, $Z_1(\lambda)$ ir standarta Brauna kustības process ar vidējo vērtību nulle un kovariāciju

$$E(Z_1(\lambda_1)Z_1(\lambda_2)) = \frac{1}{2} \{ \lambda_1^{2H} + \lambda_2^{2H} - |\lambda_1 - \lambda_2|^{2H} \}. \quad (4.15)$$

Teorēma 15. [1] Pieņemsim, ka $(\xi_i)_{i \geq 1}$ ir stacionārs Gausa process ar vidējo vērtību nulle, dispersiju 1, un autokovariāciju funkciju izskatā (4.01) ar $0 < D < \frac{1}{m}$. Vēl vairāk, pieņemsim, ka $G : \mathbb{R} \rightarrow \mathbb{R}$ ir mērojama funkcija un definēsim

$$X_k = G(\xi_k).$$

Pieņemsim, ka X_k ir nepārtraukta sadalījuma funkcija F . Ar m apzīmēsim Hermītu rangū funkciju klasei $1_{\{G(\xi_i) \leq x\}} - F(x)$, $x \in \mathbb{R}$, kā definēts vienādojumā (4.12) un pieņemsim, ka $d_n > 0$ apmierina vienādojumu (4.14). Tad

$$\frac{1}{nd_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n \left(1_{\{X_i \leq X_j\}} - \int_{\mathbb{R}} F(x) dF(x) \right), \quad 0 \leq \lambda \leq 1 \quad (4.16)$$

konverģē pēc sadalījuma uz procesu

$$\frac{1}{m!} (Z_m(\lambda) - \lambda Z_m(1)) \int_{\mathbb{R}} J_m(x) dF(x), \quad 0 \leq \lambda \leq 1.$$

Teorēmas pierādījumu skatīt literatūrā [1].

4.2 Vilkoksona rangū tests

Maiņas punkta tests, kas balstīts uz Vilkoksona rangū testu noraidīs nulles hipotēzi lielām W_n vērtībām, kur

$$W_n = \frac{1}{nd_n} \max_{1 \leq k \leq n-1} \left| \sum_{i=1}^k \sum_{j=k+1}^n \left(1_{\{X_i \leq X_j\}} - \frac{1}{2} \right) \right|. \quad (4.21)$$

Vispirms atzīmēsim testa statistikas W_n invariances īpašību.

Lemma 16. [1] Testa statistika W_n ir invarianta pie datu stingri monotonas transformācijas, tas ir,

$$\frac{1}{nd_n} \max_{1 \leq k \leq n-1} \left| \sum_{i=1}^k \sum_{j=k+1}^n \left(1_{\{G(X_i) \leq G(X_j)\}} - \frac{1}{2} \right) \right| = \frac{1}{nd_n} \max_{1 \leq k \leq n-1} \left| \sum_{i=1}^k \sum_{j=k+1}^n \left(1_{\{X_i \leq X_j\}} - \frac{1}{2} \right) \right|$$

visām stingri monotonām funkcijām $G : \mathbb{R} \rightarrow \mathbb{R}$.

Pierādījums. Ja funkcija G ir stingri augoša, tas ir acīmredzami, ka $G(X_i) \leq G(X_j)$ tad un tikai tad, ja $X_i \leq X_j$. Ja G ir stingri dilstoša, tad $G(X_i) \leq G(X_j)$ tad un tikai tad, ja $X_i \geq X_j$, tātad

$$1_{\{G(X_i) \leq G(X_j)\}} = 1 - 1_{\{X_i \leq X_j\}}.$$

No šejienes iegūstam

$$\sum_{i=1}^k \sum_{j=k+1}^n \left(1_{\{G(X_i) \leq G(X_j)\}} - \frac{1}{2} \right) = - \sum_{i=1}^k \sum_{j=k+1}^n \left(1_{\{X_i \leq X_j\}} - \frac{1}{2} \right),$$

un tādējādi lemma ir pierādīta. $\square$

Tā kā $X_i = \mu_i + G(\xi_i)$, pie nulles hipotēzes

$$W_n = \frac{1}{nd_n} \max_{1 \leq k \leq n-1} \left| \sum_{i=1}^k \sum_{j=k+1}^n \left(1_{\{G(\xi_i) \leq G(\xi_j)\}} - \frac{1}{2} \right) \right|. \quad (4.22)$$

Tad, pielietojot Teorēmu 15 un nepārtrauktās attēlošanas teorēmu (skatīt pielikumā (Teorēma 26)), pie nulles hipotēzes W_n konverģē pēc sadalījuma, kad $n \rightarrow \infty$, uz

$$\sup_{0 \leq \lambda \leq 1} \left| \frac{Z_m(\lambda)}{m!} - \lambda \frac{Z_m(1)}{m!} \right| \left| \int_{\mathbb{R}} J_m(x) dF(x) \right|.$$

Lai noteiktu asimptotiskās testa, kas balstās uz W_n , kritiskās vērtības, jāaprēķina sadalījums vienādojuma labajā pusē. Pieņemam, ka G ir stingri monotona funkcija. Savienojot (4.22) ar Lemmu 16, iegūst, ka

$$W_n = \frac{1}{nd_n} \max_{1 \leq k \leq n-1} \left| \sum_{i=1}^k \sum_{j=k+1}^n \left(1_{\{\xi_i \leq \xi_j\}} - \frac{1}{2} \right) \right|.$$

Šajā gadījumā

$$J_1(x) = E(\xi 1_{\{\xi \leq x\}}) = \int_{-\infty}^x t \varphi(t) dt = -\varphi(x),$$

kur $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ apzīmē standartnormālo blīvuma funkciju. Pēdējā solī tiek izmantots fakts, ka $\varphi'(t) = -t\varphi(t)$. Tādā veidā $J_1(x) \neq 0$ visiem x un no šejienes funkciju klasei $\{1_{\{\xi \leq x\}}, x \in \mathbb{R}\}$ ir Hermīta rangs $m = 1$, un tā kā F ir normālā sadalījuma funkcija, iegūstam

$$\int_{\mathbb{R}} J_1(x) dF(x) = - \int_{\mathbb{R}} \varphi(x) \varphi(x) dx = \int_{\mathbb{R}} -(\varphi(s))^2 ds = -\frac{1}{2\sqrt{\pi}}. \quad (4.23)$$

Tad spēkā ir sekojoša teorēma.

Teorēma 17. [1] Pieņemsim, ka $(\xi_i)_{i \geq 1}$ ir Gausa process ar vidējo vērtību nulle, $Var(\xi_i) = 1$ un autokovariāciju funkciju $\rho(k)$, kas apmierina (4.01) ar $0 < D < 1$. Vēl vairāk, $G : \mathbb{R} \rightarrow \mathbb{R}$ ir stingri monotona funkcija, kas definē $X_i = \mu_i + G(\xi_i)$. Tad, pie nosacījuma, ka izpildās nulles hipotēze $H : \mu_1 = \dots = \mu_n$, testa statistika W_n , kas definēta kā (4.21), konverģē pēc sadalījuma uz

$$\frac{1}{2\sqrt{\pi}} \sup_{0 \leq \lambda \leq 1} |Z_1(\lambda) - \lambda Z_1(1)|,$$

kur $(Z_1(\lambda))_{\lambda \geq 0}$ apzīmē standarta Brauna kustības procesu ar Hursta parametru $H = 1 - D/2 \in (1/2, 1)$. Normalizējošā konstante no vienādojuma (4.21) apmierina nosacījumu $nd_n \sim \left(\frac{L(n)}{(1-D)(1-D/2)} \right)^{1/2} n^{2-D/2}$.

4.2.1 Pielietojums

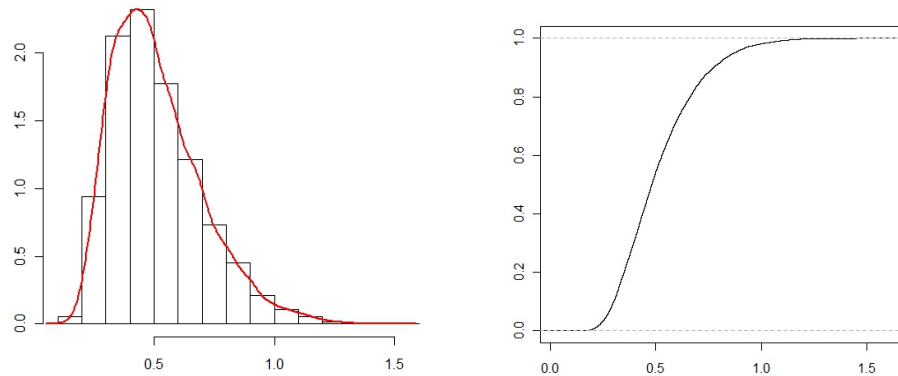
Izmantojot datorprogrammas **R** paketi *fArma* simulējam standarta Brauna kustību $(Z^{(j)}(t))_{0 \leq t \leq 1}$, kur $1 \leq j \leq 10000$, $t = \frac{i}{1000}$ un $0 \leq i \leq 1000$. Aprēķinām

$$M_j := \max_{1 \leq i \leq 1000} \left| Z^{(j)}\left(\frac{i}{1000}\right) - \frac{i}{1000} Z^{(j)}(1) \right|,$$

kas ir skaitlisks novērtējums lielumam $\sup_{0 \leq \lambda \leq 1} |Z^{(j)}(\lambda) - \lambda Z^{(j)}(1)|$. Tad empīriskā sadalījuma funkcija

$$F_M(x) := \frac{1}{10000} \# \{1 \leq j \leq 10000 : M_j \leq x\}$$

tiek izmantota kā novērtējums lieluma $\sup_{0 \leq \lambda \leq 1} |Z(\lambda) - \lambda Z(1)|$ sadalījumam. 4.1 attēlā redzam novērtēto blīvuma un empīrisko sadalījuma funkciju gadījumā, kad $H = 0.7$.



Att. 4.1: Varbūtību blīvuma un sadalījuma funkcija lielumam $\sup_{0 \leq \lambda \leq 1} |Z(\lambda) - \lambda Z(1)|$, 10000 simulācijas

4.1 tabulā aprēķinātas attiecīgās augšējās α -kvantiles

$$q_\alpha := \inf\{x : F_M(x) \geq 1 - \alpha\}, \quad (4.24)$$

kad $H = 0.6, 0.7, 0.9$, tas ir, $D = 2 - 2H = 0.8, 0.6, 0.2$. Redzams, ka kvantiļu vērtības samazinās, palielinoties datu atkarības struktūrai.

Tabula 4.1: Sadalījuma $\sup_{0 \leq \lambda \leq 1} |Z(\lambda) - \lambda Z(1)|$ augšējās α -kvantiles, kur Z ir standarta fBm , dažādiem LRD parametriem H , simulēts 10000 reižu

H/α	0.10	0.05	0.01
0.6	0.976	1.089	1.331
0.7	0.773	0.873	1.075
0.9	0.381	0.434	0.543

4.3 Vidējo vērtību starpību tests un tā robežsadalījums

Kā alternatīvu Vilkoksona testam apskatīsim testu, kas balstās uz novērojumu vidējo vērtību starpībām. Apskatām šādu testa statistiku

$$D_n := \max_{1 \leq k \leq n-1} |D_{k,n}|, \quad (4.31)$$

kur

$$D_{k,n} := \frac{1}{nd_n} \sum_{i=1}^k \sum_{j=k+1}^n (X_i - X_j).$$

Pie lielām D_n vērtībām tiek norādīta nulles hipotēze (4.02). Lai iegūtu testa statistikas asimptotisko sadalījumu, izmantojot ne-centrālo robežteorēmu no Taqqu [25], var iegūt, ka

$$\begin{aligned} \frac{1}{nd_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n (X_i - X_j) &= \frac{1}{nd_n} \left((n - [n\lambda]) \sum_{i=1}^{[n\lambda]} G(\xi_i) - [n\lambda] \sum_{j=[n\lambda]+1}^n G(\xi_j) \right) \\ &\rightarrow^d a_m \left((1 - \lambda) \frac{Z_m(\lambda)}{m!} - \lambda \left(\frac{Z_m(1)}{m!} - \frac{Z_m(\lambda)}{m!} \right) \right) \\ &= a_m \frac{1}{m!} (Z_m(\lambda) - \lambda Z_m(1)), \end{aligned}$$

kur m apzīmē $G(\xi)$ Hermīta rangu un kur $a_m = E(G(\xi)H_m(\xi))$ ir Hermītu koeficients. Pielietojot nepārtrauktās attēlošanas teorēmu, iegūstam sekojošu teorēmu, kas nosaka D_n asimptotisko sadalījumu.

Teorēma 18. [1] Pieņemsim, ka $(\xi_i)_{i \geq 1}$ ir Gausa process ar vidējo vērtību nulle, $Var(\xi_i) = 1$ un autokovariāciju funkciju $\rho(k)$, kas apmierina (4.01) ar $0 < D < 1/m$. Vēl vairāk, $G : \mathbb{R} \rightarrow \mathbb{R}$ ir mērojama funkcija, kas apmierina $E(G^2(\xi)) < \infty$ un definē $X_i = \mu_i + G(\xi_i)$. Tad, pie nosacījuma, ka izpildās nulles hipotēze $H : \mu_1 = \dots = \mu_n$, testa statistika D_n , kas definēta kā (4.31), konverģē pēc sadalījuma uz

$$\frac{|a_m|}{m!} \sup_{0 \leq \lambda \leq 1} |Z_m(\lambda) - \lambda Z_m(1)|,$$

kur $(Z_m(\lambda))$ apzīmē m -tās pakāpes Hermīta procesu ar Hursta parametru $H = 1 - Dm/2 \in (1/2, 1)$.

Piezīme 19. Stingri augošai funkcijai G Hermīta rangs ir 1, jo

$$E(G(\xi)H_1(\xi)) = \int_{\mathbb{R}} G(s)H_1(s)\varphi(s)ds = \int_0^\infty s\varphi(s)(G(s) - G(-s))ds > 0.$$

Līdzīgi arī stingri dilstošai funkcijai G mēs iegūstam $E(G(\xi)H_1(\xi)) < 0$. Tāpēc šajā gadījumā testa statistika D_n pie nulles hipotēzes konverģē uz

$$|a_1| \sup_{0 \leq \lambda \leq 1} |Z_1(\lambda) - \lambda Z_1(1)|,$$

kur Z_1 ir frakcionālā Brauna kustība ar parametru $H = 1 - D/2$. Jāievēro, ka šajā gadījumā robežsadalījumi vidējo vērtību starpību testam un Vilkoksona rangu testam atšķiras vien par normalizējošo konstanti.

5 Empīriskā ticamības funkcija ilglaicīgi atkarīga procesa vidējai vērtībai

5.1 Empīriskā ticamība divu izlašu gadījumā

No sākuma aplūkosim empīriskās ticamības funkcijas attiecību testu divu izlašu gadījumā neatkarīgiem datiem (izmantotā literatūra [26]). Apskatīsim *iid* datus $X_1, \dots, X_{n_1}$ un $Y_1, \dots, Y_{n_2}$ ar attiecīgām sadalījuma funkcijām F_1 un F_2 . Apskatīsim interesējošo parametru Δ . Pieņemsim, ka visa informācija par īstajiem parametriem Δ_0 un θ_0 atrodas nenovirzītajās funkcijās

$$E_{F_1} w_1(X, \theta_0, \Delta_0) = 0, \quad (5.11)$$

$$E_{F_2} w_2(Y, \theta_0, \Delta_0) = 0. \quad (5.12)$$

Ja $\Delta_0 = \theta_1 - \theta_0$, kur $\theta_0 = \int x dF_1(x)$ un $\theta_1 = \int y dF_2(x)$, tad izvēloties

$$w_1(X, \theta_0, \Delta_0) = X - \theta_0,$$

$$w_2(Y, \theta_0, \Delta_0) = Y - \theta_0 - \Delta_0,$$

mēs iegūsim vienādojumus (5.11) un (5.12).

Lai noteiktu ticamības intervālus parametram Δ , definēsim empīrisko ticamības attiecību funkciju

$$R(\Delta, \theta) = \sup_{p, q} \prod_{i=1}^{n_1} (n_1 p_i) \prod_{j=1}^{n_2} (n_2 q_j), \quad (5.13)$$

kur $p = (p_1, \dots, p_{n_1})$ un $q = (q_1, \dots, q_{n_2})$ ir divi varbūtību vektori, kurus nosaka ierobežojumi

$$p_i \geq 0, i = 1, \dots, n_1, \sum_{i=1}^{n_1} p_i = 1, \sum_{i=1}^{n_1} p_i w_1(X_i, \theta, \Delta) = 0,$$

$$q_j \geq 0, j = 1, \dots, n_2, \sum_{j=1}^{n_2} q_j = 1, \sum_{j=1}^{n_2} q_j w_2(Y_j, \theta, \Delta) = 0.$$

Vienādojumam (5.13) eksistē viens vienīgs atrisinājums pie nosacījuma, ka 0 pieder izliektajai čaulai, kuru veido punkti $w_1(X_i, \theta, \Delta)$ un izliektajai čaulai, kuru veido $w_2(Y_j, \theta, \Delta)$. Maksimumu iegūst, izmantojot Lagranža reizinātāju metodi, no kuras iegūst, ka

$$p_i = \frac{1}{n_1(1 + \lambda_1(\theta)w_1(X_i, \theta, \Delta))}, i = 1, \dots, n_1,$$

$$q_j = \frac{1}{n_2(1 + \lambda_2(\theta)w_2(Y_j, \theta, \Delta))}, j = 1, \dots, n_2,$$

kur Lagranža reizinātājus $\lambda_1(\theta)$ un $\lambda_2(\theta)$ nosaka no vienādojumiem

$$\sum_{i=1}^{n_1} \frac{w_1(X_i, \theta, \Delta)}{1 + \lambda_1(\theta)w_1(X_i, \theta, \Delta)} = 0, \quad (5.14)$$

$$\sum_{j=1}^{n_2} \frac{w_2(Y_j, \theta, \Delta)}{1 + \lambda_2(\theta)w_2(Y_j, \theta, \Delta)} = 0. \quad (5.15)$$

Definēsim empīrisko log-ticamības attiecību (kas pareizināta ar mīnus divi)

$$\begin{aligned} W(\Delta, \theta) = -2 \log R(\Delta, \theta) &= 2 \sum_{i=1}^{n_1} \log(1 + \lambda_1(\theta)w_1(X_i, \theta, \Delta)) \\ &+ 2 \sum_{j=1}^{n_2} \log(1 + \lambda_2(\theta)w_2(Y_j, \theta, \Delta)). \end{aligned}$$

Novērtējumu $\hat{\theta} = \hat{\theta}(\Delta)$ nepieciešamajam parametram θ , kas maksimizē $R(\Delta, \theta)$ fiksētā Δ , nosaka no vienādojuma

$$\frac{\partial W(\Delta, \theta)}{\partial \Delta} = \sum_{i=1}^{n_1} \frac{\lambda_1(\theta)\alpha_1(X_i, \theta, \Delta)}{1 + \lambda_1(\theta)w_1(X_i, \theta, \Delta)} + \sum_{j=1}^{n_2} \frac{\lambda_2(\theta)\alpha_2(Y_j, \theta, \Delta)}{1 + \lambda_2(\theta)w_2(Y_j, \theta, \Delta)} = 0, \quad (5.16)$$

kur

$$\alpha_1(X_i, \theta, \Delta) = \frac{\partial w_1(X_i, \theta, \Delta)}{\partial \theta} \text{ un } \alpha_2(Y_j, \theta, \Delta) = \frac{\partial w_2(Y_j, \theta, \Delta)}{\partial \theta}.$$

Teorēma 20 (Qin un Zhao [27]). *Izpildoties gluduma nosacījumiem funkcijām w_1, w_2, α_1 un α_2 , eksistē tāds $\hat{\theta}$, ka $\hat{\theta}$ ir būtisks θ_0 novērtējums un funkcija $R(\Delta, \theta)$ sasniedz savu lokālo maksimumu pie $\hat{\theta}$, un*

$$\sqrt{n_1}(\hat{\theta} - \theta_0) \rightarrow^d N\left(0, \frac{\beta_1\beta_2}{\beta_2\beta_{10}^2 + k\beta_1\beta_{20}^2}\right),$$

kur $k < \infty$ ir pozitīva konstante tāda, ka $n_2/n_1 \rightarrow k$, kad $n_1, n_2 \rightarrow \infty$, turklāt,

$$-2 \log R(\Delta_0, \hat{\theta}(\Delta_0)) \rightarrow^d \chi_1^2,$$

kur

$$\beta_1 = E_{F_1} w_1^2(X, \theta_0, \Delta_0), \quad \beta_2 = E_{F_2} w_2^2(Y, \theta_0, \Delta_0),$$

$$\beta_{10} = E_{F_1} \alpha_1(X, \theta_0, \Delta_0), \quad \beta_{20} = E_{F_2} \alpha_2(Y, \theta_0, \Delta_0).$$

Tad punktveida empīriskās ticamības intervāli parametram Δ nosakāmi no $\{R(\Delta, \hat{\theta}) > c\}$ īstajam Δ_0 , kur c ir nosakāms no teorēmas.

5.2 Blokveida empīriskās ticamības konstrukcija vidējai vērtībai

Praksē bieži vien nākas saskarties ar atkarīgām datu struktūrām, kurām iepriekš aplūkotie rezultāti, kas darbojas neatkarīgiem datiem, vairs nav spēkā. Atkarīgām datu struktūrām nepieciešama cita pieeja - sākumā apskatīsim vienas izlases gadījumu.

1997. gadā Kitamura savā publikācijā [28] piedāvāja blokveida EL vāji atkarīgām laikrindām (*short range dependence* jeb SRD), kurām jau līdz šim tika pielietotas tādas bloku veidošanas metodes kā blokveida butstraps un apakšizlašu veidošana [Carlstein [29], Politis un Romano [30]]. Tomēr bloku veidošanas metodes, kas der vāji atkarīgiem datiem bieži neder ilglaicīgi atkarīgiem datiem. Tā kā Kitamura aprakstīta blokveida EL paredzēta SRD procesiem, pie LRD procesiem varētu sākties problēmas, to nepieciešams modificēt.

Tātad mērķis ir sīkāk apskatīt modificēto blokveida EL metodi laikrindai $\{Y_t\}, t \in \mathbb{Z}$ ar vidējo vērtību μ un spektrālo blīvuma funkciju $f(\lambda), |\lambda| \leq \pi$, kas apmierina

$$f(\lambda) \sim c_d |\lambda|^{-2d}, \text{ kad } \lambda \rightarrow 0, \quad (5.21)$$

kur $d \in (-1/2, 1/2)$ un $c_d > 0$ ir pozitīva konstante. Tad eksistē gadījuma lielumi $\{\varepsilon_j\}$ ar vidējo vērtību nulle un dispersiju $0 < E(\varepsilon_j^2) < \infty$ tādi, ka $\{Y_t\}$ ir slīdošā vidējā reprezentācija formā

$$Y_t = \mu + \sum_{j \in \mathbb{Z}} b_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \quad (5.22)$$

kur konstantes $\{b_j\}$ apmierina $0 < \sum_{j \in \mathbb{Z}} b_j^2 < \infty$. Ja $d = 0$, tas ir SRD process, ja $d > 0$, tad LRD process.

Šim nolūkam no sākuma iepazīsimies ar Kitamura piedāvāto blokveida EL , apskatīsim dažus pieņēmumus un izlases vidējās vērtības asimptotisko sadalījumu.

5.2.1 Kitamura piedāvātā blokveida EL konstrukcija

Apskatīsim Kitamura piedāvāto blokveida EL procesa vidējai vērtībai $E(Y_t) = \mu$. Pieņemsim, ka $1 \leq l \leq n$ ir bloka garums un varam definēt blokus $B_i^l = (Y_i, \dots, Y_{i+l-1})$, $1 \leq i \leq Q \equiv n - l + 1$. Ar $M_{l,i} = \sum_{j=i}^{i+l-1} Y_j / l$ apzīmēsim vidējo vērtību B_i^l blokā.

Ieviešam varbūtības $\{p_i\}_{i=1}^Q$ katra bloka vidējai vērtībai $\{M_{l,i}\}_{i=1}^Q$, lai iegūtu ticamības funkciju. Tad blokveida EL funkcija parametram μ ir

$$L_n(\mu) = \sup \left\{ \prod_{i=1}^Q p_i : p_i > 0, \sum_{i=1}^Q p_i = 1, \sum_{i=1}^Q p_i M_{l,i} = \mu \right\}, \quad (5.23)$$

kur mēs maksimizējam reizinājumu pa varbūtībām pie trīs nosacījumiem. Funkcija $L_n(\mu)$ ir pozitīva, kad $\mu \in (\min_{1 \leq i \leq Q} M_{l,i}, \max_{1 \leq i \leq Q} M_{l,i})$, un ārpus šī intervāla mēs definējam $L_n(\mu) = -\infty$. Kad $L_n(\mu)$ ir pozitīvs, tas reprezentē maksimālo realizāciju pie svariem $p_i = Q^{-1} \{1 + \lambda_\mu(M_{l,i} - \mu)\}^{-1}$, kur $\lambda_\mu \in \mathbb{R}$ tiek noteikts kā

$$0 = \sum_{i=1}^Q \frac{M_{l,i} - \mu}{1 + \lambda_\mu(M_{l,i} - \mu)} \equiv g_\mu(\lambda_\mu). \quad (5.24)$$

Ja vienādojumā (5.23) nebūtu ierobežojuma par vidējo vērtību μ , tad reizinājums $\prod_{i=1}^Q p_i$ būtu maksimāls, kad $p_i = Q^{-1}$, kas atbilst empīriskajam sadalījumam $\{M_{l,i}\}_{i=1}^Q$. Empīriskās ticamības attiecība vidējai vērtībai μ ir

$$R_n(\mu) = \frac{L_n(\mu)}{Q^Q} = \prod_{i=1}^Q \{1 + \lambda_\mu(M_{l,i} - \mu)\}^{-1} \quad (5.25)$$

un ticamības intervāli vidējai vērtībai tiek noteikti ar tām μ vērtībām, kurām ir relatīvi augsts EL , tas ir, intervāli ir formā

$$\{\mu : R_n(\mu) \geq A\}, \quad (5.26)$$

kur $A > 0$ ir izvēlētais ticamības līmenis.

5.2.2 Pieņēmumi

Lai pielietotu EL metodi LRD procesiem, nepieciešami sekojoši pieņēmumi.

Pieņēmumi: [31]

(P.1) Stacionāram procesam $\{Y_t\}$, $t \in \mathbb{Z}$ ir spektrālā blīvuma funkcija kā (5.21). Tādiem $0 < |d| < 1/2$ procesa autokovariācijas $r(k) = Cov(Y_t, Y_{t+k})$, $k \in \mathbb{Z}$ apmierina

$$r(k) \sim c_d R_d k^{-(1-2d)}, \quad \text{kad } k \rightarrow \infty, \quad (5.27)$$

kur c_d ir konstante no (5.21) un R_d ir kāds reāls skaitlis, kas nav nulle; ja $d = 0$, tad f ir ierobežots visā intervālā $(0, \pi]$.

(P.2) Gadījuma lielumi $\{\varepsilon_j\}$ no (5.22) ir neatkarīgi un vienādi sadalīti (*iid*).

(P.3) Kad $n \rightarrow \infty$, bloka garuma nosacījums izpildās: $l^{-1} + l^2/n = o(1)$ un ja $E(Y_t) = \mu_0$ apzīmē procesa īsto vidējo vērtību, tad standartizēti bloku vidējie apmierina

$$\left(\frac{l}{n}\right)^{1/2-d} \max_{1 \leq i \leq Q} \frac{|M_{l,i} - \mu_0|}{\sqrt{\text{Var} M_{l,i}}} \rightarrow^p 0, \quad (5.28)$$

kur $\rightarrow^p$ apzīmē konvergenci pēc varbūtības.

Lai pielāgotu EL teoriju LRD procesam, mēs apskatīsim dažas atšķirības vidējās vērtības $\bar{Y}_n = \sum_{i=1}^n Y_i/n$ uzvedībā SRD un LRD gadījumos. Teorēma 21 apraksta, kā mērogošanas parametrs a_n standartizētai izlases vidējai vērtībai $(\bar{Y}_n - \mu)/a_n$ ir atkarīgs no ilglaicīgās atkarības parametra d .

Teorēma 21. [31] Pieņemsim, ka $\{Y_t\}, t \in \mathbb{Z}$ ir process, kas apmierina pieņēmumu (P.1), $d \in (-1/2, 1/2)$.

(a) Kad $n \rightarrow \infty$, $n^{1-2d} \text{Var}(\bar{Y}_n) \rightarrow \nu_d$, kur $\nu_d = c_d R_d / \{d(1 + 2d)\}$, ja $d \neq 0$ un $\nu_0 = 2\pi c_0$, ja $d = 0$.

(b) Pieņemsim, ka $a_n^2 = n^{-1+2d} \nu_d$. Ja izpildās pieņēmums (P.2), tad $(\bar{Y}_n - \mu)/a_n \rightarrow^d Z$, kad $n \rightarrow \infty$, kur Z apzīmē standartnormālo gadījuma lielumu.

5.2.3 Blokveida EL attiecības robežsadaliņums LRD novērojumiem

Tagad aplūkosim asimptotisko sadaliņumu blokveida EL attiecībai no (5.25) vispārējiem procesiem (gan SRD , gan LRD), kas apmierina (5.21). Kā minēts Kitamura publikācijā [28], nepieciešams bloka korekcijas koeficients, lai nodrošinātu to, ka EL attiecībai ir nedeģenerēta robeža. Ar $B_n = Q^{-1}(n/l)^{1-2d}$ apzīmēsim bloka korekcijas koeficientu.

Teorēma 22. [31] Pieņemsim, ka $\{Y_t\}, t \in \mathbb{Z}$ ir process, kas apmierina pieņēmumus (P.1)-(P.3). Ja $E(Y_t) = \mu_0$ ir procesa īstā vidējā vērtība, tad, kad $n \rightarrow \infty$,

$$-2B_n \log R_n(\mu_0) \rightarrow^d \chi_1^2, \quad (5.29)$$

kur χ_1^2 apzīmē Hī-kvadrāta gadījuma lielumu ar vienu brīvības pakāpi.

Piezīme 23. *Teorēmas 22 rezultāti izpildās arī nešķeļošu bloku gadījumā : $B_{i,NOL}^l = B_{1+l(i-1)^l}$ tādiem $1 \leq i \leq Q_{NOL}$, kur $Q_{NOL} = \lfloor n/l \rfloor$. Tad bloku skaits Q tiek aizvietots ar Q_{NOL} un vienādojumā (5.25) tiek izmantotas $B_{i,NOL}^l$ bloku vidējās vērtības.*

Apgalvojums 24. [31] *Empīriskās ticamības funkcijas ticamības intervāli LRD procesu vidējai vērtībai ir formā*

$$I_n(1 - \alpha) = \left\{ \mu \in \mathbb{R} : -2\hat{B}_n \log R_n(\mu) \leq \chi_{1;1-\alpha}^2 \right\}, \quad \hat{B}_n = Q^{-1}(n/l)^{1-2d_n},$$

kur $0 < \alpha < 1$.

Tagad esam apskatījuši EL metodi divu izlašu gadījumā neatkarīgiem datiem un blokveida EL vienas izlases gadījumā ilglaicīgi atkarīgiem datiem. Pielietosim šo bloku veidošanas paņēmieni katrai no divām ilglaicīgi atkarīgām izlasēm un pielietosim apskatīto teoriju par EL metodi divu izlašu gadījumā. Jāņem vērā, ka šoreiz nosacījums par izlašu elementu neatkarību un vienādo sadalījumu nav spēkā, jo mēs aprēķināsim statistiku, nevis abām sākuma izlasēm, bet gan bloku veidotajām izlasēm $B_{l,i}$. Tātad par atbilstošo robežsadalījumu nevaram pateikt, ka tas ir χ^2 sadalījums ar atbilstošajām brīvības pakāpēm vai χ^2 sadalījums, kas izdalīts ar normalizējošo konstanti B_n . Šobrīd tāds teorētisks pētījums tam, kāds ir robežsadalījums EL metodei divu izlašu gadījumā ilglaicīgi atkarīgiem datiem, nav veikts. Tomēr aplūkosim EL statistikas asimptotisko uzvedību ar empīrisku simulāciju palīdzību.

6 Praktiskā daļa

6.1 Parametra H novērtējumi programmā R

Salīdzināsim četras Hursta parametra novērtēšanas metodes datorprogrammā **R** un noskaidrosim, kura parametru H novērtē visprecīzāk. Šim nolūkam 1000 reizes simulēsim frakcionālā Gausa trokšņa procesu ar $H = 0.6$ un $H = 0.8$ apjomā $n = 2^9$ un konstruēsim apakšizlašu metodes ticamības intervālus novērtētajiem parametriem. Apakšizlašu metodei izvēlēsimies bloka lielumu $l = 2^7$, pamatojoties uz mana diplomdarba rezultātiem, kur šāda bloka garuma izvēle deva vislabākos rezultātus (skatīt mājas lapā [21]).

Tabula 6.1: Parametra H novērtējumi dažādām metodēm un novērtējumu ticamības intervālu pārklājuma precizitātes (PP). Simulēts fGn process apjomā $n = 2^9$, bloka lielums $l = 2^7$, ticamības līmenis $1 - \alpha = 0.95$, simulācijas veiktas 1000 reižu

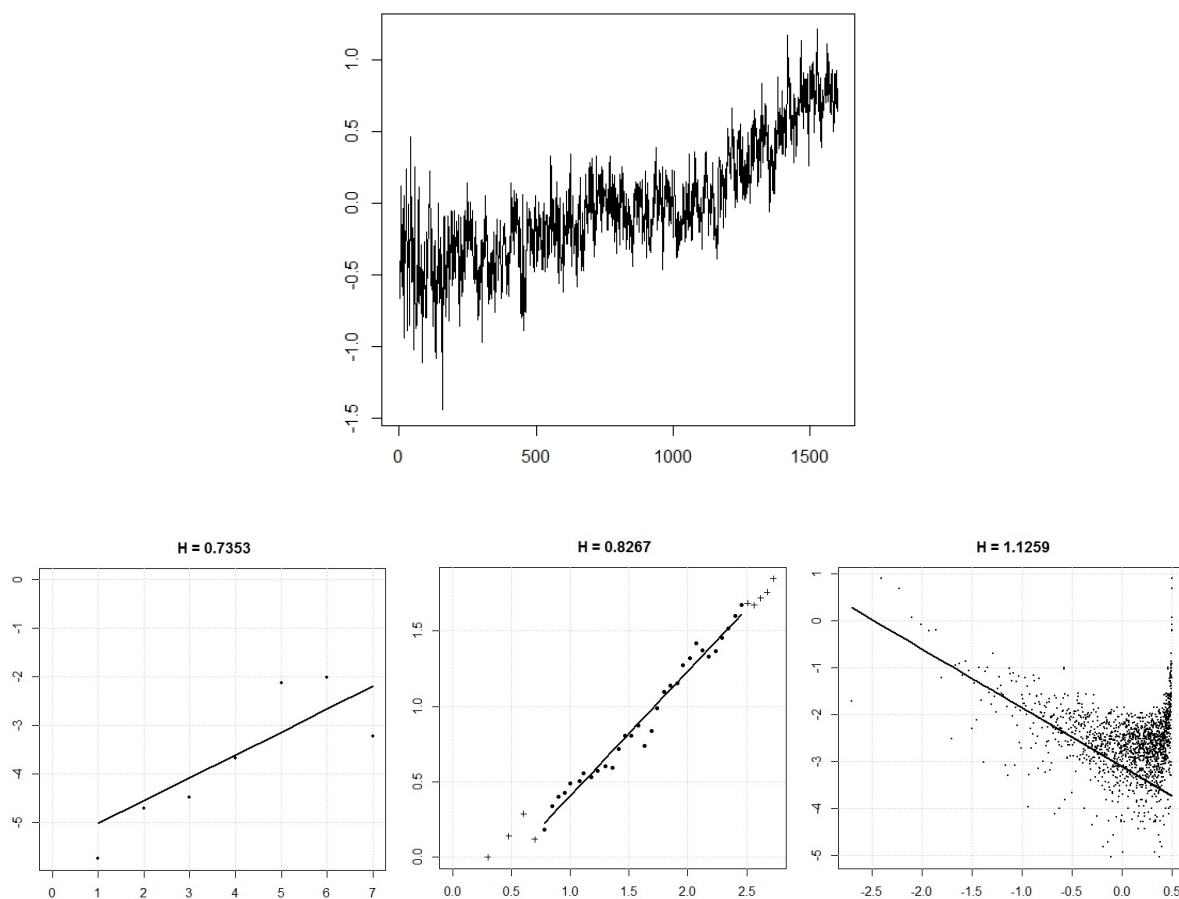
	$H = 0.6$		$H = 0.8$	
Metode	$\hat{H}$ vid.vērt.	PP	$\hat{H}$ vid.vērt.	PP
Veivletu	0.573	0.985	0.778	0.985
R/S	0.645	0.949	0.770	0.977
Period.	0.617	0.965	0.848	0.939
Kop.Var.	0.515	0.730	0.660	0.419

Kā varam redzēt 6.1 tabulā, Hursta parametra novērtēšanā kopējās variācijas metode H novērtē vissliktāk, to parāda arī zemās pārklājuma precizitātes, tātad šo metodi mēs tālāk vairs neizmantosim, lai noteiktu H . Veivletu metode novērtē H no apakšas, periodogrammas metode savukārt no augšas.

Tagad apskatīsim reālu datu piemērus.

Piemērs 3. Zemes virsmas temperatūra Celsija grādos.

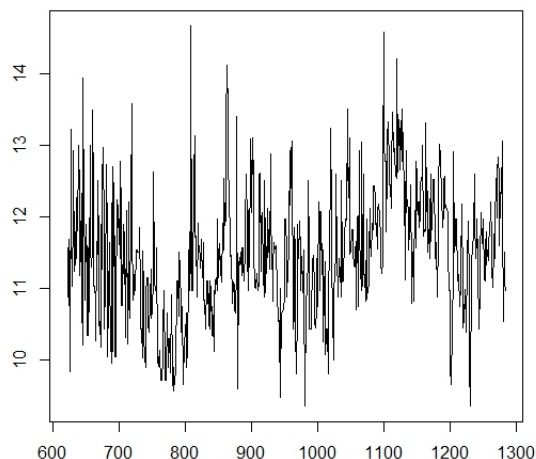
Dati par vidējo zemes virsmas temperatūru Celsija grādos no 1880. gada janvāra līdz 2013. gada aprīlim pa mēnešiem. Dati iegūti no NASA mājas lapas [32]. Kā redzam 6.1 attēlā, periodogrammas metode joprojām pārvērtē Hursta parametru, taču skatoties uz veivletu un R/S metodes novērtējumiem, H varētu būt robežās starp 0.7 un 0.85, kas liecina par ilglaicīgo atkarību.



Att. 6.1: Zemes virsmas temperatūras dati un Hursta parametra novērtējums šiem datiem ar veivletu, R/S un periodogrammas metodi

Piemērs 4. Nīlas upes dati.

6.2 attēlā dati par Nīlas upes ikgadējo minimālo plūsmu no 662. – 1284. gadam. Šie ir tie dati, kurus pētīja hidrologs Hursts (1951), ieviešot R/S statistiku. Dati pieejami **R** programmas bibliotēkā *FGN* ar nosaukumu *NileMin*.



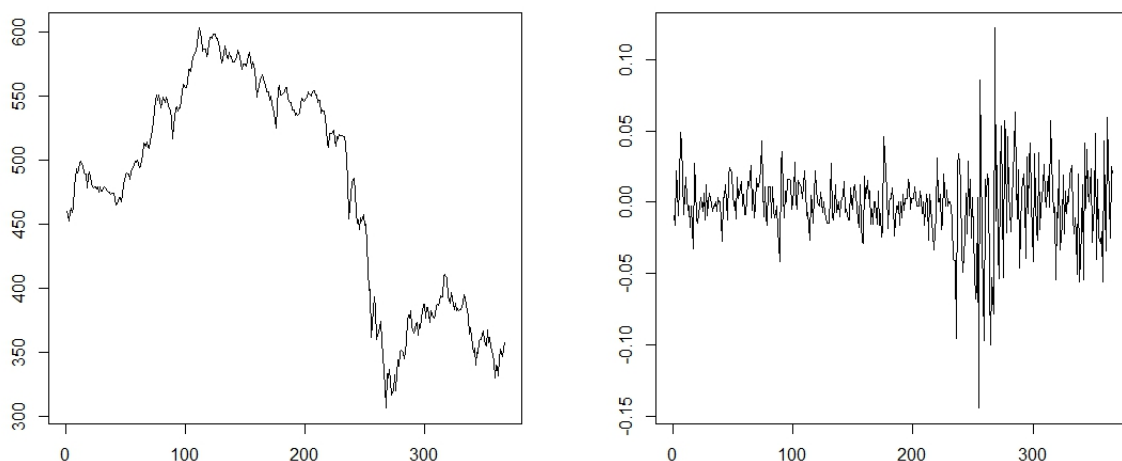
Att. 6.2: Nīlas upes plūsma no 662. – 1284. gadam

Hursta parametra novērtējums Nīlas upes datiem - veivletu metode dod $\hat{H} = 0.903$, R/S metode $\hat{H} = 0.839$ un periodogrammu metode $\hat{H} = 0.993$. Tātad, kā jau bija gaidāms, arī Nīlas upes dati ir ilglaicīgi atkarīgs process.

Piemērs 5. IBM akciju cenas.

Izmantoti dati par IBM akciju cenām dienas beigās (slēgšanas brīdī) no 17.05.1961 līdz 02.11.1962. Šie dati ņemti no publikācijas [33], kurā dati tika logaritmizēti un šo vērtību starpību procesam tika pētītas izmaiņas dispersijā. Darbā aplūkosim transformēto procesu.

Hursta parametra novērtējums transformētajiem IBM akciju cenu datiem - veivletu metode dod $\hat{H} = 0.557$, R/S metode $\hat{H} = 0.655$ un periodogrammu metode $\hat{H} = 0.572$.



Att. 6.3: IBM akciju cenas process un tā transformācija

6.2 Maiņas punkta testi ilglaicīgi atkarīgiem procesiem

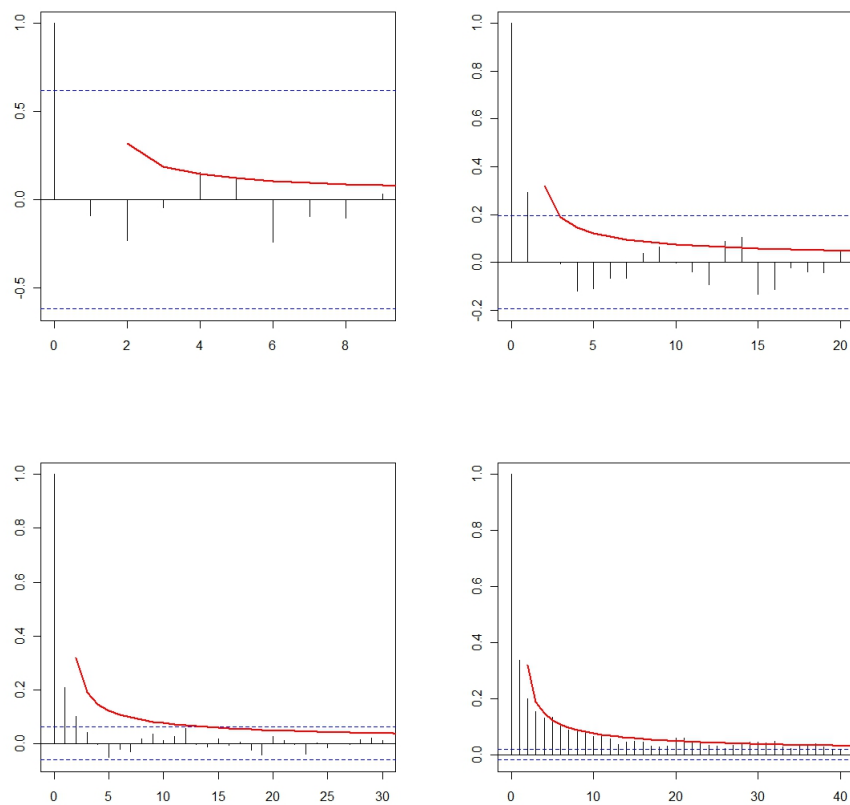
Ar simulāciju palīdzību salīdzināsim Vilksoksona tipa rangu testu (4.21) ar vidējo vērtību starpību testu (4.31). Pirmkārt, izpētīsim, kurš no abiem testiem sasniedz 5% asimptotisko līmeni, kad tos pielieto galīgas izlases gadījumā, kur izlašu apjomi mainās no $n = 10$ līdz $n = 1000$. Otrkārt, salīdzināsim abu testu jaudas izlases apjomam $n = 500$ pie dažādām alternatīvām

$$A_k : \mu_1 = \dots = \mu_k \neq \mu_{k+1} = \dots = \mu_n.$$

Ļausim gan maiņas punktam k , gan līmeņa maiņai $h := \mu_{k+1} - \mu_k$ variēt. Konkrētāk, izvēlamies $k = 25, 50, 150, 250$ un $h = 0.5, 1, 2$.

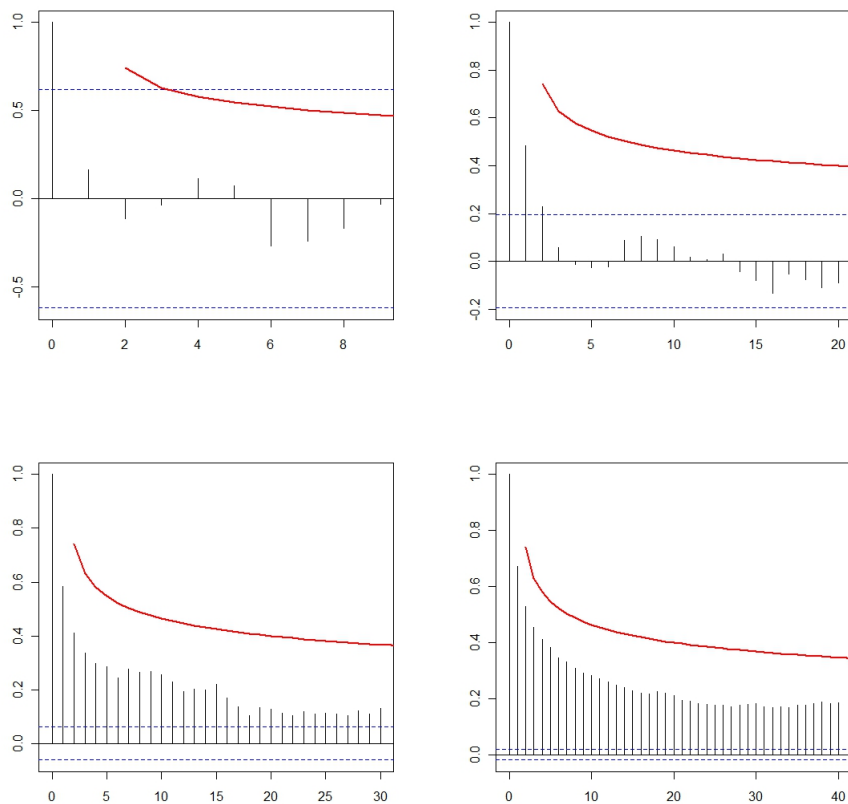
Kā pamatdatus simulācijām izvēlēsimies frakcionālā Gausa trokšņa procesu $\xi_1, \dots, \xi_n$ ar Hursta parametru H ; attiecīgi $D = 2 - 2H$. Katru simulāciju atkārtosim 10 000 reizes.

Kā redzams 6.4 attēlā, pie maziem izlases apjomiem fGn procesa autokovariāciju funkcija stipri atšķiras no teorētiskās, kas aprēķināta no (1.22). Tikai pie $n = 10000$ abas autokovariāciju funkcijas sāk sakrist. Tas nozīmē, ka priekš fGn procesa pat 1000 dati ir salīdzinoši mazs datu apjoms.

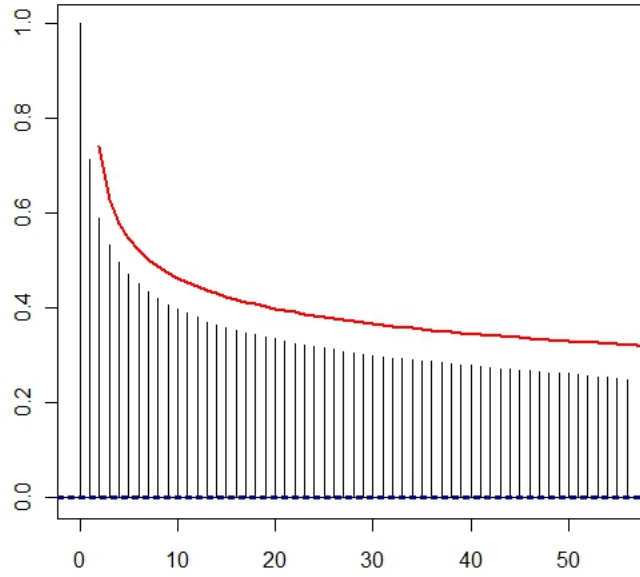


Att. 6.4: fGn procesa empīriskā un teorētiskā autokovariāciju funkcija; Hursta parametrs $H = 0.7$; izlases apjoms $n = 10, 100, 1000$ un 10000

6.5 attēlā redzams, ka gadījumā, kad fGn process simulēts ar Hursta parametru $H = 0.9$, kas nozīmē stipru atkarību starp datiem, autokovariāciju funkcijas nesakrīt arī pie izlases apjoma $n = 10000$. Gadījumu, kad $n = 500000$, skatīt 6.6 attēlā, un pat tad abas autokovariāciju funkcijas nesakrīt pie lielas atkarības struktūras.



Att. 6.5: fGn procesa empīriskā un teorētiskā autokovariāciju funkcija; Hursta parametrs $H = 0.9$; izlases apjoms $n = 10, 100, 1000$ un 10000



Att. 6.6: fGn procesa empīriskā un teorētiskā autokovariāciju funkcija; Hursta parametrs $H = 0.9$; izlases apjoms $n = 500000$

6.2.1 Normāli sadalīti dati

Simulācijas sākšim, izvēloties $G(t) = t$, tā ka $(X_i)_{i \geq 1}$ ir fGn process. Tad F ir standartnormālā gadījuma lieluma kumulatīvā sadalījuma funkcija. Lai noteiktu testu statistiku W_n un D_n kritiskās vērtības, pielietosim Teorēmu 17 un Teorēmu 18. Tā kā G ir stingri augošs, Teorēma 17 saka, ka, ja izpildās nulles hipotēze, tad W_n ir aptuveni tas pats sadalījums kā

$$\frac{1}{2\sqrt{\pi}} \sup_{0 \leq \lambda \leq 1} |Z_1(\lambda) - \lambda Z_1(1)|.$$

Tā kā G ir pirmais Hermītu polinoms, tā Hermīta rangs $m = 1$ un asociētais Hermītu koeficients $a_1 = 1$. No šejienes Teorēma 18 saka, ka, ja izpildās nulles hipotēze, tad testa statistikai D_n ir aptuveni tas pats sadalījums kā

$$\sup_{\lambda \in [0,1]} |Z_1(\lambda) - \lambda Z_1(1)|.$$

Asimptotiskās kritiskās vērtības abiem testiem, izmantojot augšējās 5%— kvantiles no sadalījuma $\sup_{\lambda \in [0,1]} |Z_1(\lambda) - \lambda Z_1(1)|$ jau aprēķinātas iepriekš un redzamas 4.1 tabulā. Tādējādi Vilkoksona tipa tests noraida nulles hipotēzi, kad $W_n \geq \frac{1}{2\sqrt{\pi}} q_\alpha$, bet vidējo vērtību starpību tests

noraidīs nulles hipotēzi, kad $D_n \geq q_\alpha$, kur q_α ir dots vienādojumā (4.24).

Tabula 6.2: Nulles hipotēzes noraidīšanas līmenis vidējo vērtību starpību testam (pa kreisi) un Vilksona rangu testam (pa labi) fGn procesam ar parametru H , simulēts 1000 reižu, abiem testiem asimptotiskais līmenis 5%

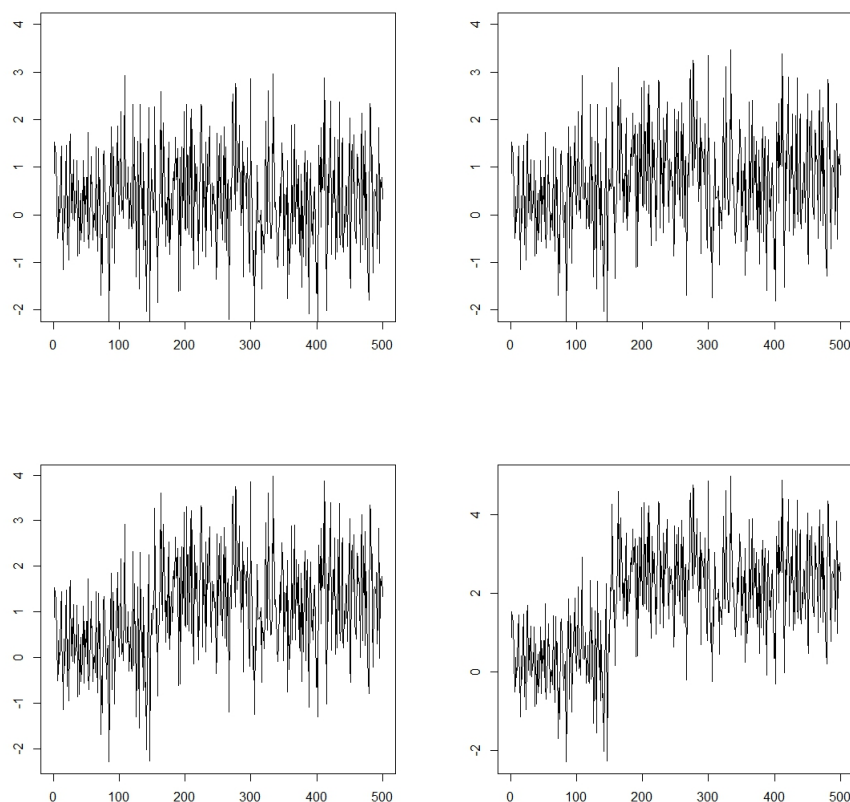
n/H	0.6	0.7	0.9	n/H	0.6	0.7	0.9
10	0.003	0.013	0.047	10	0.042	0.037	0.374
50	0.034	0.042	0.052	50	0.043	0.045	0.188
100	0.041	0.046	0.053	100	0.048	0.047	0.152
500	0.067	0.064	0.057	500	0.066	0.063	0.108
1000	0.055	0.051	0.055	1000	0.051	0.047	0.103

Tika empīriski pārbaudīts, vai abi testi sasniedz 5% līmeni, un saskaitīts, cik reizes no 10000 simulācijām testi noraida nulles hipotēzi. Kā redzams 6.2 tabulā abi testi tuvojas 5% līmenim jau pie maziem izlases apjomiem ($n = 50$), izņemot Vilksona testu gadījumā, kad $H = 0.9$, tas ir, kad ir ļoti stipra atkarība starp novērojumiem. Šajā gadījumā konverģence Teorēmā 17 ir ļoti lēna, un tāpēc galīgas izlases gadījumā asimptotiskās kritiskās vērtības netiek sasniegtas.

Lai analizētu to, cik labi testi atrod maiņas punktus, tika ieviests tāds jēdziens kā līmeņa maiņa h laika brīdī $[n\lambda]$, tas ir, apskatīta laikrinda

$$X_i = \begin{cases} \xi_i, & \text{kad } i = 1, \dots, [n\lambda] \\ \xi_i + h, & \text{kad } i = [n\lambda] + 1, \dots, n. \end{cases}$$

Konstruēsim šādas laikrindas dažādiem parametriem λ un h (piemēru skatīt 6.7 attēlā), kad izlases lielums $n = 500$. Kā var redzēt 6.3 tabulā, abi testi atrod maiņas punktus laikrindā. Jo par lielāku līmeni mēs paceļam laikrindu (lielāks h) un jo tuvāk laikrindas vidum ($\lambda = 0.5$) mēs veicam šīs izmaiņas, jo lielāka ir testu jauda. Kad maiņas punkts atrodas laikrindas vidū, abi testi dod vienādi labus rezultātus, bet, kad maiņas punkts ir tuvāk laikrindas sākumam, labāks izrādās vidējo vērtību starpības tests.



Att. 6.7: fGn process bez pārtraukumiem (augšā pa kreisi); ar lēcieni pēc 150. novērojuma ($[n\lambda]$, kur $\lambda = 0.3$) un augstums $h = 0.5$ (augšā pa labi); augstums $h = 1$ (lejā pa kreisi) un augstums $h = 2$ (lejā pa labi)

Tabula 6.3: Vidējo vērtību starpību testa (pa kreisi) un Vilkoksona rangu testa (pa labi) jauda fGn procesam ar parametru $H = 0.7$, izlases apjoms $n = 500$, dažādi pārtraukuma punkti $[n\lambda]$ un dažādiem lēcieni augstumiem h . Abiem testiem asimpotiskais līmenis 5%, simulēts 10000 reižu

h/λ	0.05	0.1	0.3	0.5
0.5	0.073	0.100	0.422	0.534
1	0.100	0.274	0.963	0.984
2	0.275	0.963	1	1

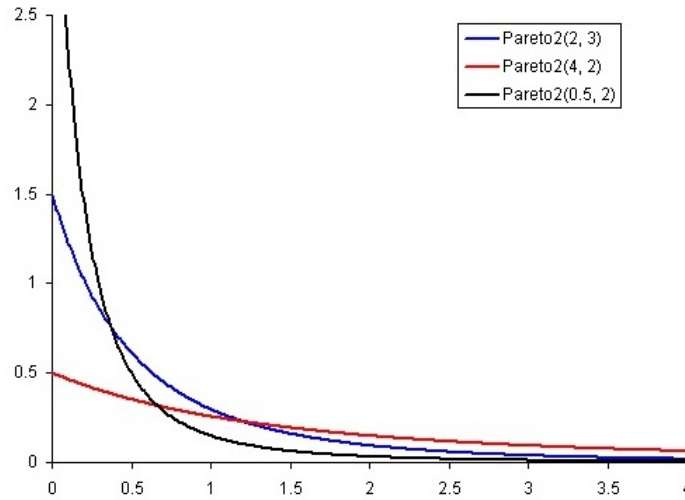
h/λ	0.05	0.1	0.3	0.5
0.5	0.073	0.099	0.420	0.527
1	0.091	0.199	0.957	0.984
2	0.129	0.761	1	1

6.2.2 Smago astu dati

Tagad apskatīsim datus, kas simulēti pēc Pareto sadalījuma [1]. $Pareto(\beta, k)$ sadalījumam ir sadalījuma funkcija

$$F(x) = \begin{cases} 1 - \left(\frac{k}{x}\right)^\beta & \text{ja } x \geq k \\ 0 & \text{citur,} \end{cases}$$

kur k ir skalas parametrs un kur β ir formas parametrs. Pareto sadalījumam ir galīga sagaidāmā vērtība, kad $\beta > 1$ un galīga dispersija, kad $\beta > 2$. Piemēru, kā izskatās dažādi Pareto sadalījumi skatīt 6.8 attēlā.



Att. 6.8: Pareto sadalījumi dažādiem parametriem β un k

Sagaidāmo vērtību un dispersiju Pareto sadalījumam aprēķina šādi:

$$\mu = E(X) = \frac{\beta k}{\beta - 1}, \quad \beta > 1,$$

$$\sigma^2 = Var(X) = \frac{\beta k^2}{(\beta - 1)^2(\beta - 2)}, \quad \beta > 2.$$

Lai iegūtu $Pareto(\beta, k)$ sadalītu $X = G(\xi)$, izvēlēsimies G kā šādu transformāciju $G(t) = k(\Phi(t))^{-1/\beta}$, kur Φ apzīmē standartnormālo sadalījuma funkciju tā, ka tādiem $x \geq k$

$$P(X_i \geq x) = P(G(\xi_i) \geq x) = P\left((\Phi(\xi_i))^{-1/\beta} \geq \frac{x}{k}\right) = P\left(\xi_i \leq \Phi^{-1}\left(\left(\frac{k}{x}\right)^\beta\right)\right) = \left(\frac{k}{x}\right)^\beta.$$

Lai X_i būtu centrēts, jāizvēlas

$$G(t) = k(\Phi(t))^{-1/\beta} - \frac{\beta k}{\beta - 1}. \quad (6.21)$$

Un, lai X_i ir standartizēts (vidējā vērtība 0 un dispersija 1), tad apskatīsim tādu $Z = (X - \mu)/\sigma$ un izvēlēsimies šādu

$$G(t) = \left(\frac{\beta k^2}{(\beta - 1)^2(\beta - 2)} \right)^{-1/2} \left(k(\Phi(t))^{-1/\beta} - \frac{\beta k}{\beta - 1} \right). \quad (6.22)$$

Tad attiecīgi Z sadalījuma funkcija ir

$$F_Z(z) = \begin{cases} 1 - \left(\frac{k}{\sigma z + \mu} \right)^\beta, & \text{ja } z \geq \frac{k - \mu}{\sigma}, \\ 0, & \text{citur,} \end{cases} \quad (6.23)$$

un tā blīvuma funkcija ir

$$f_Z(z) = \begin{cases} k^\beta \beta \sigma (\sigma z + \mu)^{-\beta-1}, & \text{ja } z \geq \frac{k - \mu}{\sigma}, \\ 0, & \text{citur.} \end{cases} \quad (6.24)$$

Pareto (3,1) dati: Tiks simulēti $Pareto(3, 1)$ dati, tas ir, smago astu dati ar galīgu dispersiju. Šajā gadījumā $\beta = 3$, $k = 1$ un $E(X) = \frac{3}{2}$, un $Var(X) = \frac{3}{4}$. Lai salīdzinātu ar simulācijām frakcionālā Gausa trokšņa datiem, dati tiks standartizēti (skatīt (6.22)):

$$G(t) = \frac{1}{\sqrt{3/4}} \left((\Phi(t))^{-1/3} - \frac{3}{2} \right).$$

Varbūtību blīvuma funkcija standartizētajam X izskatās šādi (no (6.24)):

$$f(x) = \begin{cases} 3\sqrt{\frac{3}{4}} \left(\sqrt{\frac{3}{4}}x + \frac{3}{2} \right)^{-4}, & \text{ja } x \geq -\sqrt{\frac{1}{3}}, \\ 0, & \text{citur.} \end{cases}$$

G ir stingri dilstoša funkcija, un, balstoties uz iepriekš iegūto, Hermītu rangs m funkcijai G ir 1 un $|\int_{\mathbb{R}} J_1(x) dF(x)| = (2\sqrt{\pi})^{-1}$, skatīt (4.23). Izmantojot skaitlisko integrēšanu

$$a_1 = E[\xi G(\xi)] = \sqrt{\frac{4}{3}} \int_{-\infty}^{\infty} s \Phi(s)^{-1/3} \varphi(s) ds \approx -0.6784.$$

Tabula 6.4: Nulles hipotēzes noraidīšanas līmenis vidējo vērtību starpību testam (pa kreisi) un Vilksoksona rangu testam (pa labi) $Pareto(3, 1)$ procesam, kurš iegūts no fGn procesa ar parametru H , simulēts 1000 reižu, abiem testiem asimptotiskais līmenis 5%

n/H	0.6	0.7	0.9	n/H	0.6	0.7	0.9
10	0.071	0.089	0.141	10	0.185	0.117	0.378
50	0.067	0.040	0.121	50	0.122	0.074	0.188
100	0.076	0.041	0.117	100	0.099	0.062	0.151
500	0.099	0.042	0.138	500	0.085	0.068	0.109
1000	0.109	0.046	0.112	1000	0.067	0.052	0.105

6.4 tabulā var redzēt abu testu nozīmības līmeni dažādiem izlašu apjomiem un Hursta parametriem. Redzam, ka Vilksoksona tests lēnām konverģē uz asimptotisko līmeni, palielinoties izlases apjomam. Izņēmums ir gadījums, kad $H = 0.9$, kurā abi testi neuzrāda labus rezultātus.

6.5 tabulā var redzēt abu testu jaudas pie dažādiem maiņas punktiem un to augstumiem. Smago astu datiem labāks izrādās Vilksoksona tests - situācija ir līdzīga kā normāli sadalītu datu gadījumā - jo tuvāk laikrindas vidum un jo lielākas laikrindas izmaiņas, jo lielākas testa jaudas.

Tabula 6.5: Vidējo vērtību starpību testa (pa kreisi) un Vilksoksona rangu testa (pa labi) jauda $Pareto(3, 1)$ datiem, kas iegūti no fGn procesa ar parametru $H = 0.7$, izlases apjoms $n = 500$, dažādi pārtraukuma punkti $[n\lambda]$ un dažādiem lēcienu augstumiem h . Abiem testiem asimptotiskais līmenis 5%, simulēts 10000 reižu

h/λ	0.05	0.1	0.3	0.5	h/λ	0.05	0.1	0.3	0.5
0.5	0.001	0.008	0.094	0.158	0.5	0.088	0.114	0.453	0.558
1	0.002	0.010	0.234	0.423	1	0.106	0.248	0.966	0.987
2	0.002	0.019	0.507	0.766	2	0.164	0.858	1	1

6.3 EL ilglaicīgi atkarīgiem procesiem

6.3.1 EL statistikas robežsadalījuma analīze

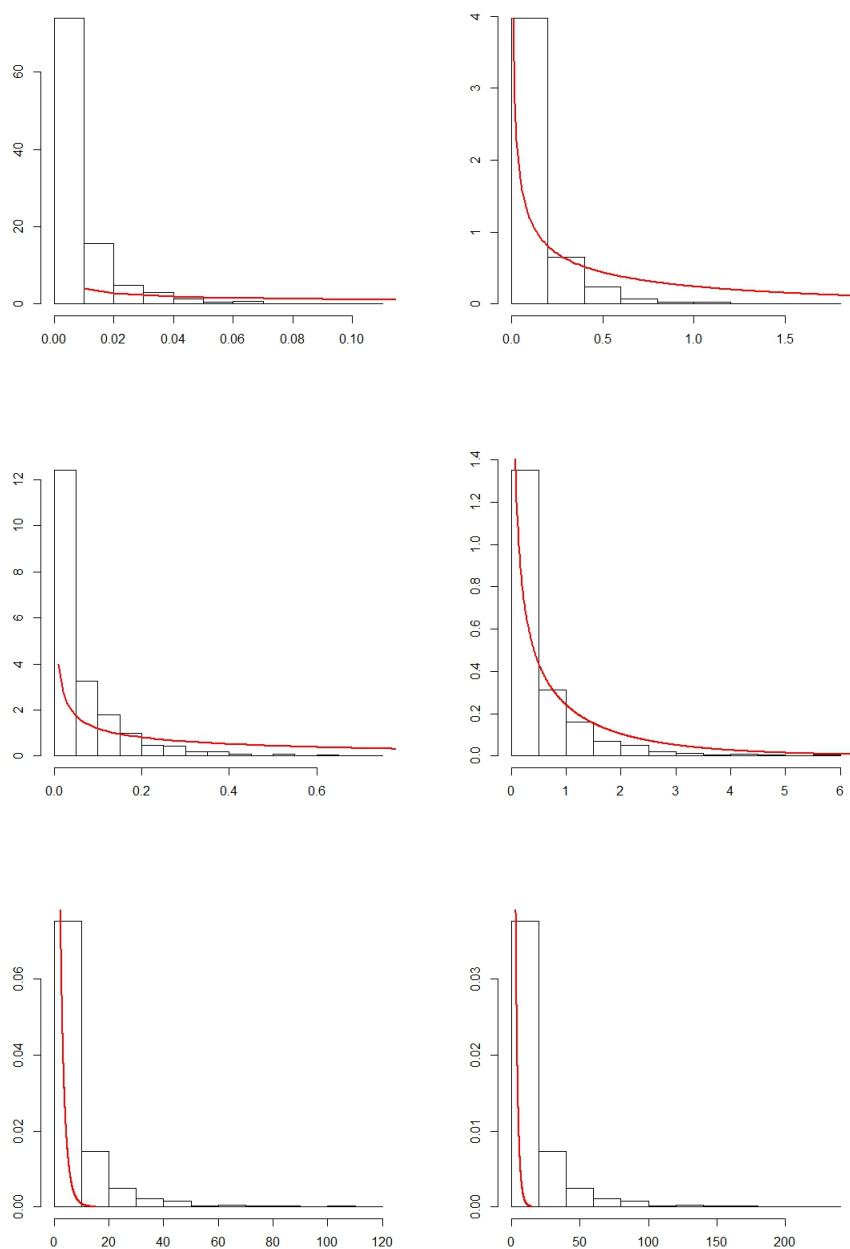
EL statistikas robežsadalījums divu izlašu gadījumā ilglaicīgās atmiņas procesiem nav izpētīts, tomēr veiksīm asimptotiskā robežsadalījuma empīrisko analīzi.

Šim nolūkam 10000 reižu tiks simulēti divi fGn procesi vienādos apjomos $n_1 = n_2 = n = 1000$ ar Hursta parametriem $H = 0.6, 0.7$ un 0.9 . Apskatām hipotēzi

$$H_0 : \mu_1 = \mu_2 \quad (6.31)$$

pret alternatīvo hipotēzi, ka vidējās vērtības abiem procesiem atšķiras. Tā kā apskatām ilglaicīgi atkarīgus procesus, nepieciešams izlases dalīt blokos. Bloku garumu izvēlēsimies $l = 2n^{0.4}$ kā minēts publikācijā [31]. Tad ar programmā **R** pieejamo funkciju *EL.means* aprēķināsim statistikas vērtības, kas tiks atspoguļotas histogrammas veidā. Robežsadalījums tiks iegūts kā vienādojumā (5.29), kur statistikas vērtībai pierēzināta konstante $B_n = Q^{-1}(n/l)^{1-2d}$ vai $\sqrt{B_n}$, kas empīriskās analīzes rezultātā šķita piemērotāka.

6.9 attēlā varam redzēt rezultātus. Kā redzams - vienas ilglaicīgi atkarīgas izlases EL statistikas robežsadalījums neder divu ilglaicīgi atkarīgu izlašu gadījumā, jo nevienā no apskatītajiem gadījumiem nevar pateikt, ka statistikas vērtības konverģētu uz χ_1^2 sadalījumu. Gadījumā, kad $H = 0.7$ izskatās, ka statistikai pierēzinātā konstante $\sqrt{B_n}$ dod mazliet labākus rezultātus nekā B_n , taču pie citām Hursta parametra vērtībām, var redzēt, ka arī šī normalizējošā konstante nebūs pareizā izvēle.



Att. 6.9: Simulētiem diviem fGn procesiem apjomā $n = 1000$ ar $H = 0.6$ (augšējie grafiki), $H = 0.7$ (vidējie grafiki) un $H = 0.9$ (apakšējie grafiki) ar EL metodes aprēķinātās statistikas vērtības un to robežsadalījumi; simulācijas veiktas 10000 reizes. Pa kreisi pierēzināta konstante B_n , pa labi - $\sqrt{B_n}$

6.3.2 *EL* metodes pārklājuma precizitāte

Kad empīriski apskatīts *EL* statistikas robežsadalījums divu izlašu gadījumā ilglaicīgi atkarīgiem novērojumiem, aplūkosim *EL* metodes pārklājuma precizitātes. Šim nolūkam simulēsim divus fGn procesus ar Hursta parametriem $H = 0.6, 0.7$ un 0.9 vienādos apjomos $n_1 = n_2 = n = 1000$. Apskatām hipotēzi H_0 kā 6.31. Ar **R** programmā *EL* bibliotēkā iebūvēto funkciju *EL.means* palīdzību aprēķināsim ticamības intervālu pārklājumu precizitātes divu izlašu vidējai vērtībai. Šeit izvēlēsimies bloku lielumus $l = c_1 n^{0.4}$, kur $c_1 \in \{1/2; 1; 2\}$.

Tabula 6.6: Nešķeļošo bloku (pa kreisi) un šķeļošo bloku (pa labi) ar atstarpi L starp blokiem ticamības intervālu pārklājuma precizitātes fGn simulētiem datiem apjomā $n = 1000$, izvēlēta precizitāte 0.95, simulācijas veiktas 1000 reizes

	$H = 0.6$	$H = 0.7$	$H = 0.9$		$H = 0.6$	$H = 0.7$	$H = 0.9$
$l = 7$	0.775	0.531	0.149	$l = 7, L = 3$	0.581	0.387	0.086
$l = 15$	0.799	0.608	0.204	$l = 15, L = 7$	0.628	0.443	0.133
$l = 31$	0.832	0.665	0.248	$l = 31, L = 15$	0.659	0.494	0.177

Kā redzams 6.6 tabulā pārklājuma precizitātes ir diezgan zemas visos gadījumos. Bloku izmēru palielināšanās dod rezultātu uzlabojumu, kas ir loģiski ilglaicīgi atkarīgiem datiem. Jo stiprāka datu atkarība (lielāks H), jo vairāk pasliktinās pārklājuma precizitātes - abu izlašu vidējās vērtības sāk stiprāk atšķirties. Arī izlašu apjoma palielināšana uz $n = 3000$ nedod nekādus uzlabojumus, skatīt 6.7 tabulā.

Tabula 6.7: Nešķeļošo bloku (pa kreisi) un šķeļošo bloku (pa labi) ar atstarpi L starp blokiem ticamības intervālu pārklājuma precizitātes fGn simulētiem datiem apjomā $n = 3000$, izvēlēta precizitāte 0.95, simulācijas 1000 reizes

	$H = 0.6$	$H = 0.7$	$H = 0.9$		$H = 0.6$	$H = 0.7$	$H = 0.9$
$l = 12$	0.725	0.449	0.148	$l = 12, L = 5$	0.519	0.298	0.100
$l = 24$	0.756	0.524	0.183	$l = 24, L = 11$	0.562	0.355	0.134
$l = 49$	0.794	0.580	0.221	$l = 49, L = 24$	0.611	0.414	0.164

6.3.3 Maiņas punkta noteikšana

Šajā nodaļā mēģināsim noteikt, kur ilglaicīgi atkarīgā laicrindā atrodas maiņas punkti. Šajā gadījumā izmantosim vienu laicrindu, kas tiks dalīta logos. Tātad kāda procesa X_T maiņas

punkta atrašanai izvēlēsimies logu

$$W_N(X_T) = (X_{T+1}, X_{T+2}, \dots, X_{T+N})$$

garumā $N = 1, 2, \dots$. Izmantojot divus blakus esošus logus $W_N(X_T)$ un $W_N(X_{T+N})$, kas nešķeļas, veiksīm hipotēžu pārbaudi par logu izveidoto laikrindu vidējo vērtību. Konstruēsim p -vērtību līknes un grafiski noteiksīm maiņas punkta atrašanās vietas - tas ir, tur, kur p -vērtību līkne pieņem vērtības, kas mazākas par nozīmības līmeni $\alpha = 0.05$, tur atrodams maiņas punkts.

Simulēsim fGn laikrindu $\{\xi_i\}$ ar Hursta parametru $H = 0.7$ apjomā $n = 1000$ un no šīs laikrindas konstruēsim procesu ar vienu maiņas punktu un ar diviem maiņas punktiem. Tātad apskatīsim

$$X_i = \begin{cases} \xi_i, & 1 \leq i \leq 500, \\ \xi_i + \mu_1, & 501 \leq i \leq 1000, \end{cases}$$

kur $\mu_1 = 3$ ir augstums, par kuru pacelta laikrinda punktā $k = 500$ (maiņas punktā), un

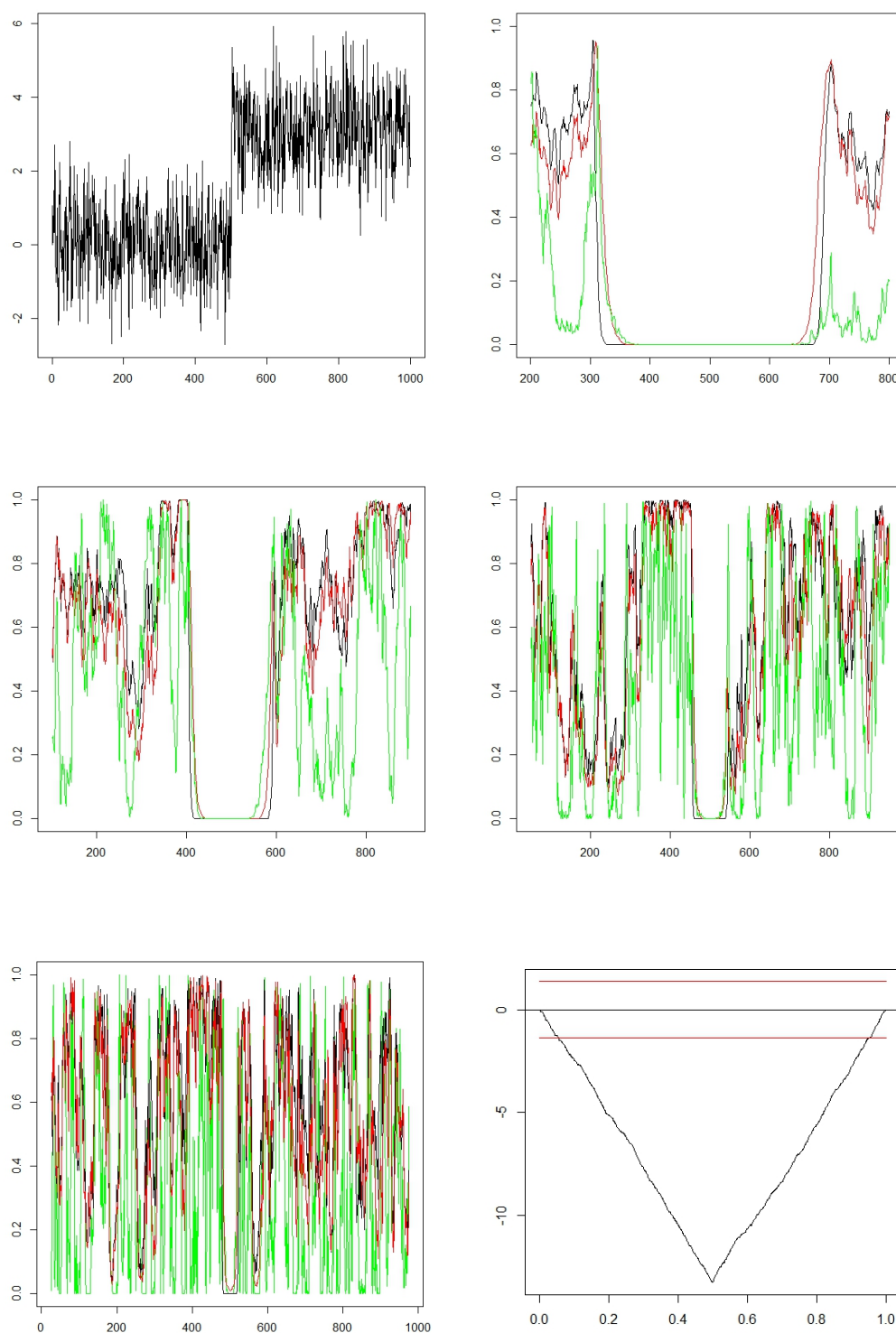
$$X_i = \begin{cases} \xi_i, & 1 \leq i \leq 350, \\ \xi_i + \mu_1, & 351 \leq i \leq 700, \\ \xi_i + \mu_2, & 701 \leq i \leq 1000, \end{cases}$$

kur $\mu_1 = 3$, $\mu_2 = 4$ un laikrindai ir divi maiņas punkti $k_1 = 350$ un $k_2 = 700$.

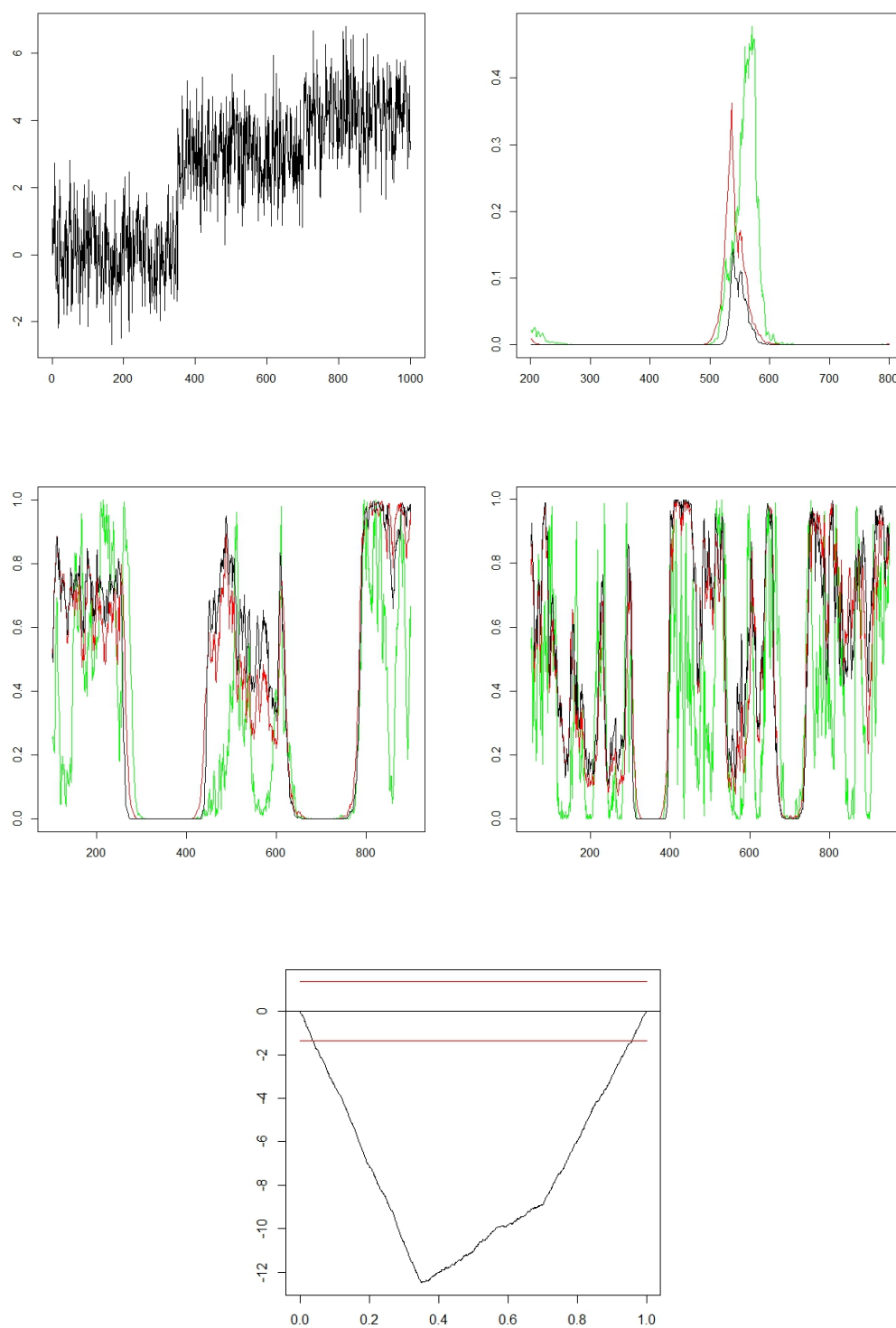
Maiņas punkta noteikšanai laikrindā konstruēsim p -vērtību līknes ar dažādiem logu garumiem jau iepriekš apskatītajiem divu izlašu rangu testiem - Vilkoksona un vidējo vērtību starpību testiem, kā arī EL metodei. Salīdzinājumam pievienosim $CUSUM$ metodi.

Kā varam redzēt 6.10 attēlā, samazinot loga garumu var aizvien precīzāk noteikt maiņas punkta atrašanās vietu, taču nevajadzētu izvēlēties arī pārāk mazu loga garumu, kā varam redzēt jau pie loga garuma $N = 25$ testi sāk uzrādīt maiņas punktus, kuri neeksistē. Salīdzinot testus savā starpā, var redzēt, ka vidējo vērtību starpību un Vilkoksona tests uzvedas līdzīgi, taču EL metode jau pie logu garuma $N = 100$ nosaka maiņas punktus, kur to nav. $CUSUM$ statistika maiņas punktu laikrindas vidū atrod.

6.11 attēlā, kur ir divi maiņas punkti interesantākais bija noskaidrot, vai tiks atrasts arī otrais maiņas punkts $k_2 = 700$, jo šeit laikrinda tika tikai mazliet pakoriģēta. Varam redzēt, ka abus maiņas punktus vidējo vērtību un Vilkoksona testu atrod pie izvēlētajiem logu garumiem, un EL metode joprojām nosaka dažus neesošus maiņas punktus, samazinoties loga garumam. Vēl varam redzēt, ka $CUSUM$ metode atrod pirmo maiņas punktu $k_1 = 350$, bet otro vairs ne tik pārliecinoši.



Att. 6.10: Laikrindas ar vienu maiņas punktu attēlojums, p –vērtību grafiki ar logu garumiem $N = 200, 100, 50$ un 25 melnā krāsā - vidējo vērtību starpību testa līknēm, sarkanā - Vilksoksona testa līknēm un zaļā - EL metodes līknēm, un $CUSUM$ grafiks



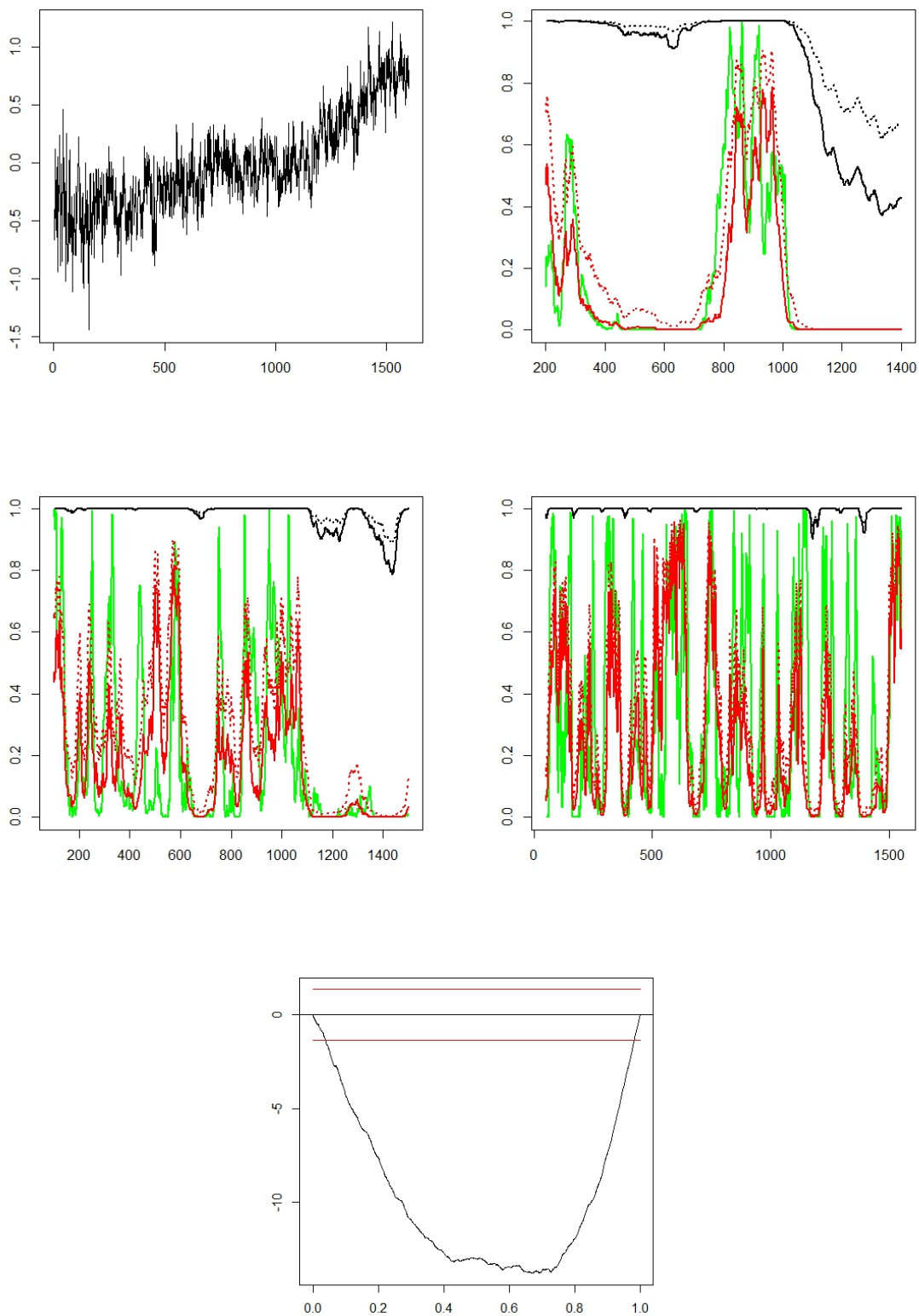
Att. 6.11: Laikrindas ar diviem maiņas punktiem attēlojums, p -vērtību grafiki ar logu garumiem $N = 200, 100$ un 50 melnā krāsā - vidējo vērtību starpību testa līknēm, sarkanā - Vilkoksona testa līknēm un zaļā - EL metodes līknēm, un $CUSUM$ grafiks

Tagad noteiksim maiņas punktus ar p –vērtību grafikiem reāliem datu piemēriem - Zemes virsmas temperatūrai, Nīlas upes datiem un IBM akciju cenu procesa transformācijai. Jāpiezīmē, ka šiem procesiem nav zināms īstais Hursta parametrs, tādēļ Vilkoksona un vidējo vērtību starpības testiem p –vērtību aprēķinam izmantosim iepriekš novērtētos Hursta parametrus ar veivletu, R/S un periodogrammas metodēm. Tos Hursta parametra novērtējumus, kas lielāki par 1 vai tuvu tam, neizmantosim.

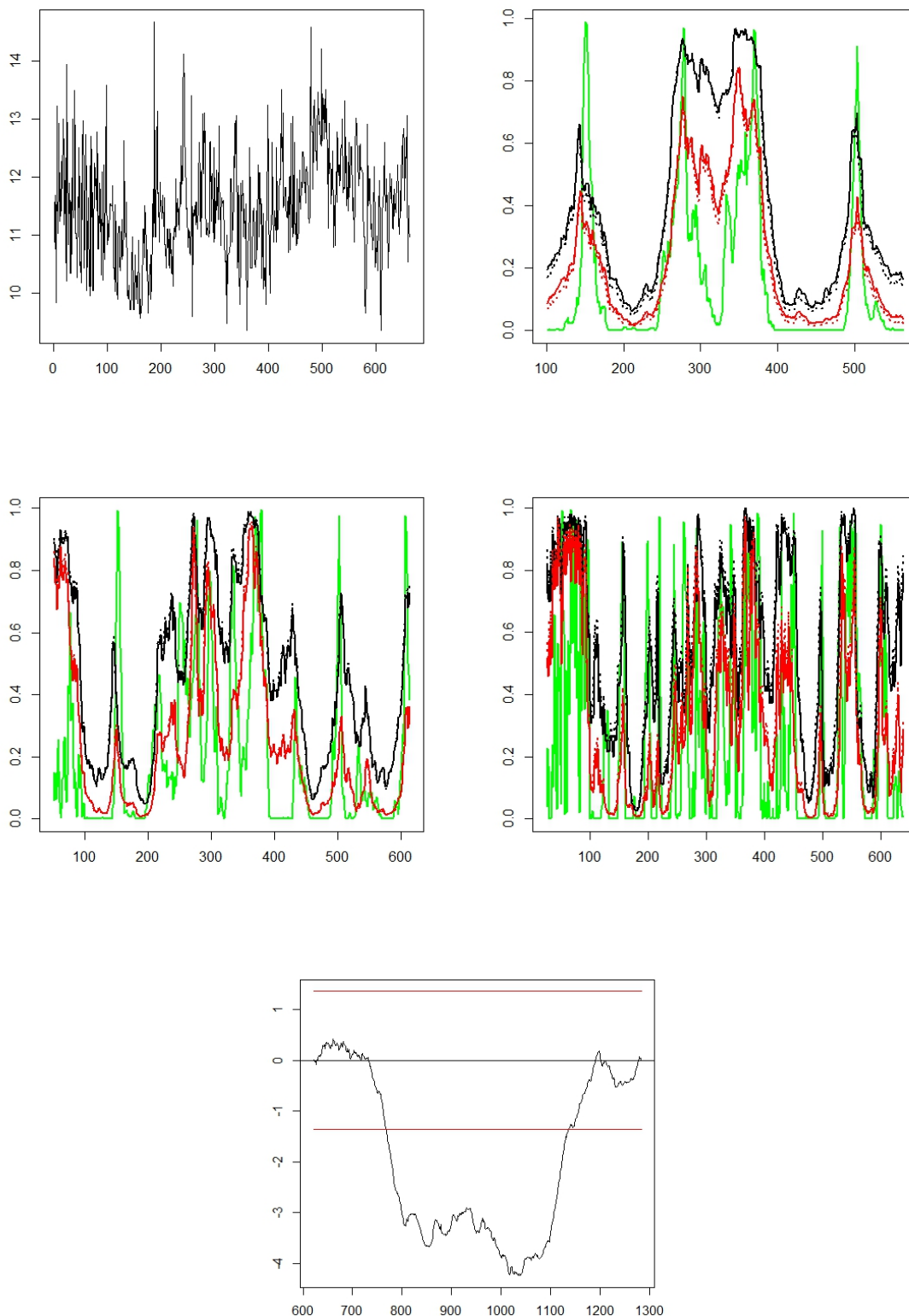
6.12.attēlā zemes virsmas temperatūras datiem loga samazināšana uzrāda aizvien jaunus maiņas punktus - ja pie loga $N = 200$ Vilkoksona tests un EL metode atrod tikai divus maiņas punktus visai laikrindai, tad pie $N = 100$ laikrindas beigās parādās jau divi maiņas punkti, bet pie $N = 50$ trīs maiņas punkti, kurus pirms tam neuzrādīja. Tas nozīmē, ka, ja vēlamies noteikt maiņas punktus, kas atrodas laikrindas sākumposmā vai beigu posmā, nepieciešams izvēlēties mazākus loga garumus, jo loga garums, kas lielāks par maiņas punkta atrašanās attālumu no laikrindas sākuma vai beigām, nespēj identificēt maiņas punktu. Redzam, ka vidējo vērtību starpību tests neatrod maiņas punktus. Parametra H izvēle būtiski nemaina rezultātus. P –vērtību līkne ar veivletu metodi novērtētajam H ir nepārtraukta, bet ar R/S metodi - raustīta. $CUSUM$ grafiks uzrāda maiņas punktus līdzīgi, kā p –vērtību līknes pie loga garuma $N = 200$.

Aplūkojot Nīlas upes datu piemēru 6.13 attēlā, var secināt, ka loga garums $N = 50$ varētu būt tuvu optimālajam, kas ir aptuveni desmitā daļa laikrindas garuma ($n = 663$). Šeit visas trīs p –vērtību līkņu metodes dod diezgan līdzīgus rezultātus. $CUSUM$ metode nosaka, ka maiņas punkti varētu būt laikrindas vidusposmā, bet tā neatrod maiņas punktus laikrindas sākumā un beigās.

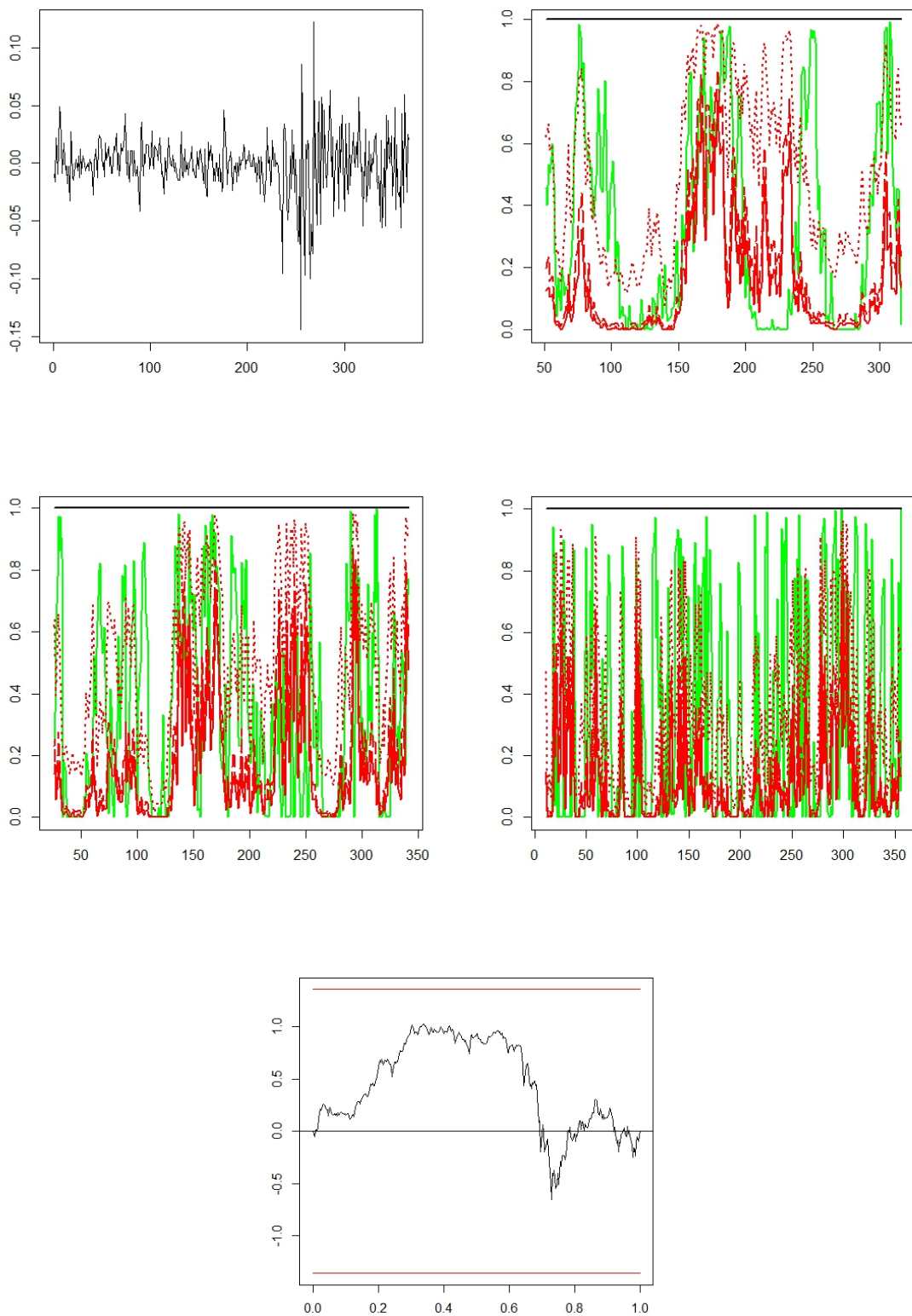
Savukārt IBM datu gadījumā 6.14 (laikrindas garums $n = 367$) loga garums $N = 50$ un $N = 25$ izskatās pieņemami. Vilkoksona testam sliktākus rezultātus dod p –vērtību līkne, kurai izmantots ar R/S metodi novērtētais Hursta parametrs $\hat{H} = 0.655$, kas ir lielāks, nekā veivletu un periodogrammas novērtējums. EL metode pie loga garuma $N = 50$ dod ticamus rezultātus.



Att. 6.12: Zemes virsmas temperatūras laikrindas attēlojums, p -vērtību grafiki logu garumiem $N = 200, 100$ un 50 melnā krāsā - vidējo vērtību starpību testa līknēm, sarkanā - Vilksoksona testa līknēm un zaļā - EL metodes līknēm, un $CUSUM$ grafiks



Att. 6.13: Nīlas upes laikrindas attēlojums, p -vērtību grafiki logu garumiem $N = 100, 50$ un 25 melnā krāsā - vidējo vērtību starpību testa līknēm, sarkanā - Vilksoksona testa līknēm un zaļā - EL metodes līknēm, un $CUSUM$ grafiks



Att. 6.14: IBM laikrindas attēlojums, p -vērtību grafiki logu garumiem $N = 50, 25$ un 10 melnā krāsā - vidējo vērtību starpību testa līknēm, sarkanā - Vilkoksona testa līknēm un zaļā - EL metodes līknēm, un $CUSUM$ grafiks

Nobeigums

Darbā aplūkoti šādi ilgaicīgās atmiņas procesi - frakcionālais Gausa troksnis (fGn), smago astu dati, kā arī trīs reālu datu piemēri. fGn procesi un smago astu process tiek simulēts programmā **R**, izmantojot dažādus Hursta parametrus H un izlašu apjomus n . fGn procesiem un reālu datu piemēriem novērtēts Hursta parametrs, izmantojot četras Hursta parametra novērtēšanas metodes. Secināts, ka Hursta parametru visprecīzāk nosaka R/S un veivletu metode. To pierāda arī ticamības intervālu pārklājuma precizitātes, kas bija augstākas tieši šīm metodēm.

Maiņas punkta noteikšanai ilglaicīgi atkarīgam procesam tika pielietoti divi neparametriskie maiņas punkta testi - Vilksoksona divu izlašu tests un vidējo vērtību starpības tests, kas ieviesti 2012. gadā [1]. Vispirms tika analizēts asimptotiskais sadalījums abu testu statistikām pie nulles hipotēzes, ka maiņas punkta nav. Šim mērķim programmā **R** tika simulēts frakcionālais Brauna process. Pielietojot abus maiņas punkta testus fGn datiem, tika secināts, ka abi testi jau pat pie maziem izlašu apjomiem ($n = 50$) tuvojas asimptotiskajam 5% līmenim, izņemot Vilksoksona testu stipras datu atkarības gadījumā ($H = 0.9$), kad asimptotiskais līmenis netika sasniegts. Analizējot abu testu jaudas maiņas punkta noteikšanā var secināt, ka, jo tuvāk laikrindas viduspunktam notikusi izmaiņa un jo lielāka tā ir, jo lielākas abu testu jaudas. Gadījumā, kad maiņas punkts atrodas tuvāk laikrindas sākumam, vidējo vērtību starpības tests izrādās labāks.

Salīdzinot šos abus testus smago astu datiem, pārāks izrādās Vilksoksona tests - testa statistika straujāk konverģē uz asimptotisko 5% līmeni, kā arī tests atpazīst maiņas punktus laikrindā. Vidējo vērtību starpību testa jaudas šiem datiem ir diezgan zemas pat, ja maiņas punkts atrodas laikrindas vidū.

Maiņas punktu atrašanai laikrindā tika konstruēti p -vērtību grafiki Vilksoksona, vidējo vērtību starpības testiem un EL metodei. Tika secināts, ka, samazinot apskatāmā loga garumu, var noteikt aizvien precīzākas maiņas punktu atrašanās vietas, taču pārāk maza loga garuma izvēle nedod nekādus rezultātus. Šajā darbā veiksmīgākā loga garuma izvēle svārstās ap $1/10$ daļu no laikrindas apjoma. Maiņas punkta noteikšanā ar p -vērtību grafikiem aplūkotā EL metode ir mazliet neprecīzāka par Vilksoksona un vidējo vērtību starpību testiem, taču izstrādājot EL divu izlašu problēmai LRD datiem, iespējams, tā uzrādītu precīzākus rezultātus. $CUSUM$ metode salīdzinājumā ar p -vērtību grafikiem nav tik efektīva, jo bieži neatpazīst maiņas punktus, ja tie ir vairāk, nekā viens.

Nobeigumā jāpiebilst, ka darbā aplūkotās metodes maiņas punkta noteikšanai ir tikai neliels

ieskats šajā jomā - Kirch (2006) disertācijā [4] minētas vēl daudzas metodes maiņas punkta noteikšanai atkarīgām datu struktūrām, piemēram, dažādas modificētas *CUSUM* un blokošanas metodes *AR*(1) procesiem, ko varētu attīstīt arī ilglaicīgi atkarīgiem datiem. Darba turpinājumā noteikti būtu nepieciešams iegūt teorētisko pamatojumu *EL* divu izlašu problēmai *LRD* datiem. Tā kā ilglaicīgi atkarīgu datu problemātikas strauji sākušas attīstīties tikai nesen, ir daudz teorētisku problēmu, kuras vēl var tikt analizētas šiem procesiem.

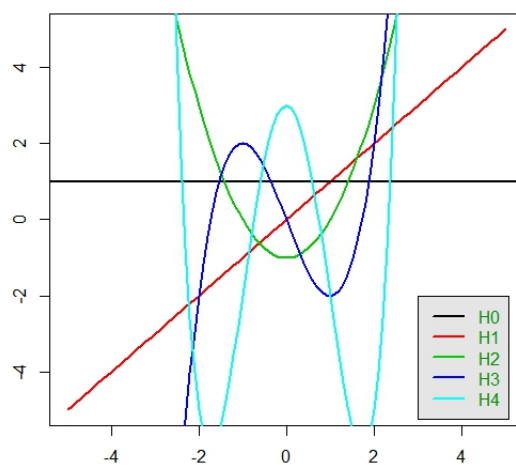
Pielikums

Definīcija 15. [34] Hermītu polinoms izskatās formā

$$He_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2} \quad (6.32)$$

Piemērs 6.

- $He_0 = 1$
- $He_1 = x$
- $He_2 = x^2 - 1$
- $He_3 = x^3 - 3x$
- $He_4 = x^4 - 6x^2 + 3$



Att. 6.15: Dažādi Hermīta polinomi

Definīcija 16. [35] Ja $\gamma(h)$ ir stacionāra procesa $\{x_t\}$ autokovariāciju funkcija un tā apmierina nevienādību

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

tad

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i \lambda h} f(\lambda) d\lambda, \quad h = 0, \pm 1, \pm 2, \dots \quad (6.33)$$

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

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \lambda(h) e^{-2\pi i \lambda h}, \quad -1/2 \leq \lambda \leq 1/2. \quad (6.34)$$

Autokovariāciju funkcija (6.33) un spektrālā blīvuma funkcija (6.34) veido *Furjē transformāciju pāri*.

Definīcija 17. [35] Ja doti dati $x_1, \dots, x_n$, tad par *diskrēto Furjē transformāciju (DFT)* sauc

$$d(\lambda_j) = n^{-1/2} \sum_{t=1}^n x_t e^{-2\pi i \lambda_j t}, \quad (6.35)$$

kad $j = 0, 1, \dots, n-1$, kur frekvences $\lambda_j = j/n$ sauc par *Furjē frekvencēm*.

Definīcija 18. [35] Dotiem datiem $x_1, \dots, x_n$ par *periodogrammu* sauc

$$I(\lambda_j) = |d(\lambda_j)|^2, \quad (6.36)$$

kad $j = 0, 1, 2, \dots, n-1$.

Apgalvojums 25. [36] *Frakcionālās Brauna kustības spektrālo blīvumu var definēt kā periodogrammas sagaidāmas vērtības robežu*

$$f_H(\lambda) \propto |\lambda|^{-2(H+1/2)}, \quad \text{kad } \lambda \rightarrow 0. \quad (6.37)$$

Definīcija 19. [15] *Diskrētos veivletu koeficientus funkcijai $x(t) \in L^2(\mathbb{R})$ aprēķina*

$$W_{j,k} = \int_{-\infty}^{+\infty} x(t) \psi_{jk}(t) dt, \quad j, k \in \mathbb{Z}, \quad (6.38)$$

kur $\psi_{jk}(t) \in L^2(\mathbb{R})$ ir *veivletu funkcijas*. Veivletu $\psi_{jk}(t)$ kopa ir atvasināta no vienas funkcijas $\psi_0(t)$, kas ir *mātes veivlets*, izmantojot *mērogošanu* un *paplašināšanos*:

$$\psi_{jk}(t) = 2^{-j/2} \psi_0(2^{-j}t - k), \quad j, k \in \mathbb{Z}, \quad (6.39)$$

kur

$$\psi_0(x) = \begin{cases} 1, & \text{ja } x \in [0, \frac{1}{2}), \\ -1, & \text{ja } x \in [\frac{1}{2}, 1), \\ 0, & \text{citur.} \end{cases} \quad (6.310)$$

Pieņēmums 1. [15] Pieņem, ka

- (a) procesam $(X_j : j \geq 1)$ ir tāda spektrālā blīvuma funkcija f , ka $\tilde{\log} f$ ir spektrālā blīvuma logaritma saistītā funkcija, kas nepārtraukta intervālā $[-\pi, \pi]$;
- (b) ψ_0 ir $m \geq 2$ zūdošie momenti ar

$$(1 + |t|)^r \{|\hat{\psi}_0(t)| + |\hat{\psi}'_0(t)| + |\hat{\psi}''_0(t)|\} \leq C_\psi \quad \forall x \in \mathbb{R},$$

kur $r \geq 2$, C_ψ ir pozitīva konstante un $\hat{\psi}_0(x) = \int \exp(-2i\pi tx) \psi_0(t) dt$ ir funkcijas ψ_0 Furjē transformācija;

- (c) $l^{-1} + n^{-1}l$ tiecas uz nulli, $j_1(n)$ tiecas uz bezgalību un $\nu_{j_1(n)}(n)/\nu_{j_1(l)}(l)$ tiecas uz nulli, ja n tiecas uz bezgalību.

Teorēma 26. (Nepārtrauktā attēlošanas teorēma) Pieņemsim, ka $f : \mathbb{R}^k \rightarrow \mathbb{R}^m$ ir funkcija, kurai $P(f(X))$ ir nepārtraukts) $= 1$. Tad, ja

- (i) $X_n \rightarrow^d X$, tad $f(X_n) \rightarrow^d f(X)$,
- (ii) $X_n \rightarrow^p X$, tad $f(X_n) \rightarrow^p f(X)$,
- (iii) $X_n \rightarrow^{g.d.} X$, tad $f(X_n) \rightarrow^{g.d.} f(X)$.

Izmantotā literatūra un avoti

- [1] M. Taqqu H. Dehling, A. Roach. Nonparametric change-point tests for long-range dependent data. *Scandinavian Journal of statistics*, 40:153–173, 2013.
- [2] G. Oppenheim P. Doukhanand and M. S. Taqqu. *Theory and applications of long-range dependence*. Birkhauser, Boston, 2003.
- [3] H. E. Hurst. *Long-term storage capacity of reservoirs*. Transactions of the American Society of Civil Engineers, 1951.
- [4] C. Kirch. Resampling methods for the change analysis of dependent data. *Inaugural-Dissertation, Univerity of Cologne*, 2006.
- [5] E. S. Page. A test for a change in a parameter occuring at an unknown point. *Biometrika*, 42:523–527, 1955.
- [6] L. Horvath M. Csorgo. Limit theorems in change-point analysis. 1997.
- [7] S. Ling. Testing for change points in time series models and limiting theorems for ned sequences. *The Annals of Statistics*, 35:1213–1227, 2007.
- [8] L. Horvath M. Reimherr A. Aue, S. Hormann. Break detection in the covariance structure of multivariate time series. *The Annals of Statistics*, 37:4046–4087, 2009.
- [9] H. Dehling D. Wied, W. Kramer. Testing for a change in correlation at an unknown point in the time using an extended functional delta method. *Econometric theory*, 2011.
- [10] C. Klieber W. Kramer, P. Sibbertsen. Long memory versus structural change in financial time series. *Allgemeines Statistisches Archiv*, 86:83–96, 2002.
- [11] A. Janic T. Ledwina J. Antoch, M. Hušková. Data driven rank tests for the cahnge point problem. *Metrika*, 68:1–15, 2008.
- [12] A. B. Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75:237–249, 1988.
- [13] A. B. Owen. Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18:90–120, 1990.
- [14] T. Dieker. *Simulation of fractional Brownian motion*. University of Twente Department of Mathematical Sciences, 2004.
- [15] S. A. Stoev P. L. Conti, L. Giovanni and M. S. Taqqu. *Confidence intervals for the long memory parameter based on wavelets and resampling*. Statistica Sirica 18, 2008.

- [16] D. R. Cox. *Long-range dependence: a review*. in Statistics: an appraisal, H.A. David and H.T. David, eds., Iowa State University Press, 1984.
- [17] J. Beran. *Statistics for Long - Memory Processes*. Chapman and Hall, New York, 1994.
- [18] <http://www.standardandpoors.com>.
- [19] M. Huskova D. Jaruskova, J. Antoch. Change point detection. *Lecture Notes of the 5th IASC Summer School*, 2000.
- [20] W. A. Taylor. Change-point analysis: A powerful new tool for detecting changes. *Taylor Enterprises, Libertyville, Illinois*, 2000.
- [21] <http://home.lu.lv/valeinis/>.
- [22] A. Rooch. *Change-point tests for Long-range dependent data*. Ruhr Universitat Bochum, 2012.
- [23] E. Lehmann. Nonparametrics: Statistical methods based on ranks. *Holden-Day, San Francisco*, 1975.
- [24] M. S. Taqqu H. Dehling. The empirical process of some long-range dependent sequences with an application to u-statistics. *The Annals of Statistics*, 17:1767–1783, 1989.
- [25] M .S. Taqqu. Convergence of integrated processes of arbitrary hermite rank. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 50:53–83, 1979.
- [26] J. Valeinis E. Cers. Extending the two-sample empirical likelihood. *preprint*, 2013.
- [27] L. Zhao Y. Qin. Empirical likelihood ratio confidence intervals for various differences of two populations. *Syst. Sci. Math. Sci.*, 13:23–30, 2000.
- [28] Y. Kitamura. Empirical likelihood methods with weakly dependent processes. *The Annals of Statistics*, 25:2084–2102, 1997.
- [29] E. Carlstein. The use of subseries methods for estimating the variance for the general statistic from a stationary time series. *The Annals of Statistics*, 14:1171–1179, 1986.
- [30] J. P. Romano D. N. Politis. On the sample variance of linear statistics derived from mixing sequences. *Stochastic Processes Application*, 45:155–167, 1993.
- [31] S. Lahiri D. Nordman, P. Sibbertsen. Empirical likelihood confidence intervals for the mean of a long-range dependent process. 2007.
- [32] <http://data.giss.nasa.gov/gistemp/tailedata v3/GLB.Ts.txt>.
- [33] Y. Wu. Inference for change-point and post-change means after a cusum test. *Lecture Notes in Statistics*, 180:31–35, 2005.
- [34] <http://uqu.edu.sa/files2/tiny mce/plugins/filemanager/files/4282164/Hermite 20polynomials.pdf>.
- [35] R. H. Shumway D .S. Stoffer. *Time Series Analysis and Its Applications 2nd Edition*. 2006.
- [36] <http://pages.stern.nyu.edu/ churvich/TimeSeries/Handouts/fBm.pdf>.

A Pielikums

A.1 Vilkoksona un vidējo vērtību starpības testu R kods

```
#### Wilkoksona statistikas iegūšana normāli sadalītiem datiem###
#####simulē FGN procesu#####
n<- 1000 # izlases apjoms
H<-0.9 # Hursta parametrs
dati<-SimulateFGN(n,H)
###teorētiskais acf FGN datiem
k<-seq(0,(n-1),by=1)
ro<-(1/2)*((k+1)^(2*H)-2*(k^(2*H))+(k-1)^(2*H))
lines(ro[0:60],col="red", lwd=2)

set.seed(1)
wilcox.fun<-function(n)
{
H<-0.9 # Hursta parametrs
dati<-SimulateFGN(n,H)
  Hermīta polinoms (m=1)
####priekš d_n
k<-seq(1,(n-1),by=1)
ro<-(1/2)*((k+1)^(2*H)-2*(k^(2*H))+(k-1)^(2*H))
summa<-c()
for (k in 1:(n-1))
{
summa[k]<-ro[k]*(n-k)
```

```

}
Var_xn<-n*var(dati)+2*sum(summa)
d_n<-sqrt(Var_xn)

##### Wilkoksona statistika

fun<-function(k,i,dati)
{
n<-length(dati)
dd<-dati[(k+1):n]
length(dd[dd>dati[i]])-(n-k)*0.5 # jābūt i no 1 līdz k
}

fun<-Vectorize(fun,vectorize.args="k")
fun<-Vectorize(fun,vectorize.args="i")
matr<-fun(1:(n-1),1:(n-1),dati)
#####summas
for (k in 1:(n-1))
{
a[k]<-sum(matr[k,(1:k)])
}
W_n<-1/(n*d_n)*max(Mod(a))
}

rez<-replicate(10000,wilcox.fun(100))
length(rez[rez>=0.1223445])/10000 ##jāmaina kvantile dažādiem H
#####

#####Vidējo vērtību starpības tests normāli sadalītiem datiem #####
#####simulē FGN procesu#####

n<- 100 # izlases apjoms
H<-0.7 # Hursta parametrs
#####H=0.9; 1000#####

```



```

vid.vert.fun<-function(n)
{
H<-0.9
dati<-SimulateFGN(n,H)

#Hermīta polinoms (m=1)
#####priekš d_n
k<-seq(1,(n-1),by=1)
ro<-(1/2)*((k+1)^(2*H)-2*(k^(2*H))+(k-1)^(2*H))
summa<-c()
for (k in 1:(n-1))
{
summa[k]<-ro[k]*(n-k)
}
Var_xn<-n*var(dati)+2*sum(summa)
d_n<-sqrt(Var_xn)

##### Vid.vērt. starpības tests

fun<-function(k,i,dati)
{
n<-length(dati)
dd<-dati[(k+1):n]
sum(dati[i]-dd)
}
fun<-Vectorize(fun,vectorize.args="k")
fun<-Vectorize(fun,vectorize.args="i")
matr<-fun(1:(n-1),1:(n-1),dati)
a<-c()
##summas
for (k in 1:(n-1))

```

```

{
a[k]<-sum(matr[k,(1:k)])
}

D_kn<-1/(n*d_n)*a
D_n<-max(Mod(D_kn))
}

rez<-replicate(10000,vid.vert.fun(1000))
length(rez[rez>=0.4337])/10000 ####jāmaina kvantile dažādiem H

#####JAUDA vidējās vērtības testam ar PARETO(3,1) datiem #####
#####h=0.5, lambda=0.05
n<-500
h<-0.5
lambda<-0.05
H<-0.7

vid.fun<-function(n)
{
H<-0.7 # Hursta parametrs
fgn<-SimulateFGN(n,H)
fgn1<-c(fgn[1:(n*lambda)],(fgn[(n*lambda+1):n]+h))
transf<-pnorm(fgn1)
beta<-3
k<-1
dati<-((((beta*(k)^2)/(((beta-1)^2)*(beta-2))))^(-1/2))*
(k*(transf)^(-1/beta)-beta*k/(beta-1))
#####Hermīta polinoms (m=1)
#####priekš d_n
k<-seq(1,(n-1),by=1)
ro<-(1/2)*((k+1)^(2*H)-2*(k^(2*H))+(k-1)^(2*H))

```

```

summa<-c()
for (k in 1:(n-1))
{
summa[k]<-ro[k]*(n-k)
}
Var_xn<-n*var(dati)+2*sum(summa)
d_n<-sqrt(Var_xn)

##### Vid.vērt. starpības statistika

fun<-function(k,i,dati)
{
n<-length(dati)
dd<-dati[(k+1):n]
sum(dati[i]-dd)
}
fun<-Vectorize(fun,vectorize.args="k")
fun<-Vectorize(fun,vectorize.args="i")
matr<-fun(1:(n-1),1:(n-1),dati)
a<-c()
##summas
for (k in 1:(n-1))
{
a[k]<-sum(matr[k,(1:k)])
}
D_kn<-1/(n*d_n)*a
D_n<-max(Mod(D_kn))

}
a1<-0.6784
rez<-replicate(10000,vid.fun(500))
length(rez[rez>=(0.87329*a1)])/10000 ###mainīt dažādiem H

```

```
#####
```

A.2 p -vērtību grafiki maiņas punkta noteikšanai

```
##### p-vērtību grafiki Nīlas upes datiem
```

```
library(EL)
```

```
library(FGN)
```

```
#####Nīlas upes dati####
```

```
data(NileMin)
```

```
data<-NileMin
```

```
n<-length(data)
```

```
#EL , kur H neko nenosaka#####
```

```
#2 izlašu tests atkarīgiem datiem - izmantojot data blocking
```

```
for (k in 10:10){
```

```
N<-k*10; #datu apjoms logos 10, 25, 50, 100, 200
```

```
l<-trunc(2*N^(2/5)) #bloka garums
```

```
L<-l #bloku atstarpe
```

```
Q<-trunc(((N-1)/L)+1) #bloku skaits
```

```
blocking<-function(X.data, Y.data)
```

```
{
```

```
X.block<-c()
```

```
mu1=mean(X.data)
```

```
#for(i in 1:Q) X.block[i]<- mean(X.data[((i-1)*L+1):((i-1)*L+1)]-mu1)
```

```
for(i in 1:Q) X.block[i]<- mean(X.data[((i-1)*L+1):((i-1)*L+1)])
```

```
Y.block<-c()
```

```

mu2=mean(Y.data)
#for(i in 1:Q) Y.block[i]<-mean(Y.data[((i-1)*L+1):((i-1)*L+1)]-mu1-delta0)
for(i in 1:Q) Y.block[i]<-mean(Y.data[((i-1)*L+1):((i-1)*L+1)])
#EL.means(X.data,Y.data)$p.value
EL.means(X.block,Y.block)$p.value
}
pp.values<-c()

for (i in 1:(n-2*N)){
F1.block<-data[i:(i+N-1)];
F2.block<-data[(i+N):(i+2*N-1)];
#mean(F1.block);mean(F2.block);
pp.values[i]<-blocking(F1.block,F2.block);
}
plot((N+1):(length(pp.values)+N),pp.values, type='lty',xlab="",ylab="",
col="green",lwd=2)
pp=round(pp.values,2)
}

#####
### p-vērtības Vidējo vērtību starpības testam###
library(fArma)
library(FGN)

#####Z(lambda)-lambda(Z(1))#####
#H<=0.9
set.seed(1)
i<-seq(1,1000, by=1)

fun<-function(n)
{
BM<-fbmSim(n = 1000, H = 0.9, method = c("mvn", "chol", "lev", "circ", "wave"),waveJ = 7

```

```

M<-max(Mod(BM[i]-(i/1000)*BM[1000]))
}

rez<-replicate(1000,fun(n))

##### H= 0.9#####
for (k in 10:10){
N<-k*10;    #datu apjoms logos 10, 25, 50, 100, 200

###p-vērt iegūšana no Vidējo vērt.starpības testa
set.seed(1)
vid.vert.fun<-function(X.dati,Y.dati)
{
dati<-c(X.dati,Y.dati)
n<-length(dati)
H<-0.9
##### Hermīta polinoms#####priekš d_n
k<-seq(1,(n-1),by=1)
ro<-(1/2)*((k+1)^(2*H)-2*(k^(2*H))+(k-1)^(2*H))
summa<-c()
for (k in 1:(n-1))
{
summa[k]<-ro[k]*(n-k)
}
Var_xn<-n*var(dati)+2*sum(summa)
d_n<-sqrt(Var_xn)

##### Vid.vērt. starpības statistika
fun<-function(k,i,dati)
{
n<-length(dati)
dd<-dati[(k+1):n]

```

```

sum(dati[i]-dd)
}
fun<-Vectorize(fun,vectorize.args="k")
fun<-Vectorize(fun,vectorize.args="i")
matr<-fun(1:(n-1),1:(n-1),dati)
a<-c()
##summas
for (k in 1:(n-1))
{
a[k]<-sum(matr[k,(1:k)])
}
D_kn<-1/(n*d_n)*a
D_n<-max(Mod(D_kn))
length(rez[rez>=D_n])/1000 #p-vērtības
}

pp.values<-c()

###sadala pa logiem
for (i in 1:(n-2*N)){
F1.block<-data[i:(i+N-1)];
F2.block<-data[(i+N):(i+2*N-1)];
pp.values[i]<-vid.vert.fun(F1.block,F2.block);
}

lines((N+1):(length(pp.values)+N),pp.values, type='lty',xlab="",ylab="",
ylim=c(0,1),lwd=2)
pp=round(pp.values,2)
}

##### H=0.9#####
### p-vērtības Wilcoxon testam###

```

```

####Izmantojam REZ no vid.vērt starpības testa!!
rezz<-rez*(1/(2*sqrt(pi)))

for (k in 10:10){
N<-k*10;    #datu apjoms logos 10, 25, 50, 100, 200

###p-vērt iegūšana no Vidējo vērt.starpības testa
set.seed(1)
wilcox.fun<-function(X.dati,Y.dati)
{
dati<-c(X.dati,Y.dati)
n<-length(dati)
H<-0.9
##### Hermīta polinoms#####priekš d_n
k<-seq(1,(n-1),by=1)
ro<-(1/2)*((k+1)^(2*H)-2*(k^(2*H))+(k-1)^(2*H))
summa<-c()
for (k in 1:(n-1))
{
summa[k]<-ro[k]*(n-k)
}
Var_xn<-n*var(dati)+2*sum(summa)
d_n<-sqrt(Var_xn)

##### Wilkoksona statistika

fun<-function(k,i,dati)
{
n<-length(dati)
dd<-dati[(k+1):n]
length(dd[dd>dati[i]])-(n-k)*0.5 # jābūt i no 1 līdz k

```



```

}

fun<-Vectorize(fun,vectorize.args="k")
fun<-Vectorize(fun,vectorize.args="i")
matr<-fun(1:(n-1),1:(n-1),dati)

#####summas
for (k in 1:(n-1))
{
a[k]<-sum(matr[k,(1:k)])
}
W_n<-1/(n*d_n)*max(Mod(a))
length(rezz[rezz>=W_n])/1000 #p-vērtības
}

pp.values<-c()

###sadala pa logiem
for (i in 1:(n-2*N)){
F1.block<-data[i:(i+N-1)];
F2.block<-data[(i+N):(i+2*N-1)];
pp.values[i]<-wilcox.fun(F1.block,F2.block);
}

lines((N+1):(length(pp.values)+N),pp.values, type='lty',xlab="",ylab="",
col="red",lwd=2)
pp=round(pp.values,2)
}

```

Maģistra darbs “Neparametriskie maiņas punkta testi ilglaicīgi atkarīgiem novērojumiem”
izstrādāts LU Fizikas un Matemātikas fakultātē.

Ar savu parakstu apliecinu, ka pētījums veikts patstāvīgi, izmantoti tikai tajā norādītie informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai.

Autors: Ieva Dasmane

(paraksts)

(datums)

Rekomendēju darbu aizstāvēšanai.

Vadītājs: doc. Dr.math. Jānis Valeinis

(paraksts)

(datums)

Recenzents: Nadežda Siņenko

(paraksts)

(datums)

Darbs iesniegts Matemātikas nodaļā _____

(datums)

(darbu pieņēma)

Darbs aizstāvēts maģistra gala pārbaudījuma komisijas sēdē

_____ prot. Nr. _____, vērtējums _____

(datums)

Komisijas sekretārs/-e: _____

(Vārds, Uzvārds)

(paraksts)