

LATVIJAS UNIVERSITĀTE
FIZIKAS UN MATEMĀTIKAS FAKULTĀTE
MATEMĀTIKAS NODAĻA

ILGLAICĪGĀS ATMIŅAS PROCESI

DIPLOMDARBS

Autors: **Ilze Rēvalde**

Stud. apl. ir05016

Darba vadītājs: doc. Dr.math. Jānis Valeinis

RĪGA 2010

Anotācija

Darba tēma ir ilglaicīgās atmiņas procesi. Apskatīta šī jēdziena izcelsme un aprakstošie modeļi. Definēts Hursta fenomens un frakcionālā Brauna kustība. Aplūkota ilglaicīgās atmiņas modeļu klase FARIMA, tās svarīgākās īpašības un parametru novērtēšanas metodes. Programmā R simulēti ilglaicīgās atmiņas procesi FARIMA un izmantotas dažādas metodes parametra H novērtēšanai.

Atslēgas vārdi: FARIMA, Hursta parametrs, ilglaicīgā atmiņa

Abstract

The subject of this thesis is long - memory processes. The origin of long memory notion and descriptive models are examined. Hurst phenomenon and fractional Brownian motion are defined. Long memory models FARIMA, the most important properties and estimation methods are considered. Long memory processes FARIMA with program R are simulated and various methods are applied to estimate the parameter H .

Keywords: FARIMA, Hurst parameter, long memory processes

Saturs

Apzīmējumi	3
Ievads	4
1. Hursta fenomens	6
2. Ilglaicīgās atmiņas definīcijas	9
3. Frakcionālā Brauna kustība un frakcionālais Gausa troksnis	12
4. FARIMA modeļi	15
4.1. Bezgalīga AR un MA reprezentācija	16
4.2. Spektrālais blīvums	16
4.3. Autokovariāciju funkcija	17
4.4. Izlases vidējā vērtība	17
4.5. Parciālās autokorelācijas	18
5. Parametra H novērtēšanas metodes	19
5.1. R/S statistika	19
5.2. Dispersiju grafiki	20
5.3. Regresiju metode	21
5.4. Detrendētā svārstību metode	22
6. Vislielākās ticamības parametru novērtēšanas metodes	23
6.1. Gausa vislielākās ticamības novērtējums	24
6.2. Whittle tuvinātais vislielākās ticamības novērtējums	25
6.3. Tuvinātie autoregresīvie vislielākās ticamības novērtējumi	26
6.4. Diskrēts Whittle novērtējums	27
7. Praktiskais pielietojums	29
Secinājumi	38
Izmantotā literatūra un avoti	39
A Pielikums	41

Apzīmējumi

ACF autokorelāciju funkcija

AR autoregresīvs process

ARMA autoregresīvs slīdošā vidējā process

$(B(t))$ Brauna kustība

$\mathcal{C}[\cdot, \cdot]$ nepārtrauktu funkciju telpa

$D[0, 1]$ no labās puses nepārtrauktu, no kreisās puses ierobežotu funkciju telpa

FARIMA frakcionāls autoregresīvs slīdošā vidējā process

fBm frakcionālā Brauna kustība

FCLT funkcionālā centrālā robežteorēma jeb Donskera teorēma

fGn frakcionālais Gausa troksnis

FI frakcionāli integrēts

FN frakcionāls troksnis

H Hursta parametrs

LRD ilglaicīgā atmiņa

MA slīdošā - vidējā process

MLE vislielākas ticamības novērtējums

PACF parciālā autokorelāciju funkcija

R/S statistika

$\text{Var}(X)$ dispersija no X

\Rightarrow konverģē

Ievads

Darbā aplūkoti ilglaicīgās atmiņas modeļi un tos raksturojošie parametri. Pēdējo dekāžu laikā, kopš atklāta ilglaicīgās atmiņas iedarbība, tā ieņem būtisku lomu laikrindu analīzē. Mūsdienās daudzus procesus var aprakstīt tieši ar šiem modeļiem. Pielietojumi ilglaicīgās atmiņas procesiem atrodami makroekonomikā, hidroloģijā, fizikā, astronomijā, telekomunikācijās, agronomijā un finanšu jomā. Procesus raksturo ļoti lēna korelāciju dilšana. Tas ietekmē novērtējumu un prognožu iegūšanas procesus, līdz ar to nepieciešams lietot atbilstošus modeļus.

Termins “ilglaicīgā atmiņa” kļuva aktuāls ap 1960.gadu. 1906.gadā Kairā ieradās Harolds Edvīns Hursts, jauns angļu ierēdnis, kurš darbojās kā hidroloģiskais konsultants. Viņš pētīja sakarības Nīlas ūdens līmeņa svārstībās. Rezultātus publicēja apjomīgos krājumos “Nile Basin” [1], kuru pielikumos bija dokumentēts nokrišņu daudzums, upes līmeņa un straumes izmaiņas 5 gadu ietvaros. Pēdējā no Hursta sējumiem - “The Future Conservation of the Nile” (1946) [2], tikai ierosināts ikgadējos ūdenstilpņu datus “uzglabāt” plašā, ilglaicīgā fondā, kuru viņš nodēvēja par “gadsimta atmiņu”. Tās koncepts balstās uz saistību starp maksimālo uzkrāto novirzi no vidējās vērtības, standarta novirzi un novērojumu periodu, to aprakstīja ar statistiku R/S . Izrādījās, ka praksē datu uzvedība atšķīrās no teorētiski paredzamās, to dilšanas pakāpe H bija 0.74 sagaidāmās 0.5 vietā. Šī parādība kļuva pazīstama kā *Hursta fenomēns*. Tika noskaidrots, ka svarīgākais aspekts ir neparasti lēnā korelāciju dilšana jeb neparasti ilgstoša atmiņa. Līdz ar to, bija nepieciešami jauni modeļi, kas atbilstu situācijai.

Stacionārās laikrindās ilglaicīgā atmiņa jeb ilglaicīgā atkarība (turpmāk apzīmēsim - LRD) izpaužas, kā korelāciju tiekšanās uz nulli kā pakāpes funkcijai, turklāt tik lēni, ka to summas diverģē. Salīdzinājumā - laikrindu modeļiem ARMA korelācijas dilst eksponenciāli. Ilglaicīgās atmiņas pakāpi raksturo parametrs H . Tas pieņem vērtības $0 < H < \frac{1}{2}$. Pēc definīcijas, procesam piemīt LRD, ja $\frac{1}{2} > H > 1$. Darba mērķis ir

- parametra H novērtēšana,
- piemērota FARIMA modeļa atrašana.

Darbā lielākoties izmantots materiāls no J. Beran grāmatas “Statistics for Long - Memory Processes” [3]. Autors tiek daudz citēts dažādos LRD modeļu pētījumos, kā arī daļa no aprēķinos izmantotajām, datorprogrammas R ilglaicīgās atmiņas procesu paketes “fArma”

komandām izveidotas pēc viņa algoritma.

Pirmajā nodaļā aprakstīta Hursta piedāvātā statistika un tās sākotnējais pierādījums, kas nesakrīta ar novērotajiem rezultātiem. Šī iemesla dēļ veica dažādus pētījumus, lai noskaidrotu, kāpēc pierādījums nav spēkā, līdz atklājās lēnās korelāciju dilšanas ietekme. Otrajā nodaļā apskatītas trīs ilglaicīgās atmiņas definīcijas un pierādīta to ekvivalence. Trešajā nodaļā aprakstīti procesi, kurus var raksturot ar Hursta parametru - frakcionālā Brauna kustība un frakcionālais Gausa troksnis. Ceturtā nodaļa ir ilglaicīgās atmiņas modeļu klases FARIMA apraksts. Definētas autokorelāciju un parciālo autokorelāciju funkcijas, spektrālais blīvums un citi raksturlielumi. Piektajā nodaļā dotas heuristiskas parametra H novērtēšanas metodes - R/S statistika, dispersiju grafiki un regresiju metode spektrālā aspektā. Sestajā nodaļā aprakstītas Whittle, Gausa un tuvinātā autoregresīvā vislielākās ticamības parametru novērtēšanas metodes. Septītā nodaļa ir praktisks R/S statistikas, dispersiju grafika un citu metožu pielietojums dažādiem datiem ar ilglaicīgo atmiņu. Pielikumā definēti darbā ietvertie, nepaskaidrotie termini un teorēmas, kā arī atrodams izmantoto programmu kods.

1. Hursta fenomens

Apskatīsim Hursta fenomena teorētisko aspektu. Statistika, kuru Harolds Hursts piemēroja Nilas datiem:

$$\frac{R}{S}(X_1, \dots, X_n) = \frac{\max_{0 \leq i \leq n} \left(S_i - \frac{i}{n} S_n \right) - \min_{0 \leq i \leq n} \left(S_i - \frac{i}{n} S_n \right)}{\left(\frac{1}{n} \sum_{i=1}^n \left(X_i - \frac{1}{n} S_n \right)^2 \right)^{\frac{1}{2}}}, \quad (1.1)$$

kur X_1, X_2, \dots, X_n ir izlase no n novērojumiem; $S_m = X_1 + \dots + X_m$ ir parciālo summu virkne, visiem $m = 0, 1, \dots$ un $S_0 = 0$. $\frac{S_n}{n}$ ir izlases vidējā vērtība; savukārt starpība $\left(S_i - \frac{i}{n} S_n \right)$ parāda, par cik parciālās summas pārsniedz to izlases vidējo vērtību; maksimuma un minimuma starpība skaitītājā ir starpība starp zemāko un augstāko parciālo summu pozīciju; dalītājs ir izlases standarta novirze.

Analizēsim statistikas (1.1) uzvedību, izmantojot materiālu no G. Samorodnitski grāmatas "Long Range Dependence" [4]. Vispirms definēsim teorēmu, kas būs nepieciešama pierādījumam.

Definīcija 1. [5] Stohastisku procesu no reālām vērtībām $\{B(t) : t \geq 0\}$ sauc par *Brauna kustību* ar sākumu punktā $x \in \mathbb{R}$, ja izpildās

1. $B(0) = x$,
2. procesam eksistē neatkarīgi pieaugumi, t.i., visiem $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$ pieaugumi $B(t_n) - B(t_{n-1}), B(t_{n-1}) - B(t_{n-2}), \dots, B(2) - B(1)$ ir neatkarīgi gadījuma lielumi,
3. visiem $t \geq 0$ un $h > 0$ pieaugumi $B(t+h) - B(t)$ ir sadalīti normāli ar matemātisko cerību nulle un dispersiju h ,
4. funkcija $t \mapsto B(t)$ ir nepārtraukta gandrīz visur.

Funkciju $\{B(t) : t \geq 0\}$ sauc par standarta Brauna kustību, ja $x = 0$.

Pieņemsim, ka virkne $\{X_n : n \geq 0\}$ ir neatkarīgi, vienādi sadalīti gadījuma lielumi ar $E(X_n) = 0$ un $\text{Var}(X_n) = 1$. Gadījuma klejošana definēta ar $S_n = \sum_{k=1}^n X_k$ ir lineāri interpolējama veselās daļās, t.i.,

$$S(t) = S_{[t]} + (t - [t])(S_{[t]+1} - S_{[t]}).$$

Tad ir uzdots gadījuma funkcija $S \in \mathcal{C}[0, \infty)$. Definējam gadījuma funkciju virkni $\{S_n^* : n \geq 1\}$ apgalabā $\mathcal{C}[0, 1]$ ar

$$S_n^*(t) = \frac{S(nt)}{\sqrt{n}}$$

visiem $t \in [0, 1]$.

Teorēma 1. *Donskera invariācijas princips (Funkcionālā centrālā robežteorēma).* [5] *Nepārtrauktu funkciju telpā $\mathcal{C}[0, 1]$ virkne $\{S_n^* : n \geq 1\}$ pēc sadalījuma konverģē uz standarta Brauna kustību $\{B(t) : t \in [0, 1]\}$*

Teorēmas pierādījumu skat. [5].

Pieņemsim, ka funkciju $f : D[0, 1] \mapsto \mathbb{R}$ definē kā

$$f(x) = \sup_{0 \leq t \leq 1} (x(t) - tx(1)) - \inf_{0 \leq t \leq 1} (x(t) - tx(1)).$$

Redzams, ka $x = (x(t), 0 \leq t \leq 1) \in D[0, 1]$ ir nepārtraukta. Pielietosim šo funkciju parciālo summu virknei ar vērtībām no apgalaba $D[0, 1]$ - parciālo summu procesam. Pieņem, ka X_1, X_2, \dots ir stacionāra gadījumu lielumu virkne ar ierobežotu dispersiju un vidējo vērtību μ . Parciālo summu procesu definē

$$S^{(n)}(t) = S_{[nt]} - [nt]\mu,$$

visiem $0 \leq t \leq 1$. Pēc FCLT varam secināt, ka, ja X_1, X_2, \dots ir neatkarīgi, vienādi sadalīti gadījuma lielumi, tad

$$\frac{1}{\sqrt{n}} S^{(n)} \Rightarrow \sigma_* B$$

vājā nozīmē telpā $D[0, 1]$, kur σ_*^2 ir novērojumu dispersija σ^2 un B ir standarta Brauna kustība intervālā $[0, 1]$.

Tad formulas (1.1) skaitītāju varam izteikt kā $f(S^{(n)})$. Tādēļ, ja FCLT izpildās, pēc nepārtrauktā attēlojuma teorēmas (Skat. pielikumu) iegūstam, ka

$$\begin{aligned} & \frac{1}{\sqrt{n}} \left(\max_{0 \leq i \leq n} \left(S_i - \frac{i}{n} S_n \right) - \min_{0 \leq i \leq n} \left(S_i - \frac{i}{n} S_n \right) \right) = f \left(\frac{1}{\sqrt{n}} S^{(n)} \right) \\ & \Rightarrow f(\sigma_* B) = \sigma_* \left[\sup_{0 \leq t \leq 1} (B(t) - tB(1)) - \inf_{0 \leq t \leq 1} (B(t) - tB(1)) \right] \\ & := \sigma_* \left[\sup_{0 \leq t \leq 1} B_0(t) - \inf_{0 \leq t \leq 1} B_0(t) \right], \end{aligned}$$

kur B_0 apzīmē Brauna tiltu intervālā $[0,1]$. Turklāt, ja stacionāra virkne X_1, X_2, \dots ir ergodiska, tad izlases standartnovirze ir būtisks populācijas standartnovirzes novērtējums. Tādēļ

$$\left(\frac{1}{n} \sum_{i=1}^n \left(X_i - \frac{1}{n} S_n \right)^2 \right)^{1/2} \rightarrow \sigma$$

ar varbūtību 1. Visbeidzot varam secināt, ka

$$\frac{1}{\sqrt{n}} \frac{R}{S} (X_1, \dots, X_n) \Rightarrow \frac{\sigma_*}{\sigma} \left[\sup_{0 \leq t \leq 1} B_0(t) - \inf_{0 \leq t \leq 1} B_0(t) \right].$$

Tas nozīmē, ka R/S statistika pieaug kā kvadrātsakne no izlases apjoma. Tomēr, kad Haralds Hursts aprēķināja R/S statistiku Nīlas datiem, atklājās, ka empīriskais pieaugums ir tuvāks $n^{0.74}$. Šo fenomenu sauc par *Hursta fenomenu*. Izrādījās, ka atrast piemērotu stohastisku modeli, kas izskaidrotu šo parādību, nav vienkārši. Vairāk informācijas atrodams grāmatā [4]. Tika izvirzīta hipotēze, ka jāņem modelis, kuram korelācijas dilst lēni un tādēļ nav spēkā FCLT. Vienkāršākais no šādiem modeļiem ir frakcionālais Gausa troksnis, kuram izpildās

$$n^{-H} \frac{R}{S} (X_1, \dots, X_n) \Rightarrow \frac{1}{\sigma} \left[\sup_{0 \leq t \leq 1} (B_H(t) - tB_H(1)) - \inf_{0 \leq t \leq 1} (B_H(t) - tB_H(1)) \right],$$

redzams, ka R/S statistika aug kā n^H . Tāpēc, izvēloties atbilstošu H , kuru dēvē par *Hursta parametru*, beidzot bija iespējams izskaidrot problēmu. Nonāca pie secinājuma, ka pats svarīgākais faktors ir lēnā korelāciju dilšana, sevišķi pie lielām H vērtībām - neparasti ilglaicīga atmiņa. Līdz ar to tika ieviesti tādi termini kā ilglaicīgā atmiņa un ilglaicīgā atkarība. Hursta parametrs parādās dažādās matemātikas sfērās, ieskaitot fraktālus un haosa teoriju, ilglaicīgās atmiņas procesus un spektrālo analīzi.

Parametrs H pieņem vērtības no 0.5 līdz 1. Procesam, kuram piemīt LRD, $H > 0.5$. Jo Hursta parametrs tuvāks 1, jo lielāka LRD pakāpe un sarežģītāka procesa struktūra.

2. Ilglaicīgās atmiņas definīcijas

Apskatīsim veidus, kā tradicionāli definē ilglaicīgās atmiņas iedarbību un pierādīsim teorēmu par definīciju ekvivalenci pie īpašiem nosacījumiem, izmantojot W. Palma grāmatu “Long - Memory Time Series: Theory and Methods”[6].

Definīcija 2. Pieņem, ka $\gamma(h) = \langle y_t, y_{t+h} \rangle$ ir autokovariāciju funkcija stacionāram procesam $\{y_t : t \in \mathbb{Z}\}$. Ilglaicīgo atmiņu definē

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| = \infty. \quad (2.1)$$

Izmanto arī citas alternatīvas definīcijas. Lai tās saprastu, vispirms definēsim dažus jēdzienus.

Definīcija 3. Pozitīvu mērojamu funkciju, definētu kādā bezgalības apkārtņē $[a, \infty)$, sauc par *lēni variējošu* Karamata izpratnē tad un tikai tad, ja katram $c > 0$ izpildās $\frac{l(cx)}{l(x)} \rightarrow 1$, kad $x \rightarrow \infty$.

Lēni variējošas funkcijas piemēri ir $l(x) = \log(x)$ un $l(x) = b$, kur b ir pozitīva konstante. Funkcija $l(x) = x^\beta$ nav lēni variējoša nevienam reālam $\beta \neq 0$.

Teorēma 2. *Volda dekompozīcijas teorēma: Jebkuru stacionāru, nedeterminētu procesu var izteikt*

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \psi(B)\varepsilon_t, \quad (2.2)$$

kur $\psi_0 = 1$, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, ε_t ir baltais troksnis ar dispersiju σ^2 .

Ilglaicīgo atmiņu definē

$$\gamma(h) \sim h^{2d-1} l_1(h), \quad (2.3)$$

$$f(\lambda) \sim |\lambda|^{-2d} l_2\left(\frac{1}{|\lambda|}\right), \quad (2.4)$$

$$\psi_j \sim j^{d-1} l_3(j), \quad (2.5)$$

kur $j > 0$, $\lambda \in 0$ apkārtnei, $l_1(\cdot), l_2(\cdot), l_3(\cdot)$ ir lēni variējošas funkcijas, d ir tā dēvētais *ilglaicīgās atmiņas parametrs* un apzīmējums $x_n \sim y_n$ jāsaprot kā $x_n/y_n \rightarrow 1$, kad $n \rightarrow \infty$.

Parametru d un Hursta parametru saista sakarība $H = d + \frac{1}{2}$. Definīcija (2.3) balstās uz autokovariāciju dilšanu, kad $h \rightarrow \infty$, plaši izmantotā (2.4) izsaka LRD caur spektrālajām īpašībām un (2.5) balstās uz procesa Volda dekompozīciju.

Definīcijas (2.1) - (2.5) ir ekvivalentas tikai pie īpašiem nosacījumiem, kurus apskatīsim sekojošā teorēmā. Definēsim divas lemmas, kas nepieciešamas pierādījumam.

Lemma 3. Pieņemsim, ka $r, s \in \mathbb{R}$ un $s < 1 < r + s$; l_1, l_2 ir lēni variējošas funkcijas un $\{f_k\}, \{g_k\}$ ir divas virknes, kuras pieder kopai \mathbb{N} , kuras apmierina $f_k \sim k^{-r}l_1(k), g_k \sim k^{-s}l_2(k)$, kad $k \rightarrow \infty$. Jebkuram $n \in \mathbb{N}$, virkne $\{f_{n+k}g_k\}$ ir summējama un

$$\sum_{k=0}^{\infty} f_{n+k}g_k \sim n^{-(r+s-1)}l_1(n)l_2(n)B(r+s-1, 1-s),$$

kad $n \rightarrow \infty$, kur $B(\cdot, \cdot)$ ir beta funkcija.

Lemma 4. (a) Ja $l(\cdot)$ ir lēni variējoša un $\alpha > 0$, tad $x^\alpha l(x) \rightarrow \infty$, kad $x \rightarrow \infty$.

(b) Ja $l(\cdot)$ ir lēni variējoša, m ir pietiekoši liels, tāds, ka $l(x)$ ir lokāli ierobežota apgabalā $[m, \infty)$ un $\alpha > -1$, tad

$$\sum_{k=m}^n k^\alpha l(k) \sim \frac{n^{\alpha+1}l(n)}{\alpha+1},$$

kad $n \rightarrow \infty$.

(c) Ja $l(\cdot)$ ir kvazi - monotona lēni variējoša, tad sekojošā rinda ir nosacīti konverģenta

$$\sum_{k=1}^{\infty} l(k) \cos(k\lambda) k^{-\mu} \sim \frac{\pi \lambda^{\mu-1} l(\frac{1}{\lambda})}{2\Gamma(\mu) \cos(\frac{\pi\mu}{2})},$$

visiem $\lambda \rightarrow 0_+$ un $0 < \mu < 1$.

Teorēma 5. Pieņem, ka $\{y_t\}$ ir stacionārs process Volda izteiksmē (2.2), $0 < d < \frac{1}{2}$, tad

(a) Ja procesam $\{y_t\}$ izpildās (2.5), tad izpildās arī (2.3).

(b) Ja procesam $\{y_t\}$ izpildās (2.3), tad izpildās arī (2.1).

(c) Ja $l_1(\cdot)$ ir kvazi - monotona lēni variējoša funkcija, tad (2.3) nozīmē (2.4).

Pierādījums. a) No Volda dekompozīcijas (2.2) seko [6], ka procesa $\{y_t\}$ autokovariāciju funkciju var uzrakstīt formā

$$\gamma(h) = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+|h|}.$$

Tādēļ, pēc nosacījuma (2.5) un Lemmas 3, mēs varam secināt, ka

$$l_1(h) = \sigma^2 l_3(h)^2 B(1-2d, d).$$

b) Jebkuram veselam skaitlim $0 < m < n$

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| \geq \sum_{h=m}^n |\gamma(h)|.$$

Tagad, no un Lemmas 4 (b) pie lieliem n iegūsim

$$\sum_{h=m}^n |\gamma(h)| \sim \frac{n^{2d} l_1(n)}{2d}.$$

Tā kā $d > 0$, tad no Lemmas 4 seko, ka $n^{2d} l_1(n) \rightarrow \infty$, kad $n \rightarrow \infty$. Tas nozīmē, ka nosacījums (2.3) apmierina nosacījumu (2.1).

c) Tā kā l_1 ir kvazi - monotona lēni variējoša, tad no Lemmas 4 punkta (c) seko, ka (2.4):

$$l_2(\lambda) = \frac{l_1(\lambda)}{2\Gamma(1-2d)\sin(\pi d)} = \frac{\Gamma(1-d)\Gamma(d)}{2\pi\Gamma(1-2d)} l_1(\lambda)$$

visiem $\lambda > 0$.

□

3. Frakcionālā Brauna kustība un frakcionālais Gau- sa troksnis

Frakcionālo Brauna kustību turpmāk apzīmēsim ar fBm. Izmantots materiāls no grā-
matas “Theory and Applications of Long Range Dependence” [7].

Definīcija 4. Stohastisku procesu $Z = \{Z(t)\}_{t \in \mathbb{R}}$, kas pieņem reālas vērtības, sauc par
sev līdzīgu ar indeksu $H > 0$, visiem $a > 0$, ja

$$\{Z(at)\}_{t \in \mathbb{R}} =^d \{a^H Z(t)\}_{t \in \mathbb{R}}, \quad (3.1)$$

kur ar $=^d$ apzīmē galīgdimensionālu sadalījumu ekvivalenci. Procesu (3.1) turpmāk dē-
vēsim par H - se procesu. Kaut arī tie nav stacionāri, starp sev līdzīgiem un stacionāriem
procesiem pastāv zināma sakarība.

Apgalvojums 6. Ja process $\{X(t)\}_{t > 0}$ ir H - se, tad process $Y(t) = e^{-tH} X(e^t), t \in \mathbb{R}$,
ir stacionārs. Kā arī otrādi, ja $\{Y(t)\}_{t \in \mathbb{R}}$ ir stacionārs, tad $X(t) = t^H Y(\ln t), t > 0$, ir
 H - se process.

Definīcija 5. Saka, ka procesam $Z = \{Z(t)\}_{t \in \mathbb{R}}$ ir stacionāri pieaugumi, ja, visiem $h \in \mathbb{R}$,

$$\{Z(t+h) - Z(h)\}_{t \in \mathbb{R}} =^d \{Z(t) - Z(0)\}_{t \in \mathbb{R}}.$$

Procesus H - se plaši izmanto, jo no tiem var izveidot virknes ar svarīgām īpašībām. Ja
 $\{Z(t)\}_{t \in \mathbb{R}}$ ir H - ss process, tad

$$X_j = \Delta Z(j) = Z(j+1) - Z(j), j \in \mathbb{Z},$$

ir stacionāra virkne, saukta par laikrindām.

H - se procesam $Z = \{Z(t)\}_{t \in \mathbb{R}}$ ar ierobežotu dispersiju piemīt sekojošas īpašības:

- $Z(0) = 0$.
- Ja $H \neq 1$, tad $EZ(t) = 0$ visiem $t \in \mathbb{R}$.
- $Z(-t) = -Z(t)$.
- $EZ^2(t) = EZ^2(|t| \text{ sign}t) = |t|^{2H} EZ^2(\text{sign}t) = |t|^{2H} EZ^2(1) = |t|^{2H} \sigma^2$. Ja $\sigma^2 = EZ^2(1) = 1$, tad procesu $Z = \{Z(t)\}_{t \in \mathbb{R}}$ sauc par standarta.

- H -ss procesa Z kovariāciju funkcija ir

$$\begin{aligned} \text{Cov}(Z(s), Z(t)) &= \Gamma_H(s, t) = \frac{1}{2} \{EZ^2(s) + EZ^2(t) - E(Z(t) - Z(s))^2\} \\ &= \frac{\sigma^2}{2} \{|t|^{2H} + |s|^{2H} - |t - s|^{2H}\}. \end{aligned}$$

- Procesas Z parametrs $H \leq 1$.

Ja $0 < H \leq 1$, tad kovariāciju funkcija ir nenegatīvi definīta.

Definīcija 6. Gausa H - se procesu $\{B_H(t)\}_{t \in \mathbb{R}}$, kur $0 < H < 1$, sauc par *frakcionālo Brauna kustību*. To dēvē par standarta fBm, ja $\sigma^2 = \text{Var}B_H(1) = 1$.

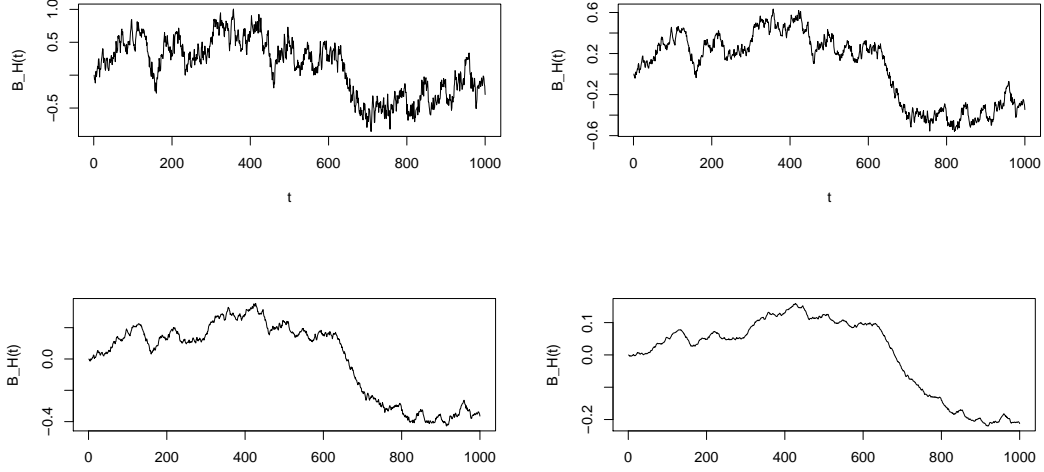
Piezīme 7. Ja $H = \frac{1}{2}$, tad fBm $\{B_{\frac{1}{2}}(t)\}_{t \in \mathbb{R}}$ ir parastā Brauna kustība (Definīcija 1) ar $\text{Cov}(B(s), B(t)) = EB_{\frac{1}{2}}(s)B_{\frac{1}{2}}(t) = \min(|s||t|)$, ja s un t ir ar vienādām zīmēm un $\text{Cov}(B(s), B(t)) = 0$, ja zīmes ir pretējas.

Apgalvojums 8. Pieņemsim, ka $\{X(t)\}_{t \in \mathbb{R}}$

1. ir Gausa process ar vidējo vērtību 0, $X(0) = 0$,
2. $EX^2(t) = \sigma^2|t|^{2H}$ kādam $\sigma > 0$ un $0 < H < 1$,
3. ir stacionāri pieaugumi;

tad $\{X(t)\}_{t \in \mathbb{R}}$ ir frakcionālā Brauna kustība. Frakcionālā Brauna kustība ir vienīgais Gausa H - se process.

Parametra H vērtība ietekmē fBm trajektorijas izskatu. Attēlā 1. redzamas viena fBm procesa simulācijas dažādiem H . Pie lielākām H vērtībām, fBm trajektorija kļūst līdzeknāka. Tas skaidrojams ar kovariāciju struktūru. Kad $0 < H < \frac{1}{2}$, pieaugumi no fBm ir ar pretējām zīmēm un to kovariācijas ir negatīvas. Savukārt, gadījumā $\frac{1}{2} < H < 1$, tās pieņem pozitīvas vērtības.



1. att. Frakcionālā Brauna kustība ar $H = 0.3$, $H = 0.5$, $H = 0.7$ un $H = 0.9$

Labi pazīstams LRD process ir *frakcionālais Gausa troksnis* (turpmāk lietošim saīsinājumu fGn), kas būtībā ir diskrētu soļu pieaugumu process no fBm.

Definīcija 7. Ja $\{Z(t)\}_{t \in \mathbb{R}}$ ir fBm, tad procesu $\{X_k\}_{k \in \mathbb{Z}}$ sauc par frakcionālo Gausa troksni, kuram izpildās

$$X_k = Z(k+1) - Z(k), k \in \mathbb{Z}.$$

Apgalvojums 9. Augošai virknei $\{X_k\}_{k \in \mathbb{Z}}$ (Definīcija 7) piemīt sekojošas īpašības:

1. $\{X_k\}_{k \in \mathbb{Z}}$ ir stacionāra.
2. $EX_k = 0$.
3. $EX_k^2 = \sigma^2 = EZ(1)^2$.
4. Procesa $\{X_k\}_{k \in \mathbb{Z}}$ autokovariāciju funkcija ir dota ar

$$\gamma(k) = EX_i X_{i+k} = \frac{\sigma^2}{2} (|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}) = \frac{\sigma^2}{2} \Delta^2 |k|^{2H},$$

kur Δ^2 apzīmē otrās kārtas diferenci.

5. Pieņem, ka $k \neq 0$. Tad $\gamma(k) = 0$, ja $H = \frac{1}{2}$, $\gamma(k) < 0$, ja $0 < H < \frac{1}{2}$ un $\gamma(k) > 0$, ja $\frac{1}{2} < H < 1$.
6. Ja $H \neq \frac{1}{2}$, tad

$$\gamma(k) \sim \sigma^2 H(2H-1) |k|^{2H-2},$$

kad $k \rightarrow \infty$.

Apgalvojumu pierādījumi pieejami [7].

4. FARIMA modeļi

Plaši pazīstama ilglaicīgās atmiņas modeļu klase ir frakcionālie autoregresīvie slidošā - vidējā jeb *FARIMA* (dažkārt arī - *ARFIMA*) modeļi. Būtībā tie ir ARIMA procesu vispārinājums, tikai FARIMA var piemist arī ilglaicīgā atmiņa [3].

Definīcija 8. [3] Stacionāru procesu X_t sauc par FARIMA (p, d, q) procesu, ja

$$\phi(B)(1-B)^d X_t = \psi(B)\epsilon_t, \quad (4.1)$$

visiem $-\frac{1}{2} < d < \frac{1}{2}$, kur $(1-B)^d$ ir diferencu operators, $\{\epsilon_t\}$ - baltais troksnis ar ierobežotu dispersiju, $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ un $\psi(B) = 1 + \psi_1 B + \dots + \psi_q B^q$ ir attiecīgi autoregresīvā un slidošā vidējā operatori.

Procesam piemīt ilglaicīgā atmiņa, ja $0 \leq d < \frac{1}{2}$. Nepieciešams definēt augšējo robežu $d < \frac{1}{2}$, jo, gadījumā $d \geq \frac{1}{2}$, process nav stacionārs. Jebkuram $d \in \mathbb{R}$, diferencu operatoru $(1-B)^d$ definē

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k,$$

kur

$$\binom{d}{k} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}$$

Vienādojumu (4.1) var izteikt dažādās formās. Piemēram,

$$(1-B)^d X_t = \tilde{X}_t, \quad (4.2)$$

kur \tilde{X}_t ir ARMA process, definēts kā

$$\tilde{X}_t = \phi^{-1}(B)\psi(B)\epsilon_t.$$

Tas nozīmē, ka, pielietojot X_t frakcionālo diferencu operatoru (bezgalīgu lineāru filtru) $(1-B)^d$, iegūst ARMA procesu. No otras puses, var pierakstīt

$$X_t = \phi(B)^{-1}\psi(B)X_t^*,$$

kur

$$X_t^* = (1-B)^d \epsilon_t \quad (4.3)$$

ir FARIMA $(0, d, 0)$ process. Tad X_t tiek iegūts, lietojot ARMA filtru procesam FARIMA $(0, d, 0)$.

4.1. Bezgalīga AR un MA reprezentācija

Ja process FARIMA(p, d, q) ir stacionārs un apgriezams (skat. [6] 44.lpp), tad

$$X_t = (1 - B)^{-d} \phi(B)^{-1} \psi(B) \epsilon_t = a(B) \epsilon_t,$$

un

$$\epsilon_t = (1 - B)^d \phi(B) \psi(B)^{-1} X_t = \pi(B) y_t.$$

MA (∞) koeficienti, ψ_j , un AR (∞) koeficienti, π_j , apmierina sekojošas asimptotiskas sakarības:

$$a_j = \frac{\psi(1)k^{d-1}}{\phi(1)\Gamma(d)} + \mathcal{O}(k^{-1}),$$

$$\pi_j = \frac{\phi(1)k^{-d-1}}{\psi(1)\Gamma(-d)} + \mathcal{O}(k^{-1}),$$

kad $k \rightarrow \infty$.

Frakcionāliem trokšņa procesiem ar ilglaicīgās atmiņas parametru d attiecīgie koeficienti ir doti ar vienādojumiem

$$a_k = \prod_{t=1}^k \frac{t-1+d}{t} = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)},$$

$$\pi_k = \prod_{t=1}^k \frac{t-1-d}{t} = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)},$$

visiem $k \geq 1$ un $a_0 = \pi_0 = 1$.

4.2. Spektrālais blīvums

Procesa (4.1) spektrālo blīvumu var pierakstīt sekojošā veidā

$$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} |1 - e^{i\lambda}|^{-2d} \frac{|\psi(e^{i\lambda})|^2}{|\phi(e^{i\lambda})|^2} = \frac{\sigma_\epsilon^2}{2\pi} \left[2 \sin \frac{\lambda}{2} \right]^{-2d} \frac{|\psi(e^{i\lambda})|^2}{|\phi(e^{i\lambda})|^2},$$

un tas ir

$$f(\lambda) \sim \frac{\sigma_\epsilon^2}{2\pi} \frac{|\psi(1)|^2}{|\phi(1)|^2} |\lambda|^{-2d},$$

visiem $|\lambda| \rightarrow 0$.

Procesa X_t spektrālo blīvumu raksturo arī ar periodigrammu

$$I(\lambda_j) = \frac{1}{2\pi n} \left| \sum_{t=1}^n X_t e^{it\lambda_j} \right|^2. \quad (4.4)$$

4.3. Autokovariāciju funkcija

No formulas (2.3) zināms, ka, visiem $|k| \rightarrow \infty$,

$$\gamma(k) \sim c_\gamma(d, \phi, \psi)|k|^{2d-1}, \quad (4.5)$$

kur

$$c_\gamma(d, \phi, \psi) = \frac{\sigma_\epsilon^2 |\psi(1)|^2}{\pi |\phi(1)|^2} \Gamma(1 - 2d) \sin \pi d$$

un

$$\rho \sim c_\rho(d, \phi, \psi)|k|^{2d-1},$$

kur

$$c_\rho(d, \phi, \psi) = \frac{c_\gamma(d, \phi, \psi)}{\int_{-\pi}^{\pi} f(\lambda) d\lambda}.$$

Procesa FARIMA $(0, d, 0)$ autokovariāciju funkcija ir

$$\gamma_0(k) = \sigma_\epsilon^2 \frac{(-1)^k \Gamma(1 - 2d)}{\Gamma(k - d + 1) \Gamma(1 - k - d)},$$

savukārt, autokorelāciju funkcija izsakāma ar

$$\rho_0(k) = \frac{\Gamma(1 - d)}{\Gamma(k + d)} \frac{\Gamma(d)}{\Gamma(k + 1 - d)}.$$

Vispārīgā gadījumā, procesa FARIMA (p, d, q) autokovariāciju funkciju var pierakstīt

$$\gamma(k) = \sum_{j,l=0}^{\infty} \beta_j \beta_l \gamma^*(k + j - l),$$

kur

$$\beta(B) = \sum_{j=0}^{\infty} \beta_j B^j = \phi(B) \psi^{-1}(B)$$

un $\gamma^*(k)$ ir autokovariāciju funkcija procesam X_t^* (4.3).

4.4. Izlases vidējā vērtība

Pieņemam, ka X_1, X_2, \dots, X_n ir izlase no procesa FARIMA (p, d, q) un \bar{X} ir izlases vidējā vērtība. Tad \bar{X} dispersija

$$\text{Var}(\bar{X}) = \frac{1}{n} \left[2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n} \right) \gamma(k) + \gamma(0) \right]. \quad (4.6)$$

No (4.5) seko, ka $\gamma(k) \sim c_\gamma k^{2d-1}$ pie lieliem k . Tādēļ, pie lielām n vērtībām, iegūsim

$$\begin{aligned}\text{Var}(\bar{X}) &\sim 2c_\gamma n^{2d-1} \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \left(\frac{k}{n}\right)^{2d-1} \frac{1}{n} \\ &\sim 2c_\gamma n^{2d-1} \int_0^1 (1-t)t^{2d-1} dt \\ &\sim \frac{c_\gamma}{d(2d+1)} n^{2d-1}.\end{aligned}$$

4.5. Parciālās autokorelācijas

Vispārīgam FARIMA (p, d, q) modelim, PACF noteikšana ir sarežģīta, tomēr koeficientu β_{kj} iegūšanas princips ir tāds pats, kā ARMA procesu gadījumā. Parciālos korelācijas koeficientus nosaka ar Durбина - Levinsona algoritmu. Frakcionālā trokšņa gadījumā Hoskings (1981) [8] apgalvo:

Apgalvojums 10. *Pieņem, ka X_t ir FARIMA $(0, d, 0)$ process, kur $-\frac{1}{2} < d < \frac{1}{2}$, un*

$$\hat{X}_t = \sum_{j=1}^k \beta_{kj} X_{t-j}$$

ir X_t labākās lineārās prognozes. Tad koeficienti β_{kj} minimizē sagaidāmo kvadrātisko prognozes kļūdu $E \left[(X_t - \hat{X}_t)^2 | X_{t-1}, \dots, X_{t-k} \right]$ un

$$\beta_{kj} = - \binom{k}{j} \frac{\Gamma(j-d)\Gamma(k-d-j+1)}{\Gamma(-d)\Gamma(k-d+1)}.$$

Parciālās autokorelācijas koeficienti β_{kk} procesam FARIMA $(0, d, 0)$

$$\beta_{kk} = \frac{d}{k-d}.$$

Gadījumā, kad $j, k \rightarrow \infty$ un $j/k \rightarrow 0$,

$$\beta_{kj} \sim \frac{1}{\Gamma(-d)} j^{-d-1}.$$

Inoue(2002) [9] ir pierādījis, ka, ja β_{kk} ir stacionāra procesa FARIMA parciālās autokorelācijas, tad

$$|\beta_{kk}| \sim \frac{d}{k},$$

kad $k \rightarrow \infty$.

5. Parametra H novērtēšanas metodes

Ilglaicīgās atmiņas klātbūtni praksē novēroja ilgi pirms bija zināmi piemēroti stohastiski modeļi, tādēļ parametra H novērtēšanai lietoja dažādas uz empīriskiem novērojumiem balstītas metodes. Tās galvenokārt izmanto ilglaicīgās atmiņas diagnosticēšanā.

5.1. R/S statistika

Metode pamatojas uz Hursta statistiku (1.1), kuru jau apskatījām pirmajā nodaļā. Šajā gadījumā pieņem, ka $Q = Q(t, k) = R(t, k)/S(t, k)$, kur

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

un

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

ja $Y_j = \sum_{i=1}^j X_i$ un $\bar{X}_{t,k} = k^{-1} \sum_{i=t+1}^{t+k} X_i$. Lai novērtētu parametru, aprēķina Q visām (vai arī - pietiekoša skaita) t un k vērtībām, grafiski attēlo $\log Q$ pret $\log k$ un zīmē taisni $y = a + b \log k$.

Teorēma 11. [3] 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 fBm, kad t tiecas uz bezgalību. Tad, visiem $k \rightarrow \infty$

$$k^{-H} Q \rightarrow_d \xi, \tag{5.1}$$

kur ξ ir nedeģenerēts gadījuma lielums.

Statistiskiem pielietojumiem tas nozīmē, ka grafika $\log Q$ pret $\log k$ punkti būs vienmērīgi izkaisīti ap taisni ar slīpumu H . Ilglaicīgās atmiņas gadījumā ap $H > 0.5$, pretējā gadījumā - ap $H = 0.5$. Metode ir populāra un plaši izmantota, tomēr tai piemīt savi trūkumi. Piemēram, liela nozīme ir k izvēlei. Lietojot dažādas k vērtības vienai datu kopai, iespējams iegūt krasi atšķirīgus H novērtējumus, pat ārpus intervāla $0 < H < 1$. Turklāt, gadījumos, kad procesam nepiemīt ilglaicīgā atmiņa, bet ir lēni dilstošs trends, metode var dot aplamu slēdzienu par ilglaicīgās atmiņas klātbūtni. Sīkākas detaļas atrodamas Mandelbrot (1972) [10] un citos darbos.

5.2. Dispersiju grafiki

Iepriekš noskaidrojām (4.6), ka $\text{Var}(\bar{X}) \sim cn^{2d-1}$, pie lieliem n un $c > 0$.

Pieņem, ka k ir vesels skaitlis. Sadalām izlasi m_k blokos ar garumu k , tad

$$\log \text{Var}(\bar{X}(k)) \sim c + (2d - 1) \log k. \quad (5.2)$$

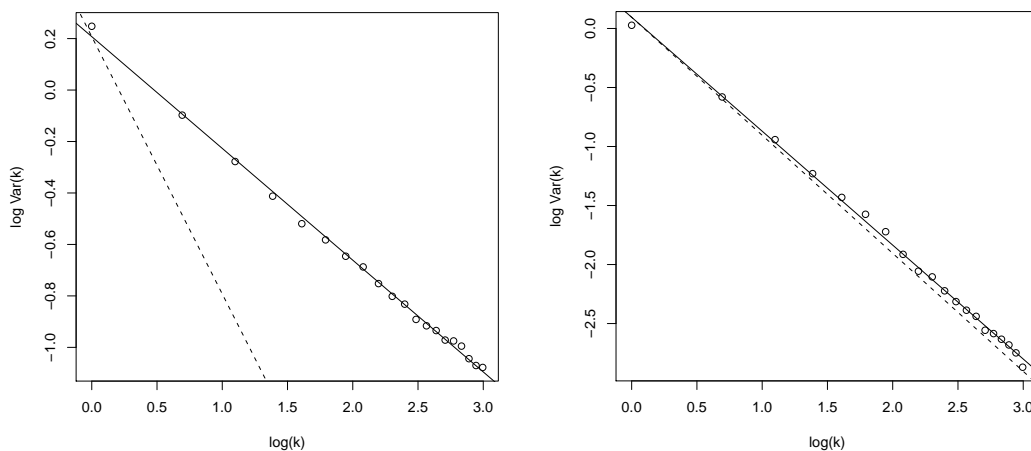
Pietiekami lielam skaitam bloku ar dažādām k vērtībām intervālā $2 \leq k \leq \frac{n}{2}$, aprēķinām bloku vidējās vērtības $\bar{X}_1(k), \bar{X}_2(k), \dots, \bar{X}_{m_k}(k)$ un izlases kopējo vidējo vērtību

$$\bar{X}(k) = m_k^{-1} \sum_{j=1}^{m_k} \bar{X}_j(k).$$

Katram k aprēķina dispersiju no bloku vidējām vērtībām $\bar{X}_j(k)$, $j = 1, \dots, m_k$ un izlases vidējās vērtības:

$$s^2(k) = (m_k - 1)^{-1} \sum_{j=1}^{m_k} (\bar{X}_j(k) - \bar{X}(k))^2.$$

Grafiski attēlo $\log s^2(k)$ pret $\log k$. Pie lielām k vērtībām, sagaidāms, ka grafika punkti būs izkļiedēti ap taisni ar negatīvu slīpumu $2d - 1$ (jeb $2H - 2$). Gadījumā, kad procesam nepiemīt ilglaicīgā atmiņa ($d = 0$), taisnes slīpums ir -1 un tā būs lēzenāka. Izdevīgi šādu taisni piezīmēt dispersiju grafikam. Jo tā stāvāka, jo izteiktāka ilglaicīgās atmiņas klātbūtne. Parametra H vērtība ir iegūtais taisnes slīpums $\times 0.5 + 1$. Metodes trūkumi būtībā ir tādi paši kā R/S statistikai.



2. att. Procesu FARIMA(0, 0.3, 0) un ARMA (1, 0) dispersiju grafiki

Attēla 2. kreisajā pusē redzams dispersiju grafiks simulētam frakcionālā trokšņa procesam ar $d = 0.3$, savukārt, labajā pusē - pirmās kārtas autoregresīvajam modelim, ar $\phi = 0.3$. Abos gadījumos izlases garums $n = 10000$ un bloku garums $k = 20$. Nepārtrauktā līnija raksturo vienādojumu (5.2), raustītā līnija attēlo taisni ar slīpumu -1 . ARMA(1,0) gadījumā, abas taisnes gandrīz pārklājas, līdz ar to metode noraida ilglaicīgās atmiņas klātbūtni.

5.3. Regresiju metode

Pieņem, ka stacionāra procesa spektrālais blīvums ir

$$f(\lambda) = f_0(\lambda) \left[2 \sin \left(\frac{\lambda}{2} \right) \right]^{-2d}. \quad (5.3)$$

Logaritmējot vienādojumu (5.3) punktos $\lambda_j = 2\pi j/n$, iegūstam

$$\log f(\lambda_j) = \log f_0(0) - d \log \left[2 \sin \frac{\lambda_j}{2} \right]^2 + \log \left[\frac{f_0(\lambda_j)}{f_0(0)} \right]. \quad (5.4)$$

No otras puses, periodigrammu $I(\lambda_j)$ var izteikt kā

$$\log I(\lambda_j) = \log \left[\frac{I(\lambda_j)}{f(\lambda_j)} \right] + \log f(\lambda_j). \quad (5.5)$$

Apvienojot (5.4) un (5.5),

$$\log I(\lambda_j) = \log f_0(0) - d \log \left[2 \sin \frac{\lambda_j}{2} \right]^2 + \log \left[\frac{I(\lambda_j) [2 \sin(\lambda/2)]^{2d}}{f_0(0)} \right].$$

Definējot $y_j = \log I(\lambda_j)$, $\alpha = \log f_0(0)$, $\beta = -d$, $x_j = \log [2 \sin(\lambda_j/2)]^2$ un

$$\epsilon_j = \log \left\{ \frac{I(\lambda_j) [2 \sin(\lambda/2)]^{2d}}{f_0(0)} \right\},$$

iegūst regresijas vienādojumu

$$y_j = \alpha + \beta x_j + \epsilon_j.$$

Varam sagaidīt, ka, visiem $j = 1, \dots, m$, kur $m \ll n$,

$$f(\lambda_j) \sim f_0(0) [2 \sin(\lambda_j/2)]^{-2d},$$

tādēļ

$$\epsilon_j \sim \log \left[\frac{I(\lambda_j)}{f(\lambda_j)} \right].$$

Tad mazāko kvadrātu metodes ilglaicīgās atmiņas parametra d novērtējums ir

$$\hat{d}_m = - \frac{\sum_{j=1}^m (x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^m (x_j - \bar{x})^2},$$

kur $\bar{x} = \sum_{j=1}^m x_j/m$ un $\bar{y} = \sum_{j=1}^m y_j/m$. 1

5.4. Detrendētā svārstību metode

Metodi izstrādājis Pengs [11]. Pieņem, ka X_1, \dots, X_n ir izlase no kāda stacionāra ilglaicīgās atmiņas procesa un Y_t ir parciālo summu virkne $Y_t = \sum_{j=1}^t X_j$, visiem $t = 1, \dots, n$. Tad procesa X_t , $t \in \mathbb{Z}$, parametra d novērtējumu iegūst sekojoši. Izlasi X_t sadala k nepārklājošos blokos, kuri satur $m = n/k$ novērojumus. Katrā blokā X_t piemēro lineārās regresijas modeli attiecībā pret $t = 1, \dots, m$ un novērtē atlikumu dispersijas σ_k^2

$$\sigma_k^2 = \frac{1}{m} \sum_{t=1}^m (Y_t - \hat{\alpha}_k - \hat{\beta}_k t)^2,$$

kur $\hat{\alpha}_k$ un $\hat{\beta}_k$ ir mazāko kvadrātu metodes koeficientu novērtējumi.

Pieņem, ka $F^2(k)$ ir šo dispersiju vidējā vērtība

$$F^2(k) = \frac{1}{k} \sum_{j=1}^k \sigma_j^2.$$

Kā aprakstījis Pengs [11], gadījuma klejošanai,

$$F(k) \sim ck^{1/2},$$

savukārt procesiem ar LRD,

$$F(k) \sim ck^{d+1/2}. \quad (5.6)$$

Tad, logaritmējot (5.6), iegūst

$$\log F(k) \sim \log c + (d + \frac{1}{2}) \log k.$$

Piemērojot mazāko kvadrātu regresijas modeli

$$\log F(k) = \alpha + \beta \log k + \epsilon_k$$

visām $k \in K$ vērtībām, iegūst d novērtējumu

$$\hat{d} = \hat{\beta} - \frac{1}{2},$$

kur $\hat{\beta}$ ir mazāko kvadrātu novērtējums parametram β . Ir dažādas pieejas indeksa K izvēlei. Ja $k_0 = \min\{K\}$ un $k_1 = \max\{K\}$, daži autori piedāvā izmantot $k_0 = 4$ un k_1 - daļa no izlases apjoma n . [6]

6. Vislielākās ticamības parametru novērtēšanas metodes

Definēsim dažus jēdzienus, kas būs nepieciešami turpmākai analīzei. Vairāk informācijas par nodaļā aplūkotajām metodēm un teorēmu pierādījumus skatīt materiālā [3].

Pieņem, ka X_1, X_2, \dots, X_n ir stacionārs process ar vidējo vērtību μ un dispersiju σ^2 , kā arī - izpildās (2.3) un (2.4). Spektrālo blīvumu raksturo ar nezināmu galīgdimensionālu parametru vektoru

$$\theta^0 = (\sigma_0^2, H^0, \theta_3^0, \dots, \theta_M^0).$$

Tad $f(\lambda) = f(\lambda; \theta)$, kur $\theta \in \Theta \subset R^M$. Mērķis ir novērtēt nezināmo parametru vektoru θ^0 dotajiem novērojumiem X_1, \dots, X_n . Apzīmē: $X = (X_1, X_2, \dots, X_n)^t$,

$\Sigma_n(\theta) = [\gamma(j-l)]_{j,l=1,\dots,n}$ = kovariāciju matrica no X ,

$|\Sigma_n|$ = determinants no Σ_n .

Pieņemsim, ka μ ir zināms un vienāds ar nulli. Tas seko no fakta, ka Gausa vislielākās ticamības novērtējuma parametram θ asimptotiskais sadalījums nemainās, neatkarīgi no tā, vai μ tiek aizvietots ar izlases vidējo vērtību. Procesu X_t var izteikt formās

$$X_t = \sum_{s=1}^{\infty} b(s)X_{t-s} + \epsilon_t \quad (6.1)$$

un

$$X_t = \sum_{s=0}^{\infty} a(s)\epsilon_{t-s}, \quad (6.2)$$

kur ϵ_s ir nekorelēti gadījuma lielumi ar vidējo vērtību nulle un dispersiju $\text{var}(\epsilon_s) = \sigma_\epsilon^2$.

Lemma 12. *Pieņem, ka eksistē tādas konstantes $0 < c_b < \infty$ un $0 < c_a < \infty$, ka*

$$b(s) \sim c_b k^{-H-\frac{1}{2}}$$

un

$$a(s) \sim c_a k^{H-\frac{3}{2}},$$

kad $k \rightarrow \infty$. Dotajā gadījumā,

$$\sum_{s=1}^{\infty} b(s) < \infty$$

un

$$\sum_{s=1}^{\infty} a(s) = \infty.$$

Pieņemam, ka vēlamies iegūt viena soļa prognozi no X_t neierobežotām pagātnes vērtībām $X_s, s \leq t-1$, tad vidējo kvadrātisko prognozes kļūdu

$$MSPE = E \left[(X_t - \hat{X}_t)^2 | X_s, s \leq t-1 \right]$$

minimizē ar

$$\hat{X}_t = E [X_t | X_s, s \leq t-1] = \sum_{s=1}^{\infty} b(s) X_{t-s}.$$

Minimālā prognozes kļūda

$$MSPE = \sigma_\epsilon^2 = 2\pi \exp \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda; \theta^0) d\lambda \right].$$

No tā seko [3], ka

$$\int_{-\pi}^{\pi} \log f_1(\lambda) d\lambda = 0,$$

visiem

$$f_1(\lambda) = \frac{2\pi}{\sigma_\epsilon^2} f(\lambda).$$

6.1. Gausa vislielākās ticamības novērtējums

Pieņem, ka X_t ir Gausa process, kam izpildās (6.1) un (6.2). Tad $X = (X_1, X_2, \dots, X_n)^t$ kopējā sadalījuma funkcija ir

$$h(x; \theta^0) = (2\pi)^{-\frac{n}{2}} |\Sigma(\theta^0)|^{-\frac{1}{2}} e^{-\frac{1}{2} x^t \Sigma^{-1}(\theta^0) x},$$

kur $x = (x_1, \dots, x_n)^t \in R^n$. Log - ticamības funkcija ir

$$\begin{aligned} L_n(x; \theta^0) &= \log h(x; \theta^0) \\ &= -\frac{n}{2} \log 2\pi - \frac{1}{2} \log |\Sigma(\theta^0)| - \frac{1}{2} x^t \Sigma^{-1}(\theta^0) x. \end{aligned} \quad (6.3)$$

Definē M - dimensiju vektoru

$$\begin{aligned} L'_n(x; \theta) &= \frac{\partial}{\partial \theta_j} L_n(x; \theta) \\ &= -\frac{1}{2} \frac{\partial}{\partial \theta_j} \log |\Sigma(\theta)| - \frac{1}{2} x^t \left[\frac{\partial}{\partial \theta_j} \Sigma^{-1}(\theta) \right] x, \quad (j = 1, 2, \dots, M). \end{aligned}$$

Vislielākās ticamības novērtējumu (MLE) iegūst, maksimizējot $\log h(x; \theta)$ pēc M - dimensiju parametra θ . Attiecīgo maksimizācijas problēmu var formulēt ar pirmās kārtas parciālajiem atvasinājumiem. Tad $MLE \hat{\theta}$ ir M vienādojumu sistēmas

$$L'_n(x; \hat{\theta}) = 0 \quad (6.4)$$

atrisinājums. $\hat{\theta}$ asimptotisko sadalījumu var iegūt no L'_n Teilora izvirzījuma. Otrās kārtas parciālo atvasinājumu matricu no L_n apzīmē ar

$$L''_n(x; \theta) = \frac{\partial^2}{\partial \theta_j \partial \theta_l} L_n(x; \theta), (j, l = 1, \dots, M)$$

No(6.4) iegūstam, ka

$$L'_n(x; \hat{\theta}) = 0 = L'_n(x; \theta^0) + L''_n(x; \theta^0)(\hat{\theta} - \theta^0) + r_n. \quad (6.5)$$

Teorēma 13. Pieņem, ka X_t ir Gausa process un $\hat{\theta}$ ir definēts ar (6.4). Definē $M \times M$ - matricu $D = [D_{ij}]_{i,j=1,\dots,M}$ kā

$$D_{ij}(\theta^0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_i} \log f(x) \frac{\partial}{\partial \theta_j} \log f(x) dx |_{\theta=\theta^0}$$

un

$$C(\theta^0) = 2D^{-1}(\theta^0).$$

Tad, pie nosacījuma $n \rightarrow \infty$, ir spēkā

$$\hat{\theta} \rightarrow \theta^0$$

gandrīz visur un

$$n^{\frac{1}{2}}(\hat{\theta} - \theta^0) \rightarrow_d \zeta,$$

kur ζ ir M - dimensionāls normāls gadījuma vektors ar vidējo vērtību nulle un kovariāciju matricu C .

6.2. Whittle tuvinātais vislielākās ticamības novērtējums

Lai iegūtu Whittle metodes novērtējumus, veic divus tuvinājumus. Tā kā

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |\Sigma_n(\theta)| = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda; \theta) d\lambda,$$

tad $\log |\Sigma_n(\theta)|$ no vienādojuma (6.3) var aizstāt ar $n(2\pi)^{-1} \int_{-\pi}^{\pi} \log f(\lambda; \theta) d\lambda$.

Otru tuvinājumu veic $x^t \Sigma^{-1}(\theta) x$. Matricu $\Sigma^{-1}(\theta)$ aizstāj ar citu, vieglāk aprēķināmu $n \times n$ matricu

$$A(\theta) = [\alpha(j-l)]_{j,l=1,\dots,n},$$

kura sastāv no

$$\alpha(j-l) = (2\pi)^{-2} \int_{-\pi}^{\pi} \frac{1}{f(\lambda; \theta)} e^{i(j-l)\lambda} d\lambda. \quad (6.6)$$

Tad no abiem tuvinājumiem un (6.5) iegūstam

$$L_n^* = -\frac{n}{2} \log 2\pi - \frac{n}{2} \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda; \theta) d\lambda - \frac{1}{2} x^t A(\theta) x.$$

Tuvinātu parametra θ^0 vislielākās ticamības novērtējumu iegūst, minimizējot funkciju

$$L_W(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda; \theta) d\lambda + \frac{x^t A(\theta) x}{n} \quad (6.7)$$

atkarībā no θ .

6.3. Tuvinātie autoregresīvie vislielākās ticamības novērtējumi

Citu tuvināto *MLE* iegūst, izmantojot X_t labāko lineāro prognozi, kas dota neierobežota skaita pagātnes vērtībām X_{t-1}, X_{t-2}, \dots . Ja zināmi $X_s (s \leq t)$, tad neatkarīgiem, vienādi sadalītiem $\epsilon_s (s \leq t)$,

$$\epsilon_t = X_t - \sum_{s=1}^{\infty} b(s) X_{t-s}.$$

Ja X_t ir Gausa process, tad $\epsilon_1, \dots, \epsilon_n$ log - ticamības funkcija ir

$$L_n = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma_\epsilon^2 - \frac{1}{2} \sum_{t=1}^n \left(\frac{\epsilon_t}{\sigma_\epsilon} \right)^2. \quad (6.8)$$

Ja zināmas tikai ierobežota skaita pagātnes vērtības, tad $\epsilon_2, \dots, \epsilon_n$ var novērtēt ar

$$u_t = X_t - \sum_{s=1}^{t-1} b(s) X_{t-s}, \quad (t = 2, \dots, n).$$

Tuvinātu log - ticamības funkciju var definēt, formulā (6.8) ϵ_t aizstājot ar u_t . Parametriskā modelī koeficienti $b(s)$ ir atkarīgi no parametru vektora η^0 . Tuvinātu $\theta^0 = (\sigma_\epsilon^2, \eta)$ *MLE* iegūst, maksimizējot tuvinātu log - ticamības funkciju attiecībā pret θ . Pieņem, ka

$$r_t(\theta) = \frac{u_t(\eta)}{\sqrt{\theta_1}},$$

$$\dot{r}_t(\theta) = \left(\frac{\partial}{\partial \theta_1} r_t(\theta), \dots, \frac{\partial}{\partial \theta_M} r_t(\theta) \right)^t,$$

un

$$\dot{u}_t(\eta) = \left(\frac{\partial}{\partial \eta_1} u_t(\eta), \dots, \frac{\partial}{\partial \eta_{M-1}} u_t(\eta) \right)^t.$$

Tuvināto $\hat{\theta}$ *MLE* iegūst, minimizējot summu no funkcijas

$$n \log \theta_1 + \sum_{t=2}^n r_t^2(\theta)$$

attiecībā no θ . Ņemot pirmās kārtas parciālos atvasinājumus no θ , iegūstam M nelineāru vienādojumu sistēmu

$$\sum_{t=2}^n \{r_t(\theta)\dot{r}_t(\theta) - E[r_t(\theta)\dot{r}_t(\theta)]\} = 0,$$

kuru var sadalīt divās daļās

$$\sum_{t=2}^n \dot{u}_t(\eta)u_t(\eta) = 0$$

un

$$\sigma_\epsilon^2 = \frac{1}{n-1} \sum_{t=2}^n u_t^2(\theta).$$

6.4. Diskrēts Whittle novērtējums

Whittle tuvinātā MLE iegūšana var būt laukietilpīgs process, sevišķi, ja ir lielu izmēru izlases un θ ar lielu dimensiju. Integrāli (6.6) var aizvietot ar Rīmaņa summu

$$\hat{\alpha}(k) = 2 \frac{1}{(2\pi)^2} \sum_{j=1}^m \frac{1}{f(\lambda_{j,m})} e^{ik\lambda_{j,m}} \frac{2\pi}{m},$$

kur

$$\lambda_{j,m} = \frac{2\pi j}{m}, (j = 1, \dots, m^*) \quad (6.9)$$

un m^* ir $(m-1)/2$ veselā daļa. Ja $I(\lambda)$ ir procesa X_t periodigramma (4.4), tad (6.7) var izteikt kā

$$L_W = \frac{1}{2\pi} \left[\int_{-\pi}^{\pi} \log f(\lambda; \theta) d\lambda + \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda)} d\lambda \right].$$

No (6.9) iegūstam tuvinājumu

$$\tilde{L}_W(\theta) = 2 \frac{1}{2\pi} \left[\sum_{j=1}^{m^*} \log f(\lambda_{j,m}; \theta) \frac{2\pi}{m} + \sum_{j=1}^{m^*} \frac{I(\lambda_{j,m})}{f(\lambda_{j,m}; \theta)} \frac{2\pi}{m} \right].$$

Turklāt, gadījumā, kad $\theta = (\theta_1, \eta)$ definēts tāds, ka

$$f(\lambda; \theta) = \theta_1 f(\lambda; \theta^*),$$

$$\int_{-\pi}^{\pi} \log f(\lambda; \theta^*) d\lambda = 0,$$

un izvēlas $m = n$, novērtējumu $\hat{\theta}$ iegūst, minimizējot funkciju

$$\tilde{Q}(\eta) = \sum_{j=1}^{m^*} \frac{I(\lambda_{j,n})}{f(\lambda_{j,n}; \theta^*)}$$

pēc η un pielīdzinot

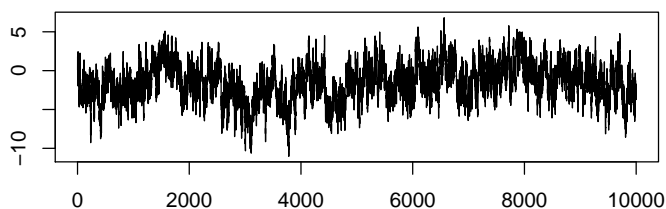
$$\hat{\sigma}_\epsilon^2 = 2\pi\hat{\theta}_1 = \frac{4\pi}{n}\tilde{Q}(\hat{\eta}).$$

7. Praktiskais pielietojums

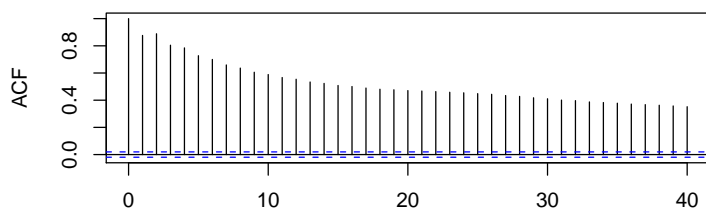
Apskatīsim parametra H (vai ekvivalenti - parametra d) novērtējumu metodes piemēros. Aprēķiniem izmantota datorprogramma R. Tajā pieejamas speciālas ilglaicīgās atmiņas procesu paketes "fArma" [12], "fracdiff" un "longmemo", kurās iebūvētas funkcijas FARIMA(p, d, q), fBm, fGn procesu simulēšanai pēc dažādām pieejām (Berana, Durbina - Levinsona algoritma u.c.) un parametra H novērtēšanai.

Piemērota FARIMA modeļa atrašanai, izmanto procesa vienādojumu formā (4.2). Tad, zināmai d vērtībai, aprēķina $\tilde{X}_t = (1 - B)^d X_t$ (\tilde{X}_t dēvē arī par "frakcionāli diferencētām rindām"[13]), pielāgo \tilde{X}_t atbilstošu ARMA procesu un veic atlikumu pārbaudi.

Piemērs 1. Apskatīsim FARIMA modeļa atrašanu simulētiem datiem. Izmantojot programmā R pieejamo funkciju "fracdiff.sim", ģenerē procesu FARIMA (2, 0.4, 0), kur $d = 0.4$, $\phi_1 = 0.1$ un $\phi_2 = 0.5$, apjomā $n = 10000$. Meklē piemērotu FARIMA modeli. Šajā gadījumā zinām, kā dati simulēti, tāpēc izvēlamies AR procesu ar kārtu 2 un, ar komandas "fracdiff" palīdzību, iegūstam vislielākās ticamības metodes koeficientu novērtējumus.



(a)



(b)

3. att. FARIMA(2, 0.4, 0) laikrinda (a) un ACF (b)

1. tabula Parametru novērtējumi

Parametrs	Novērtējums	St.kļūda
d	0.40625	0.02010
ϕ_1	0.08482	0.02026
ϕ_2	0.50274	0.01194

Zinot, ka $(1 - B)^d$ var izteikt formā

$$(1 - B)^d = 1 - dB + \frac{d(d-1)}{2!}B^2 - \frac{d(d-1)(d-2)}{3!}B^3 + \dots$$

un izmantojot iegūto $\hat{d} = 0.406251572$, līdz zināmam lagam (šajā gad. - 30) ar frakcionālo diferencu operatoru tiek aprēķināts \tilde{X}_t

$$\tilde{X}_t = X_t - 0.406251572X_{t-1} - 0.120605616X_{t-2} - \dots - 0.002283513X_{t-30}.$$

Tad, frakcionāli diferencētajam procesam \tilde{X}_t , piemērotāko AR modeli atrod pēc Akaike kritērija (AIC). Iebūvētā komanda "ar(x)" izvēlas modeli ar minimālo

$$AIC = 2k - 2 \ln(L)$$

vērtību, kur k - modeļa novērtēto parametru skaits, L - ticamības funkcija.

Šajā gadījumā, tas ir AR(2) ar novērtējumiem $\hat{\phi}_1 = 0.090$ un $\hat{\phi}_2 = 0.505$, $\sigma^2 = 0.993$.

Novērtējuma koeficienti praktiski neatšķiras no pirmajā modelī novērtētajiem.

Nepieciešams pārbaudīt, ka atlikumi ϵ_t ir neatkarīgi, vienādi sadalīti gadījuma lielumi.

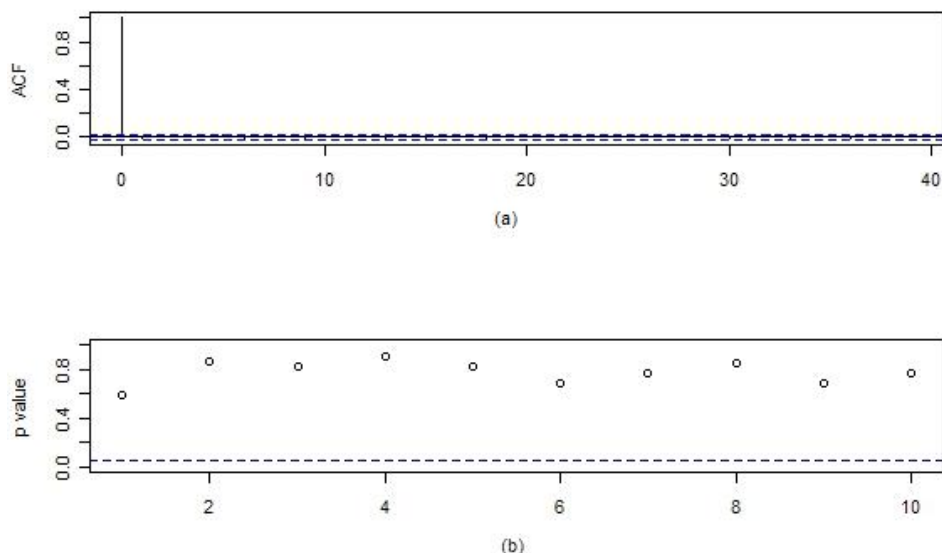
Izmanto Ljung - Box statistiku

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k},$$

kur n - izlases apjoms, $\hat{\rho}_k$ - autokorelācija k - tajā lagā, h - pārbaudāmo lagu skaits.

Hipotēze - dati ir neatkarīgi, alternatīva - nav neatkarīgi. Pēc iegūtajām p - vērtībām 4.

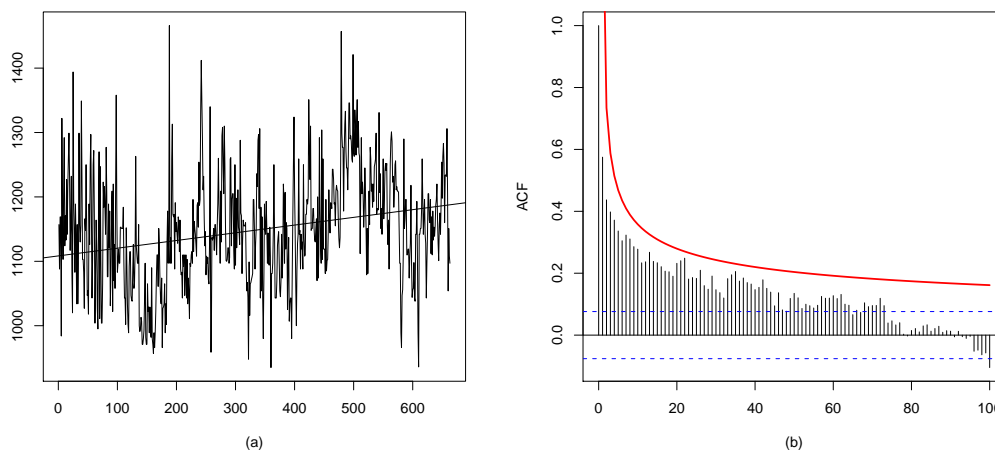
hipotēzi nenoraida, arī atlikumu ACF liecina par nekorelētiem datiem.



4. att.: Diferenču procesam pielāgotā AR(2) atlikumu ACF (a) un Ljung - Box statistikas p - vērtības(b).

Piemērs 2. Nīlas minimuma dati

Veiksim H novērtēšanu un modeļa izvēli Nīlas minimuma datiem [14]. Tie raksturo ikgadējo minimālo Nīlas ūdens līmeni (mm) no 622. līdz 1284. gadam. Dati satur 663 mērījumus, kas veikti Rodas salā, netālu no Kairas un tos bieži izmanto ilglaicīgās atmiņas procesu analīzes piemēros. Šiem datiem literatūrā izpētīts, ka parametra sagaidāmā vērtība $H \approx 0.84$ ($d \approx 0.34$) (J. Beran [3] 118.lpp)



5. att.: Nīlas minimuma laikrinda un trendu raksturojošā taisne (a), ACF un teorētiskā ACF (nepārtrauktā līnija) pie $H = 0.84$ (b)

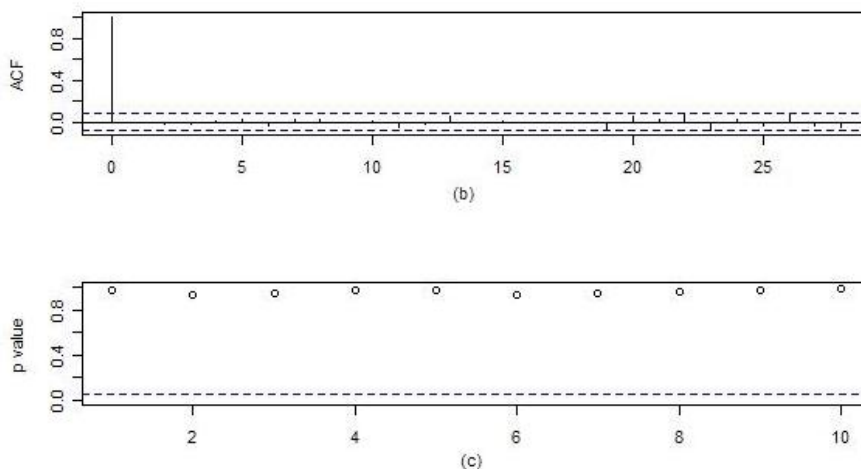
Attēlā 5. redzama LRD modeļiem raksturīgā lēnā korelāciju dilšana. Pēc Nīlas minimuma laikrindu attēla secināts, ka datiem varētu piemist trends. Novērtētie lineārā

trenda koeficienti ir nozīmīgi, tādēļ turpmāk strādā ar datiem, no kuriem atņemts trends. Izmantojot komandu "adf.test", pēc paplašinātā Dickey - Fuller testa (ADF) [15] ar p - vērtību=0.01, noraidām hipotēzi par vienības sakni. Tātad, process ir stacionārs. Pēc Akaike kritērija, izvēlas procesu AR(3). Tabulā 2. atbilstošie parametru novērtējumi.

2. tabula Koeficientu novērtējumi modelim FARIMA(3, d , 0)

	Novērtējums	St.kļūda
d	0.366277	0.029298
ϕ_1	0.055489	0.049182
ϕ_2	-0.038626	0.041963
ϕ_3	0.006719	0.039730

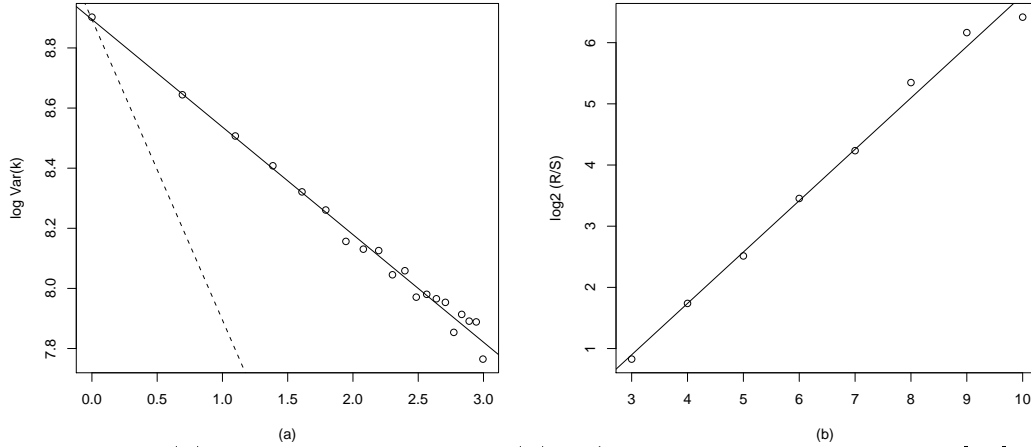
Iegūtajai \hat{d} vērtībai aprēķina diferencu operatoru un no tā iegūstamās fracionāli diferencētās rindas. Ar komandu "ar(x)" pēc AIC tiek izraudzīts modelis AR(1) ar $\phi = 0.082$. Pārbaudot Ljung - Box testa rezultātus un spriežot pēc ACF attēla, atlikumi ir nekorelēti. Parametrs $\hat{d} = 0.3420386$, līdz ar to $\hat{H} = 0.84204$.



6. att. Procesa AR(1) atlikumu ACF (b) un Ljung - Box statistikas p - vērtības (c)

Salīdzināsim šo H novērtējumu ar piektajā nodaļā aprakstītajām R/S statistikas un dispersiju grafika metodēm.

Attēlā 7. (a), redzami dispersiju grafika rezultāti, bloku skaits $k = 20$. Novērojama ilglaicīgās atmiņas klātbūtne - nozīmīga nobīde no taisnes ar slīpumu -1 . Koeficienta novērtējums ir -0.3579410 . Atbilstošais parametra novērtējums $\hat{H} = 0.5(-0.3579410) + 1 = 0.8210295$. Formulu R/S statistikas iegūšanai skatīt arī materiālā [16]. Iegūtie $\log(R/S)$



7. att. (a) Dispersiju grafiks un (b) R/S metodes grafiks datiem [14]

attiecībā pret $\log(k)$ attēloti 7. (b). Šajā gadījumā taisnes slīpuma novērtējums ir pozitīvs skaitlis, kas raksturo $\hat{H} = 0.839833$.

Izmantojot trīs iepriekš aplūkotās metodes, esam ieguvuši samērā tuvus novērtējumus. Dispersiju grafiku novērtējums ir visneprecīzākais. Tomēr jāpiezīmē, ka abu grafisko metožu rezultātus ietekmē k izvēle. Programmā R pieejamas 9 funkcijas H vērtības noteikšanai. R/S , dispersiju grafiku, DFA un Whittle algoritms jau apskatīts teorijas sadaļā. Periodigrammu un modificētās periodigrammu metodes arī ir grafiska pieeja, kas balstās uz procesa periodigrammu (4.4) asimptotisko uzvedību. Plašāk par iebūvētajām metodēm skatīt [17].

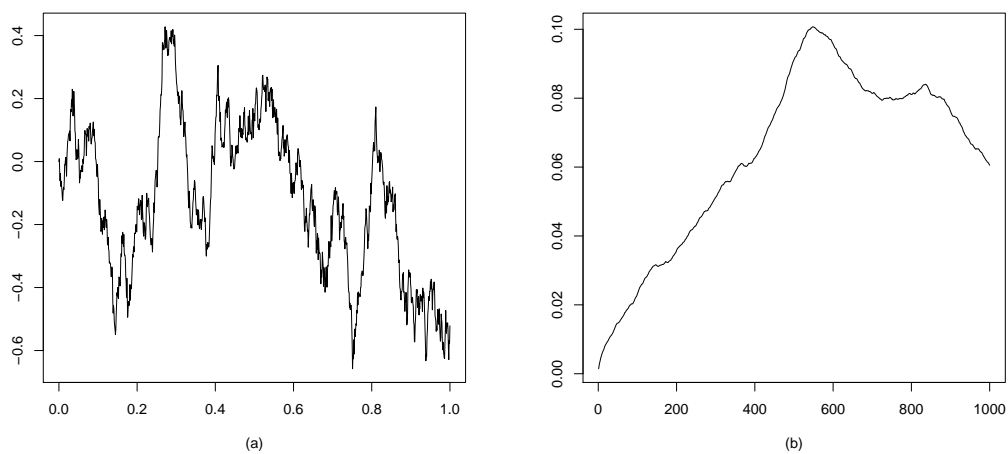
3. tabula R iebūvēto funkciju parametra H novērtējumi datiem [14]

metode	R komanda	\hat{H}
R/S statistika	rsFit	0.8394554
Dispersiju grafiki	aggvarFit	0.8466963
DFA	pengFit	0.8962124
Periodigrammu met.	perFit	0.9926786
Modificētā per.met.	boxperFit	0.7722782
Whittle novērtējums	whittleFit	0.837425

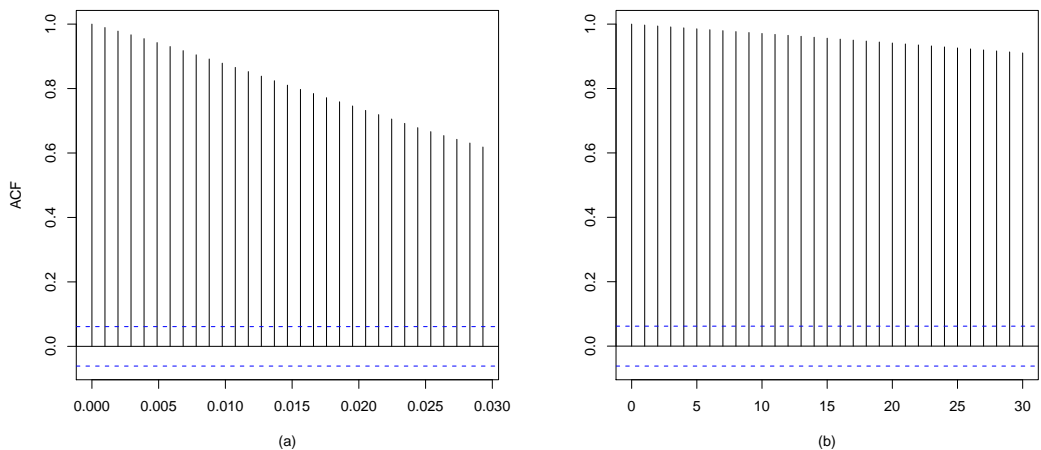
Piemērs 3. Dots fizikālos eksperimentos izstrādās process, kurā pēta daļiņu kustību īpašā šķīdumā. Uzdevums ir noskaidrot, vai tā ir Brauna kustība, analizēt uzvedību. Turpmāk datus apzīmēsim - "Kustība". Par procesu un tā izstrādi vairāk informācijas pieejams M. Kozlovskā maģistra darbā "Ūdeņraža izdalīšanās kinētikas pētījumi uz dažādu sastāvu un struktūras elektrodiem" [18].

Salīdzināšanas nolūkā, ar datorprogrammu R simulēsim sekojošus procesus:

- Brauna kustība (turpmāk - "Brauna"),
- Neatkarīgu, vienādi sadalīti gadījuma lielumu process (apz. - "iid"),
- AR(1) ar $\phi = 0.3, \phi = -0.3, \phi = 0.8, \phi = -0.8$.



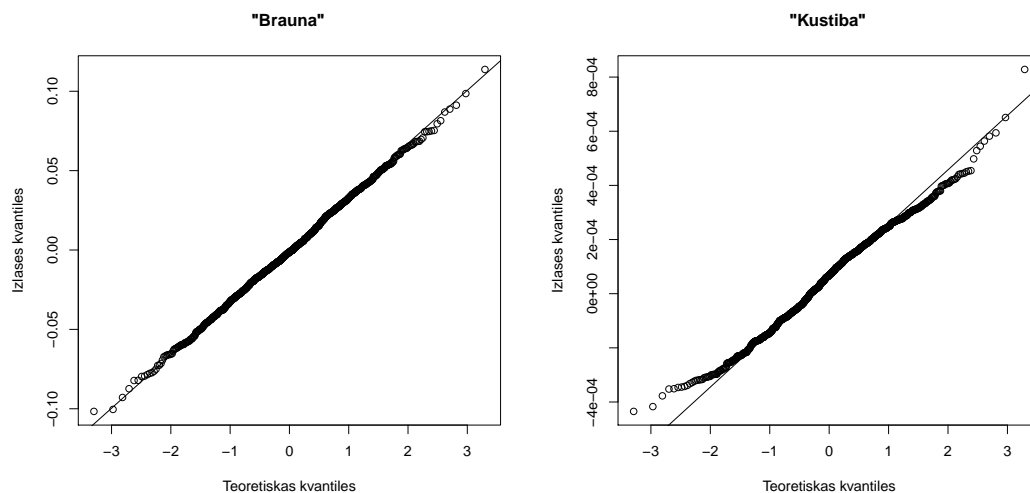
8. att. Process "Brauna" (a); process "Kustība" (b)



9. att. Procesa "Brauna" ACF, procesa "Kustība" ACF

No pirmajā nodaļā minētās definīcijas zināms, ka Brauna kustība ir process, kuram eksistē neatkarīgi, normāli sadalīti pieaugumi. Gadījumā, kad parametrs $H = 0.5$, tā ir standarta Brauna kustība, ja $0.5 < H < 1$ - frakcionālā Brauna kustība.

Analīzes pirmais solis būtu pārbaudīt "Kustība" pieaugumu normalitāti. Tam izmantosim Shapiro - Wilk testu [19] un kvantiļu - kvantiļu grafikus.



10. att.: Procesa "Brauna" pieaugumu kvantiļu - kvantiļu grafiks; procesa "Kustība" pieaugumu kvantiļu - kvantiļu grafiks

4. tabula Shapiro - Wilk testa rezultāti

Pieaugumu process	testa vērtība W	p - vērtība
"Brauna"	0.9993	0.9692
"Kustība"	0.9932	0.0001676

No kvantiļu - kvantiļu grafikiem 10. redzams, ka normālā sadalījuma kvantile labi aproksimē procesa "Brauna" pieaugumus, arī Shapiro - Wilk testa rezultātā 4., nevaram noraidīt hipotēzi par pieaugumu normālo sadalījumu. Procesa "Kustība" pieaugumu normalitāti noraidām.

Izmantojot Hursta R/S statistiku, iegūtie procesu parametra H novērtējumi 5.

5. tabula R/S statistikas novērtējums parametram H

Dati	\hat{H}
"Kustība"	1.023200
"Brauna"	1.031004
"iid"	0.5201048
AR($\phi = 0.3$)	0.5122390
AR($\phi = -0.3$)	0.42372275
AR($\phi = 0.8$)	0.63974779
AR($\phi = -0.8$)	0.29773241
"Brauna" pieaugumi	0.5595665
"Kustība" pieaugumi	0.983231

Redzams, ka abu nestacionāro procesu - "Kustība" un "Brauna" gadījumā, H vērtība pārsniedz pēc definīcijas apskatīto augšējo robežu 5.. Procesu "iid" parametra novērtējums atbilst Teorēmas 11 (19.lpp) par neatkarīgu, vienādi sadalītu datu uzvedību- ja spēkā centrālā robežteorēma, statistika $k^{-\frac{1}{2}}Q$ konverģē uz nedeģenerētu gadījuma lielumu (Mandelbrot 1975 [20]). Tas nozīmē, ka statistikas $Q(t, k)$ vērtības šiem datiem izkliedētas ap taisni ar slīpumu ≈ 0.5 .

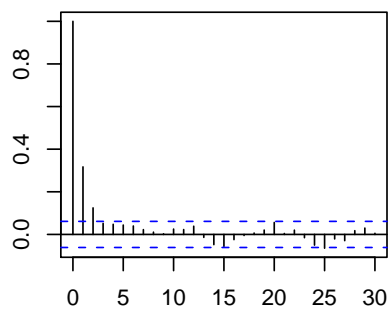
Tipiski ir arī autoregresīvā pirmās kārtas procesa rezultāti (11.). Korelāciju

$$\rho(i, j) = \rho(i - j) = a^{|i-j|}$$

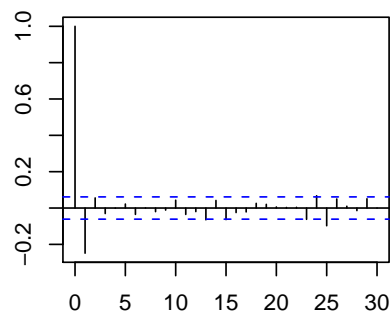
summas ir lielas pie lielām a vērtībām. Tāpēc $\rho(k) = a^{|k|}$ norāda uz augstu korelāciju pie pirmā laga [3].

6. tabula Nestacionaritātes pārbaude ar ADF testu

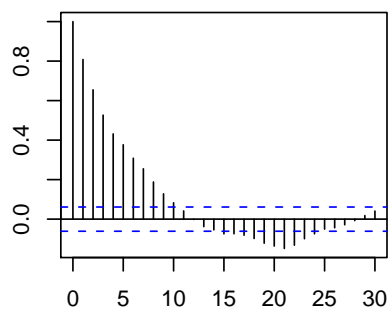
Process	testa vērtība	p - vērtība
"Kustība"	0.2746	0.99
"Brauna"	-2.8161	0.2328
"Kustība" pieaugumi	-4.1303	0.01
"Brauna" pieaugumi"	-8.9624	0.01



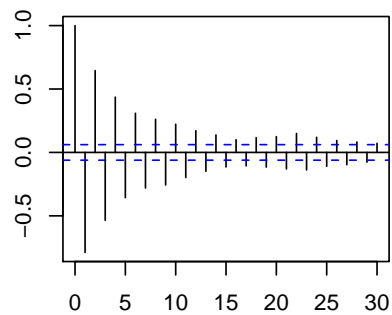
(a)



(b)



(c)



(d)

11. att. Procesas AR(1) ACF pie: $\phi = 0.3$ (a), $\phi = -0.3$ (b) $\phi = 0.8$ (c), $\phi = -0.8$ (d)

Stacionāram procesam no Brauna kustības ("Brauna" pieaugumi 5.) H tuvs teorētiskajai Brauna kustības H vērtībai ≈ 0.5 . Stacionāram procesam no "Kustība" ("Kustība" pieaugumi 5.) ir liela $H = 0.983231$ vērtība, kas arī bija sagaidāms pēc ACF attēla 9..

Secinājumi

Diplomdarbā aprakstīti procesi, kam piemīt ilglaicīgā atmiņa, to raksturojošais Hursta parametrs un tā novērtēšanas metodes. Simulētiem un reāliem datiem veikta piemērota FARIMA modeļa atrašana. Veikta ilglaicīgās atmiņas parametra H novērtēšana, izmantojot R pieejamās funkcijas, kā arī pēc dispersiju grafiku un R/S statistikas izveidotās programmas. Simulētiem datiem iegūtie rezultāti ir samērā tuvi zināmajām vērtībām. Veicot daudzkārtējas datu simulācijas, novērots, ka metodes būtiski ietekmē pētāmā procesa apjoms un izvēlēto bloku skaits. Pieejamas dažādas publikācijas, labāko metožu izvērtēšanai, piemēram, Pilar Grau - Carles (2005) [21] un Murphy un Izzeldin (2009) [22]. Novērtēt Hursta parametru ir būtiski, lai izvēlētos atbilstošu modeli datu analīzei. Pielietojot frakcionālo diferencēšanu, modeļu analīzi var reducēt uz jau pazīstamo ARMA modeļu analīzi.

Analizēta fizikālos eksperimentos iegūtas kustības uzvedība. Secināts, ka tā nav Brauna kustība.

Izmantotā literatūra un avoti

- [1] Sutcliffe J. V. *Obituary: Harold Edwin Hurst*, volume 24, pages 539–541. 1979.
- [2] Simaika Y.P. Hurst H.E., Black R.M. *The future conservation of the Nile, Cairo*, volume 7. 1946.
- [3] Beran J. *Statistics for Long - Memory Processes*. Chapman & Hall/CRC, New York, 1994.
- [4] Samorodnitsky G. *Long Range Dependence*. Now Publishers,INC., Hanover, 2007.
- [5] Peres Y. Morters P. *Brownian Motion*. 2006.
- [6] Palma W. *Long - Memory Time Series: Theory and Methods*. John Wiley & Sons,INC., New Jersey, 2007.
- [7] Taqqu M.S. Doukhan P., Oppenheim G. *Theory and Applications of Long Range Dependence*. Birkhauser Boston,INC., New York, 2003.
- [8] Hosking J.R.M. *Some models of persistence in time series: theory and practice 1*, pages 641–654. 1982.
- [9] Inoue A. *Asymptotic behavior for partial autocorrelation functions of fractional ARIMA processes*, volume 12, pages 1471–1491. 2002.
- [10] Mandelbrot B.B. *Statistical Methodologyfor nonperiodic cycles: From the covariance to R/S analysis*, volume 1, pages 259–290. 1972.
- [11] Havlin S. Simons M. Stanley H.E. Goldberger A.L. Peng C.K., Buldyrev S.V. *Mosaic organization of DNA nucleotides*, volume 49, pages 1685–1689. 1994.
- [12] <http://cran.r-project.org/web/packages/fArma/fArma.pdf>.

- [13] Metcalfe A.V. Cowpertwait P.S.P. *Introductory Time Series with R*. Springer, New York, 2009.
- [14] <http://www.massey.ac.nz/~pscowper/pub/data/nile.dat>.
- [15] http://en.wikipedia.org/wiki/Augmented_Dickey-Fuller_test.
- [16] http://www.bearcave.com/misl/misl_tech/wavelets/hurst/index.html.
- [17] Porter-Hudak S. Geweke J. *The estimation and application of long memory time series models*, volume 4, pages 221–238. 1983.
- [18] Kozlovskis M. *Ūdeņraža izdalīšanās kinētikas pētījumi uz dažādu sastāvu un struktūras elektrodēm*. 2010.
- [19] http://en.wikipedia.org/wiki/Shapiro-Wilk_test.
- [20] Mandelbrot B.B. *Limit theorems of the self - normalized range for weakly and strongly dependent processes*, volume 31, pages 271–285. 1975.
- [21] Grau Carles P. *Tests of Long Memory: A Bootstrap Approach*, volume 25, pages 103–113. 2005.
- [22] Izzeldin M. Murphy A. *Bootstrapping long memory tests: Some Monte Carlo results*, volume 53, pages 2325–2334. 2009.

A Pielikums

Teorēma 14. Nepārtraukts attēlojums. Pieņem, ka $\{X_n\}_{n=1}^{\infty}$ ir gadījuma lielumu virkne un X ir gadījuma lielums, definēti vienā metriskā telpā \mathcal{X} . Pieņem, ka $X_n \xrightarrow{D} X$. Pieņem, ka \mathcal{Y} ir metriskā telpa un $g : \mathcal{X} \rightarrow \mathcal{Y}$. Definē $C_g = \{x : g \text{ ir nepārtraukts uz } x\}$. Pieņem, ka $P(X \in C_g) = 1$. Tad $g(X_n) \xrightarrow{D} g(X)$.

Definīcija 9. [6] Stohastisku procesu $\{y_t : t \in \mathbb{Z}\}$ sauc par *otrās kārtas jeb stacionāru vājā nozīmē*, ja eksistē simetriska autokovariāciju funkcija $\gamma(\cdot)$, kurai $\gamma(0) = \text{Var}(y_t) < \infty$, tāda, ka $\langle y_t, y_{t+h} \rangle = \gamma(h)$ visiem $t, h \in \mathbb{Z}$.

Definīcija 10. [6] Otrās kārtas procesa autokovariāciju funkcija $\gamma(\cdot)$

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} dF(\lambda),$$

kur funkcija F ir no labās puses nepārtraukta, nedilstoša, ierobežota $[-\pi, \pi]$ un izpildās nosacījums $F(-\pi) = 0$. F sauc par $\gamma(\cdot)$ *spektrālo sadalījumu*. Turpmāk, ja

$$F(\lambda) = \int_{-\pi}^{\lambda} f(\omega) d(\omega),$$

tad $f(\cdot)$ sauc par $\gamma(\cdot)$ *spektrālo blīvumu*.

Definīcija 11. Pieņem, ka $\{X_t : t \in T\}$ ir stohastisks process, kuram $\text{Var}(X_t) < \infty$ visiem $t \in T$. Tad $\{X_t\}$ *autokovariāciju funkcija* ir

$$\gamma_X(r, s) := \text{Cov}(X_r, X_s) = E[(X_r - E(X_r))(X_s - E(X_s))]$$

visiem $r, s \in T$.

Definīcija 12. [7] Funkciju $\Gamma(p)$ sauc par *gamma funkciju*, ja

$$\Gamma(p) = \begin{cases} \int_0^{\infty} t^{p-1} e^{-t} dt, & p > 0, \\ \infty, & p = 0, \\ p^{-1} \Gamma(1 + p), & p < 0 \end{cases}$$

Definīcija 13. Funkciju $B(x, y)$ sauc par *beta funkciju*, ja

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)}.$$

A Izveidoto programmu kods

1) Darbā izmantotās R paketes

```
library(fracdiff)
library(tseries)
library(stats)
library(fArma)
library(sde)
```

2) Nilas minimuma piemērota FARIMA modeļa atrašana

```
dati.nile<-na.omit(scan(file="nile.txt",sep=",",))
plot.ts(dati.nile)
regr<-lm(dati.nile~c(1:length(dati.nile)))
summary(regr)
plot.ts(dati.nile)
abline(regr$coef)
trends<-regr$coef[[1]] + regr$coef[[2]]*c(1:length(dati.nile))
nile.notrend<-trends-dati.nile
plot(nile.notrend,type="l")
adf.test(nile.notrend)
nile.fit<-fracdiff(nile.notrend,nar=1,nma = 0)
AIC(nile.fit)
nile.fit$d
summary(nile.fit)
n<-length(nile.notrend)
L<-30
d<-nile.fit$d
fdc<-d
fdc[1]<-fdc
for (k in 2:L) fdc[k] <-fdc[k-1] * (d+1-k)/k
y<-rep(0,L)
for (i in (L+1):n) {
  csm<-y.notrend[i]
  for (j in 1:L) csm<-csm + ((-1)^j)*fdc[j]*nile.notrend[i-j]
  y[i]<-csm
}
y<-y[(L+1):n]
karta.ar<-ar(y)
```

```

n.resid<-1+karta.ar$order
residuals<-karta.ar$res[n.resid:length(y)]
fit<-arima(x,c(1,0,0))
tsdiag(fit)
plot.ts(dati.nile)
abline(regr$coef)
acf(y)
acf(nile.notrend)
acf(z)
plot.ts(y)

```

3) Dispersiju grafiku metode

```

log.disp<-function(bloku.g)
{
bloku.sk<-floor(n/bloku.g)
rez<-c()
for (j in 1:bloku.sk )
{
rez[j]<-mean(dati.nile[((j-1)*bloku.g+1):(j*bloku.g)])
}
log(var(rez))
}
log.disp2<-Vectorize(log.disp)
x.vec<-c(1:100)
y.dati<-log.disp2(x.vec)
x.dati<-log(x.vec)
plot(x.dati,y.dati, xlab="a", ylab="log Var(k)")
lin.reg<-lm(y.dati~x.dati)
lin.reg$coef
abline(lin.reg$coef)
abline(lin.reg$coef[1],-1,lty=2)
H<-0.5*lin.reg$coef[2]+1
H

```

4) R/S metode

```

f.range<-function(n,tau,dati)
{
sum(dati[1:n]-mean(dati[1:tau]))
}
f.range2<-Vectorize(f.range,vectorize.args="n")

```

```

RS.range<-function(dati)

```

```

{
n<-length(dati)
rez<-max(f.range2(1:length(dati),length(dati),dati))-min(f.range2(1:length(dati),length(dati),dati))
rez/sd(dati)

```

```

}
n<-length(dati)
RS.ave<-function(dati,m)
{
  nn<-length(dati)/m
  rez<-c()
  for (i in 1:m)
  {
    rez[i]<-RS.range(dati[((i-1)*nn+1):(i*nn)])
  }
  mean(rez)
}
mean(c(RS.range(dati[1:(n/2)]),RS.range(dati[(n/2+1):n])))
RS.ave<-Vectorize(RS.ave,vectorize.args="m")
x.data<-3:10
y.data<-sort(log(RS.ave(dati,c(1,2,4,8,16,32,64,128)),base=2))
plot(x.data,y.data,xlab="b",ylab="log2(R/S)")
fit<-lm(y.data~x.data)
fit$coef
abline(fit$coef[[1]],fit$coef[[2]])
h.nov<-lm(y.data~x.data)$coef[[2]]

```

5) Iebūvētās H novērtēšanas metodes

```

## R/S tests
rsFit(x, levels = 50, minnpts = 3, cut.off = 10^c(0.7, 2.5),doplot = TRUE)

## Dispersiju grafiku metode
aggvarFit(x, levels = 50, minnpts = 3, cut.off = 10^c(0.7, 2.5),doplot = TRUE)

## Diferencētā dispersiju grafiku metode
diffvarFit(x, levels = 50, minnpts = 3, cut.off = 10^c(0.7, 2.5),doplot = TRUE)

## Aggregated absolūto vērtību metode
absvalFit(x, levels = 50, minnpts = 3, cut.off = 10^c(0.7, 2.5), moment = 2,doplot = TRUE)

## Higuchi fraktāļu dimensiju metode
higuchiFit(x, levels = 50, minnpts = 2, cut.off = 10^c(0.7, 2.5),doplot = TRUE)

## Penga DFA metode
pengFit(x, levels = 50, minnpts = 3, cut.off = 10^c(0.7, 2.5),method ="mean",doplot = TRUE)

## Periodogrammu metode
perFit(x, cut.off = 0.1, method = c("per", "cumper"),doplot = TRUE)

## Pārveidotā periodogrammu metode
boxperFit(x, nbox = 100, cut.off = 0.10,doplot = TRUE)

## Whittle novērtējums
whittleFit(x, order = c(1, 1), subseries = 1, method = c("fgn", "farma"))

## Vaveletu metode
waveletFit(x, length = NULL, order = 2, octave = c(2, 8),doplot = TRUE)

## Procesa FARIMA simulācija
farimaSim(n = 1000, model = list(ar = c(0.5, -0.5), d = 0.3, ma = 0.1),method = c("freq", "time"))

## Procesa fBm simulācija
fbmSim(n = 100, H = 0.7, method = c("mvn", "chol", "lev", "circ", "wave"),waveJ = 7,doplot = TRUE,fgn = FALSE)

```


Diplomdarbs "Ilglaicīgās atmiņas procesi" 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: Ilze Rēvalde

(paraksts) (datums)

Rekomendēju darbu aizstāvēšanai.

Vadītājs: doc. Dr.math. Jānis Valeinis

(paraksts) (datums)

Recenzents: doc. Dr.math. Nadežda Siņenko

(paraksts) (datums)

Darbs iesniegts Matemātikas nodaļā _____
(datums)

(darbu pieņēma)

Darbs aizstāvēts diplomdarbs valsts pārbaudījuma komisijas sēdē

_____ prot. Nr. _____, vērtējums _____
(datums)

Komisijas sekretārs/-e: Ingrīda Uļjane _____
(paraksts)