LATVIJAS UNIVERSITĀTE
FIZIKAS UN MATEMĀTIKAS FAKULTĀTE
MATEMĀTIKAS NODAĻA

NEIMAŅA TESTS SALIKTĀM HIPOTĒZĒM

MAĢISTRA DARBS

Autors: **Anna Jansone**

Stud. apl. aj08070

Darba vadītājs: LU docents Dr. Math. Jānis Valeinis

RĪGA 2013

Anotācija

Maģistra darbā tiek apskatīts Neimaņa tests saliktām hipotēzēm, kuru izmanto datu ekponencialitātes un normalitātes pārbaudei. Darba teorētiskajā daļā tiek analizēta Neimaņa testa uzbūve saliktām hipotēzēm neatkarīgiem datiem, kā arī veikta Neimaņa testa modifikācija atkarīgiem procesiem. Praktiskajā daļā tiek veiktas Monte Karlo simulācijas Neimaņa testam saliktām hipotēzēm neatkarīgiem datiem, iegūstot robežsadalījumus, kritiskās vērtības un jaudas pie dažādām alternatīvām, kā arī veiktas simulācijas Neimaņa testam saliktām hipotēzēm AR(1) procesam, lai iegūtu robežsadalījumu. Darbam pievienotas programmēšanas valodā *R* izveidotās programmas un programmā *Maple* veiktie aprēķini.

Atslēgas vārdi: Nogludināšanas tests, Neimaņa tests, AR(1), Švarca selekcijas kritērijs, datu nosakošā metode

Abstract

This work is devoted to Neyman smooth test for composite hyphotheses used for testing exponentiality and normality. In theoretical part of the work the Neyman smooth test for composite hypotheses for independent data is analyzed and the modification of Neyman smooth tests for dependent processes ir made. In practical part of the work Monte Carlo simulations to Neyman smooth test for composite hypotheses for independent data are made in order to get marginal distributions, critical values and empirical powers. Monte Carlo simulations for Neyman smooth test for composite hypotheses for AR(1) process are made.

Keywords: Smooth test, Neyman's tests, AR(1), Schwartz's rule, data driven procedure

Saturs

Ievads 2

1. Teorētiskā daļa 4
- 1.1. Neimaņa tests vienkāršām hipotēzēm 4
- 1.2. Neimaņa tests saliktām hipotezēm 6
- 1.3. Neimaņa statistikas izveduma teorētiskā analīze 12
- 1.4. Neimaņa tests atkarīgiem novērojumiem 17
  - 1.4.1. AR(1) process 18
  - 1.4.2. Neimaņa testa modifikācija AR(1) procesam 18

2. Praktiskā daļa 23
- 2.1. Neimaņa testa implementācija programmēšanas valodā R 23
- 2.2. Neimaņa tests saliktām hipotēzēm neatkarīgiem datiem 25
  - 2.2.1. Robežsadalījums 25
  - 2.2.2. Kritiskās vērtības 26
  - 2.2.3. Alternatīvie sadalījumi un simulētās jaudas eksponenciālajam sadalījumam 27
  - 2.2.4. Alternatīvie sadalījumi un simulētās jaudas normālajam sadalījumam . 28
- 2.3. Neimaņa tests saliktām hipotēzēm AR(1) procesam 29

Secinājumi 32

Izmantotā literatūra un avoti 33

A Neimaņa tests saliktām hipotēzēm neatkarīgiem datiem 35

B Neimaņa tests saliktām hipotēzēm AR(1) procesam 38

C Programma *Maple* veiktie aprēķini 40

Ievads

Viena no galvenajām problēmām matemātiskajā statistikā ir pārbaudīt dotās izlases atbilstību noteiktam sadalījumam.

Visbiežāk statistikā lietotie testi ir Karla Pīrsona ieviestais Hī kvadrāta tests (ieviests 1900. gadā), Kolmogorova - Smirnova tests (ieviests 1933. gadā) un Krāmera - von Mises tests (ieviests 1928. gadā). Šos testus var pielietot gan gadījumos, kad sadalījuma parametri ir zināmi (vienkāršās hipotēzes), gan gadījumos, kad sadalījuma parametri nav zināmi (saliktās hipotēzes). Saliktu hipotēžu gadījumos šo testu jaudas nav lielas [1], tādēļ to vietā eksponencialitātes pārbaudei iesaka izvēlēties Gini statistiku, bet normalitātes pārbaudei izvēlēties Šapiro-Vilkes testu. Kā laba alternatīva šiem testiem ir Neimaņa nogludināšanas tests.

1937. gadā Neimanis [2] ieviesa nogludināšanas testu vienmērīgā sadalījuma intervālā [0,1] pārbaudei. Aplūkosim vienkāršu hipotēzi vienmērīgā sadalījuma [0,1] pārbaudei

$$H_0: \quad X \sim v.s.[0,1] \quad \text{pret} \quad H_1 :\nsim v.s[0,1]. \tag{0.1}$$

Ja $X_1, X_2, \ldots, X_n \sim F$, kur F ir datu sadalījuma funkcija, tad lietojot transformāciju $Y_i = F(X_i)$, iegūsim, ka $Y_i \sim v.s.[0,1]$. Tāpēc jebkuru vienkāršu hipotēzi par datu sadalījuma veidu var reducēt uz hipotēzi 0.1.

Neimaņa tests balstās uz ortonormāliem polinomiem (Ležandra polinomiem). Līdz pat 1994. gadam nebija zināms, cik ortonormālie polinomi jāizvēlas, lai testam būtu maksimālā jauda. 1994. gadā Tadeušs Inglots [3] ieteica un pamatoja izvēlēties tikai pirmās no ortonormālās sistēmās komponentēm, H_0 hipotēzi pārbaudot ar maksimāli mazu komponenti . Tajā pašā gadā Terēza Ledvina ieteica ortonormālās sistēmās komponenšu skaitu izvēlēties, balstoties uz izlases datiem [4]. Šo pieeju var īstenot, izmantojot Švarca selekcijas likumu. Švarca Beijesa informācijas kritērijs (BIC) tiek izmantots, lai noteiktu ekponenciālā modeļa dimensiju (tādējādi arī ortonormālo polinomu skaitu). Šī pieeja tiks izmantota arī šajā darbā.

Švarca selekcijas likumu ar tā vienkāršākām modifikācijām [5] izmanto gan vienkāršām, gan saliktām hipotēzēm. Neimaņa testu saliktām hipotēzēm izmanto, lai noteiktu eksponencialitāti vai normalitāti. Literatūrā ir parādīts, kā izskatās Neimaņa tests saliktām hipotēzēm neatkarīgiem datiem [1], taču pagaidām nav sīkāk pētīts un modificēts Neimaņa tests saliktām hipotēzēm atkarīgiem datiem.

Darba galvenais mērķis ir modificēt Neimaņa testu atkarīgiem procesiem. Lai to sasniegtu, tika izvirzīti šādi darba uzdevumi: iepazīties ar atbilstošo statistisko literatūru par Neimaņa testu vienkāršām hipotēzēm un par Neimaņa testu saliktām hipotēzēm neatkarīgiem novērojumiem, izveidot programmu programmēšanas valodā **R**, kas realizē Neimaņa testu saliktām hipotēzēm neatkarīgiem novērojumiem, veikt Neimaņa statistikas izveduma toerētisko analīzi neatkarīgiem novērojumiem saliktu hipotēžu gadījumā, un, izmantojot veikto analīzi, modificēt Neimaņa testu saliktām hipotēzēm atkarīgiem procesiem.

Darbs sastāv no divām galvenajām nodaļām—darba teorētiskās un praktiskās daļas. Darba teorētiskajā daļā 1. apakšnodaļa veltīta Neimaņa testam vienkāršām hipotēzēm (tā uzbūvei un datu nosakošajai metodei, lai izvēlētos atbilstošo komponenšu skaitu Neimaņa testā). 2. apakšnodaļa veltīta Neimaņa testam saliktām hipotēzēm neatkarīgiem datiem, 3. apakšnodaļā tiek veikta Neimaņa testa saliktām hipotēzēm neatkarīgiem datiem analīze, savukārt 4. apakšnodaļā tiek veikta Neimaņa testa modifikācija atkarīgiem procesiem.

Darba praktiskajā daļā 1. apakšnodaļa veltīta Neimaņa testam saliktām hipotēzēm neatkarīgiem datiem. Šajā apakšnodaļā ar autores izveidoto programmu realizēts Neimaņa tests saliktām hipotēzēm, ir doti robežsadalījumi, kritiskās vērtības, kā arī veiktas Monte Karlo simulācijas eksponencialitātes un normalitātes pārbaudīšanai pie dažādām alternatīvām. Praktiskās daļās 2. apakšnodaļa veltīta Neimaņa testam AR(1) procesam, tajā ar autores izveidotu programmu realizēts Neimaņa tests AR(1) procesam un doti robežsadalījumi, savukārt 3. apakšnodaļā aprakstīts, kā tieši izveidotas programmas programmēšanas valodā **R**, kas realizē Neimaņa testu. Darba beigās pielikumā programmas **R** kods un programmā *Maple* veiktie matemātiskie aprēķini.

1. Teorētiskā daļa

1.1. Neimaņa tests vienkāršām hipotēzēm

$X_1, X_2, \ldots, X_n$ ir neatkarīgi un vienādi sadalīti gadījuma lielumi. Nogludinošais Neimaņa tests vienkāršām hipotēzēm pārbauda, vai dati ir sadalīti vienmērīgi intervālā $[0, 1]$.

$$H_0 : X \sim v.s.[0, 1] \text{ pret } H_1 : X \nsim v.s.[0, 1].$$

Tests noraida H_0 pie lielām šādas statistikas vērtībām

$$N_k = \sum_{j=1}^{k} \left(n^{-1/2} \sum_{i=1}^{n} \phi_j(X_i) \right)^2,$$

kur $\phi_1, \phi_2, \ldots, \phi_n$ ir ortonormāli Ležandra polinomi slēgtā intervālā $L_2([0, 1])$ un $\phi_0 \equiv 0$.

Pirmais, otrais un trešais Ležandra polinoms ir šādi:

$$\phi_1 = \sqrt{12}(x - 0.5),$$

$$\phi_2 = \sqrt{5}(6(x - 0.5)^2 - 0.5),$$

$$\phi_3 = \sqrt{7}(20(x - 0.5)^3 - 3(x - 0.5)).$$

Konstruējot Neimaņa testu, vienmērīguma pārbaude (H_0 hipotēze) tiek aizstāta ar parametrisku problēmu $H_0 : \theta = 0$ pret $H_1 : \theta \neq 0$ eksponenciālajā saimē ar kārtu k, kura ir

$$p_\theta(x) = \exp\{\theta \circ \phi(x) - \phi_k(\theta)\},$$

kur

$$\theta = (\theta_1, \theta_2, \ldots \theta_k), \quad \phi = (\phi_1, \phi_2, \ldots, \phi_k),$$

$$\phi_k(\theta) = \log \int_0^1 \exp\{\theta \circ \phi(x)\, \mathrm{d}x\},$$

kur ar $\circ$ apzīmē skalāro reizinājumu $\mathbb{R}^k$, bet ar "$\log$" apzīmē naturālo logaritmu.

Eksponenciālās saimes augšējā robeža ir $d(n)$. Pie augšējās robežās svarīgi ir noteikt eksponenciālās saimes kārtu k. Lai to izdarītu, tiek izmantots Švarca selekcijas likums jeb Švarca BIC (*Bayesian information criterion* [6]).

Definē

$$Y_n = (\hat{\phi}_1, \ldots, \hat{\phi}_k),$$

$$\hat{\phi}_j = n^{-1/2} \sum_{i=1}^{n} \phi_j(X_i),$$

$$L_s = n \sup_{\theta \in \Omega_s} \{Y_n \circ \theta - \phi_s(\theta)\} - \frac{1}{2} s \log n.$$

Švarca BIC kritērijā tiek izvēlēts modelis ar dimensiju S tādu, ka

$$S = \min\{j, 1 \leq j \leq K : L_j = \max_{1 \leq i \leq K} L_s\}$$

Tikai 1996. gadā Ledvina un Inglots [5] parādīja, ka $n \sup_{\theta \in \Omega_s} \{Y_n \circ \theta - \phi_s(\theta)\}$ ir ekvivalenta $\frac{1}{2} n(\|Y_n(\beta)\|)^2$, līdz ar to tika ieviesta Švarca likuma modifikācija, kuru ir vieglāk matemātiski aprēķināt,

$$S2 = \{k : 1 \leq k \leq K \quad n\|Y_n\|^2_{(k)} - k \log n \geq n\|Y_n\|^2_{(j)} - j \log n, \quad j = 1, \ldots, K\} \quad (1.1.1)$$

Līdz ar to Neimaņa testa statistika tiek definētā kā

$$N_S = \sum_{j=1}^{S} \left(n^{-1/2} \sum_{i=1}^{n} \phi_j(X_i) \right)^2,$$

kur S uzdots ar (1.1.1).

Lai pierādītu Neimaņa testa N_S atbilstību, ir jāzina testa uzvedība pie nulles hipotēzes un pie alternatīvām. Vispirms aplūkosim uzvedību pie nulles hipotēzes. Tiek aplūkots sākotnējais gadījums un S asimptotika, kur S var būt neatkarīgs. P_0 norāda, ka X_i ir vienmērīgi sadalīts intervālā $[0, 1]$.

$$P_0(S = 1) = 1 - \sum_{k=2}^{d(n)} P_0(S = k)$$

un, izmantojot, ka $\phi_1(0) = 0$,

$$P_0(S = k) \leq P_0(L_k \geq L_1) \leq P_0 \left(n \sup_{\vartheta \in \mathbb{R}^k} \{Y_n \circ \vartheta - \phi_k(\vartheta)\} \geq \frac{1}{2}(k-1) \log n \right).$$

Tālāk definē

$$V_k = \max_{1 \leq j \leq k} \sup_{x \in [0,1]} |\phi_j(x)|.$$

Ortonormāliem Ležandra polinomiem intervālā [0,1]

$$V_k = (2k+1)^{1/2}.$$

S asimptotika ir dota sekojošā teorēmā.

Teorēma 1. *Pieņemsim, ka*

$$\lim_{n\to\infty} d(n) V_{d(n)} (n^{-1}\log n)^{-1/2} = 0. \tag{1.1.2}$$

Tad

$$\lim_{n\to\infty} P_0(S=1) = 1.$$

Teorēma 2. *Ja 1.1.2, tad pie* H_0

$$N_S \longrightarrow_d \chi_1^2$$

Tagad aplūkosim uzvedību pie alternatīvām. Doti $X_1, X_2, ..., X_n$ neatkarīgi un vienādi sadalīti gadījuma lielumi,katrs no tiem ar sadalījumu P intervālā [0,1]. Pieņemsim, ka

$$E_P\phi_1(X) = \ldots = E_P\phi_{K-1}(X) = 0, \quad E_P\phi_K(X) \neq 0, \tag{1.1.3}$$

$K = K(P)$. N_S atbilstība pierādīta jebkurai alternatīvai formā 1.1.3. Pieņem, ka

$$\lim_{n\to\infty} \inf d(n) \geq K,$$

ja $\lim_{n\to\infty} = \infty$ pie fiksēta K.

Teorēma 3. *Ja 1.1.3, tad*

$$\lim_{n\to\infty} P(S \geq K) = 1.$$

Teorēma 4. *Ja 1.1.3, tad*

$$N_S \longrightarrow_d \infty, \quad \textit{kad} \quad n \longrightarrow \infty$$

Teorēmu pierādījumi doti [7] 1598.-1601. lpp.

1.2. Neimaņa tests saliktām hipotezēm

Ar Neimaņa testu tiek pārbaudīta datu piederība kādai blīvuma funkciju kopai, piemēram, eksponenciālajai vai normālajai. Pieņemsim, ka $X_1, X_2, \ldots, X_n$ ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar blīvuma funkciju $f(x)$. Tad pārbaudāmā nulles hipotēze ir

$$H_0 : f(x) \in \{f(x;\beta), \beta \in B\} \quad \text{pret} \quad H_1 : f(x) \notin \{f(x;\beta), \beta \in B\}. \tag{1.2.1}$$

kur

$$\{f(x;\beta), \beta \in B\}$$

ir dotā blīvuma funkciju kopa, piemēram, normālā vai eksponenciālā blīvuma funkciju kopa.

Eksponenciālajai blīvuma funkciju kopai:

$$f(x;\beta) = \begin{cases} 0, & x < 0 \\ \beta e^{-\beta x}, & x \geq 0 \end{cases}$$

Normālajais blīvuma funkciju kopai:

$$f(x;\beta) = (\sqrt{2\pi}\sigma)^{-1} \exp -\frac{(x-\mu)^2}{2\sigma^2},$$

kur $\beta = (\mu, \sigma)$.

Konstruējot Neimaņa testsu saliktām hipotēzēm, 1.2.1 hipotēze tiek aizstāta ar parametrisku problēmu $H_0 : \theta = (\theta_1, \ldots, \theta_k)' = (0, 0, \ldots, 0)$ pret $H_1 : \theta = (\theta_1, \ldots, \theta_k)' \neq (0, 0, \ldots, 0)$ eksponenciālā funkciju saimē ar kārtu k. Šīs saimes blīvuma funkcija ir

$$g_k(x;\theta,\beta) = \exp\left(\theta \circ \phi[F(x,\beta)] - \psi_k(\theta)\right) f(x,\beta), \tag{1.2.2}$$

kur

$\theta = (\theta_1, \ldots, \theta_k)'$, $\phi = (\phi_1, \ldots, \phi_k)'$. $\phi_1, \phi_2, \ldots, \phi_k$ ir ortonormāli Ležandra polinomi slēgtā intervālā $L_2([0,1])$, $\phi_0 \equiv 0$. Ar "′" apzīmē transponēto matricu vai vektoru, $\psi_k(\theta)$ ir normalizējošā konstante, kas definēta ar $\psi_k(\theta) = \log \int_0^1 \exp\left(\theta \circ \psi(y)\right) \mathrm{d}y$.

Lai novērtētu nezināmos parametrus β, tiek izmantota maksimālās ticamības funkcija. Maksimālās ticamības funkcija G neatkarīgiem un vienādi sadalītiem $X_1, X_2, \ldots, X_n$ ar blīvuma funkcijām (1.2.2) ir

$$G = \prod_{i=1}^{n} g_k(x;\theta,\beta) = \exp\{n[\theta \circ Y_n(\beta) - \psi_k(\theta)]\} \prod_{i=1}^{n} f(X_i;\beta)$$

Izmantojot šo funkciju (to maksimizējot)varam novērtēt parametru β. Eksponenciālajai blīvuma funkciju kopai novērtētais parametrs $\hat{\beta} = n(\sum_{i=1}^{n} X_i)^{-1}$, bet normālajai blīvuma funkciju kopai novērtētais parametrs
$\hat{\beta} = \left(\bar{X}, \sqrt{n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)$, kur $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$

Kad β ir nezināma un k ir fiksēts, pārbauda saliktu hipotēzi $H_0 : \theta = (\theta_1, \dots, \theta_k)' = (0, 0, \dots, 0)$ pret $H_1 : \theta = (\theta_1, \dots, \theta_k)' \neq (0, 0, \dots, 0)$ [2]. Statistika ir formā:

$$W_k = nY_n(\hat{\beta}) \left(I + R(\hat{\beta}) \right) Y_n(\hat{\beta}), \tag{1.2.3}$$

kur

$$Y_n(\beta) = (\bar{\phi_1}(\beta), \dots \bar{\phi_j}(\beta))' = n^{-1} \sum_{i=1}^{n} \left(\phi_1[F(X_i; \beta)], \dots, \phi_j[F(X_i; \beta)] \right)',$$

$$I_\beta = -E_\beta \frac{\partial}{\partial \beta_t} \phi_j \left[F(X, \beta) \right]_{t=1,\dots,q \quad j=1,\dots,k},$$

$$I_{\beta\beta} = -E_\beta \frac{\partial^2}{\partial \beta_t \beta_u} \log f(X, \beta)_{t=1,\dots,q \quad u=1,\dots,q},$$

$$R(\beta) = I'_\beta \left(I_{\beta\beta} - I_\beta I'_\beta \right)^{-1} I_\beta.$$

I ir $k \times k$ vienības matrica un E_β apzīmē matemātisko cerību, kad X blīvuma funkcija ir $f(x; \beta)$ (t.i., kad izpildās H_0).

Ja β ir zināma, tad hipotēze H_0 saimē (1.2.2) reducējas par hipotēzi $H_0 : \theta = 0$ un statistika ir formā:

$$T_k(\beta) = \sum_{j=1}^{k} \left[n^{-1/2} \sum_{i=2}^{n} \phi_j \left[F(X_i, \beta) \right] \right]^2, \; k = 1, 2, \dots,$$

kur

$$Y_n(\beta) = (\bar{\phi_1}(\beta), \dots \bar{\phi_j}(\beta))' = n^{-1} \sum_{i=1}^{n} \left(\phi_1[F(X_i; \beta)], \dots, \phi_j[F(X_i; \beta)] \right)'.$$

Būtībā, kad β ir zināma, tad Neimaņa tests saliktām hipotēzēm reducējas uz Neimaņa testu vienkāršām hipotēzēm.

Lai izvēlētos atbilstošu dimensiju k, tiek izmantots Švarca selekcijas likums (ieviests 1978. gadā)[5].

$$S(\beta) = \min\{k; 1 \leq k \leq d(n), L_k(\beta) \geq L_j(\beta), j = 1, \dots, d(n)\}, \tag{1.2.4}$$

kur

$$L_k(\beta) = n \sup_{\theta \in R^k} \{\theta \circ Y_n(\beta) - \psi_k(\theta)\} - \frac{1}{2} k \log n$$

Lai gan tas nav pieminēts apzīmējumos, tomēr $S(\beta)$ ir atkarīgs no eksponenciālās saimes augšējās robežas $d(n)$. Ar $\hat{\beta}$ tiek apzīmēts β maksimālās ticamības funkcijas novērtējums pie H_0. Tiek definēts

$$S = S(\hat{\beta})$$

Tādējādi datu nosakošā statistika ir formā [8]

$$W_S = W_{S(\hat{\beta})},$$

kur W_k uzdota vienādojumā (1.2.3), $S(\beta)$ (1.2.4) un $\hat{\beta}$ ir β maksimālās ticamības funkcijas novērtējums.

Švarca likums $S(\beta)$ (1.2.4) salīdzina maksimālās ticamības funkcijas vērtības. Ir pierādīts [1] , ka maksimālās ticamības funkcija ir lokāli ekvivalenta statistikai $\frac{1}{2}n\|Y_n(\beta)\|^2_{(k)}$. Izmantojot šo faktu, ir ieviesta vienkāršāka (vieglāk aprēķināma) Švarca likuma modifikācija

$S2 = S2(\hat{\beta}) =$

$\min\{k : 1 \leq k \leq d(n), n\|Y_n(\hat{\beta})\|^2_k - k\log n \geq n\|Y_n(\hat{\beta})\|^2_j - j\log n,\ j = 1, \dots, d(n)\}.$

Atbilstošā testa statistika ir:

$$W_{S2} = W_{S2(\hat{\beta})}$$

Lai pierādītu Neimaņa testa W_{S2} atbilstību, ir jāzina testa uzvedība pie nulles hipotēzes un pie alternatīvām. Vispirms aplūkosim uzvedību pie nulles hipotēzes. Ar P_β apzīmē, ka X_i blīvuma funkcija ir $f(x;\beta)$, bet ar E_β un var_β apzīmē attiecīgi atbilstošo matemātisko cerību un dispersiju. Funkciju saimei $\{f(x;\beta) : \beta \in B\}$ nepieciešami šādi 4 nosacījumi.

(R1) Katram $t, u = 1, \dots, q$, $(\partial/\partial\beta_t)f(x;\beta)$ un $(\partial^2/\partial\beta_t\partial\beta_u)f(x;\beta)$ eksistē gandrīz visur un ir tādi, ka katrai $\beta_0 \in B_0$ vienmērīgums β_0 apkārtnē

$$|(\partial/\partial\beta_t)f(x;\beta)| \leq H_t(x)$$

un

$$|(\partial^2/\partial\beta_t\partial\beta_u)f(x;\beta)| \leq G_tu(x),$$

kur $\int_{\mathbb{R}} H_t(x)dx < \infty$ un $\int_{\mathbb{R}} G_tu(x)dx < \infty$

(R2) Katram $t, u = 1, \dots, q$, $(\partial/\partial\beta_t)\log f(x;\beta)$ un $(\partial^2/\partial\beta_t\partial\beta_u)\log f(x;\beta)$ eksistē gandrīz visur un ir tādi, ka Fišera informācijas matrica

$$I_{\beta\beta} = E_\beta\left\{\left[\frac{1}{2}\log f(X_1;\beta)\right]\left[\frac{1}{2}\log f(X_1;\beta)\right]'\right\},$$

ir nepārtraukta,ierobežota un pozitīvi definēta, un kad $\delta \to 0$

$$E_\beta = \left\{\sup_{h:\|h\|\leq\delta} \| \frac{\partial^2}{\partial\beta\partial\beta'}\log f(X_1;\beta + h) - \partial\beta\partial\beta'\log f(X_1;\beta) \|\right\} \longrightarrow 0$$

(R3) Katrai $\beta_0 \in B_0$ eksistē $\eta = \eta(\beta_0) > 0$ ar

$$\sup_{\|\beta-\beta_0\|}\sup_{x\in\mathbb{R}}|\frac{\partial^2}{\partial\beta_t\beta_u}F(x;\beta)| < \infty, \quad t, u, = 1, \dots, q$$

un

$$\sup_{x\in\mathbb{R}}|\frac{\partial}{\partial\beta_t}F(x;\beta)|_{\beta=\beta_0} < \infty, \quad t, u, = 1, \dots, q.$$

(R4) Eksistē pozitīvas konstantes c_1, c_2, p_1 un n_1 tādas, ka $\beta \in B_0$ novērtējums $\hat{\beta}$ apmierina

$$P_\beta = (\sqrt{n}) \parallel \hat{\beta} - \beta \parallel \geq n) \leq c_1 \exp(-c_2 r^2)$$

visiem $r = p\sqrt{\log n}$ ar $0 < p \leq p_1$ un $n \leq n_1$

Ortonormālai sistēmai $\{\phi_j\}_{j=0}^{\infty}$ jāizpildās šādiem nosacījumiem.

(S1) Katram $j = 1, 2, \ldots, d(n)$ un $c_3 > 0, m_1 > 0$

$$\sup_{x\in[0,1]} |\phi_j'(x)| \leq c_3 j^{m_1}$$

(S1) Katram $j = 1, 2, \ldots, d(n)$ un $c_4 > 0, m_2 > 0$

$$\sup_{x\in[0,1]} |\phi_j''(x)| \leq c_4 j^{m_2}$$

Eksponenciālās saimes dimensijai $d(n)$ jāizpildās šādiem nosacījumiem.

(D1) $\{d(n)V_{d(n)}\}^2 n^{-1} \log n \longrightarrow 0$, kad $n \longrightarrow \infty$, kur $V_k = \max\limits_{1\leq j\leq k} \sup\limits_{x\in[0,1]} |\phi_j(x)|$.

(D2) $d(n) = o(\{n/\log n\}^{(2m)^{-1}})$, kad $n \longrightarrow \infty$, kur $m = \max(m_1, m_2)$

(D3) $d(n) = o(n^c)$, kad $n \longrightarrow \infty$

$$c < c_2 b^{-2},$$

ja $p_1 b \geq 1$, un ar $c = c_2 p_1^2$, pretējā gadījumā, kur c_2 un p_1 ir uzdotas ar (R4) un

$$b = \left\{ \sum_{i=1}^{n} var_\beta \frac{\partial}{\partial \beta_t} \log f(X;\beta) \right\}^{-1/2}$$

Nākamajās 2 teorēmās tiek parādīta S sākotnējā sadalījuma asimptotika. P_β nozīmē, ka X_i blīvuma funkcija ir $f(x;\beta)$. Pirmā teorēma rāda, cik tuvu Y_n ir $Y_n(\hat{\beta})$.

Teorēma 5. *Pieņem, ka (R1)-(R4), (S1), (S2),(D2), (D3). Tad eksistē tāds $\epsilon < 0$, ka*

$$\lim_{n\to\infty} \sum_{k=2}^{d(n)} P_\beta(\parallel Y_n(\hat{\beta}) - Y_n\beta) \parallel \geq (1-\epsilon)\{(k-1)n^{-1}\log n\}^{1/2}) = 0 \tag{1.2.5}$$

Otrā teorēma rāda,ka pie H_0 Švarca selekcijas kritērijs un tā modifikācija asimptotiski koncentrējas uz dimensiju 1.

Teorēma 6. *Pieņem (R1)-(R4), (S1), (S2), (D1), (D2), (D3), tad*

$$\lim_{n\to\infty} P_\beta(S(\hat{\beta} \geq 2) = 0 \qquad (1.2.6)$$

Teorēma 7. *Pie Teorēmas 6. nosacījumiem*

$$W_{S2(\hat{\beta})} \longrightarrow_d \chi_1^2 \qquad (1.2.7)$$

Teorēmu pierādījumi doti raksts, 1232., 1237. lpp.

Tagad aplūkosim statistikas uzvedību pie alternatīvām. Pieņemsim, ka $X_1, \ldots, X_n$ ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar sadalījumu P. Pie alternatīvām β ir tikai novērtējums $\hat{\beta}$. Pie alternatīvā sadalījuma novērtētais $\hat{\beta}$ parasti konverģēs uz kādu B_0 elementu, šis elements arī būs β.

Apskatīsim P kā saimes $\{f(x;\beta)\}$ alternatīvu, ja eksistē $K(\beta)$ tāds, ka

$$E_P\phi_1(F(X;\beta)) = \ldots = E_P\phi_{K(\beta)-1}(F(X;\beta)) = 0, \quad E_P\phi_{K(\beta)}(F(X;\beta)) \neq 0 \qquad (1.2.8)$$

Pieņemsim, ka

$$\lim_{n\to\infty} \inf d(n) \geq K. \qquad (1.2.9)$$

Ja 1.2.8 neizpildās, tad $E_P\phi_{K(\beta)}(F(X;\beta)) = 0$ visiem j un jebkurai alternatīvai. Pie nulles hipotēzes S koncentrējas uz dimensiju 1, bet pie alternatīvām nozīmīgas kļūst arī augstākas dimensijas.

Teorēma 8. *Pieņem, ka 1.2.8 ir spēkā un β ir saistīta ar $\hat{\beta}$, ka*

$$\hat{\beta} - \beta \longrightarrow_d 0 \qquad (1.2.10)$$

Teorēma 9. *Pie Teorēmas 8. nosacījumiem*

$$W_{S2(\hat{\beta})} \longrightarrow_d 0 \qquad (1.2.11)$$

Teorēmu pierādījumi doti [8] 1234., 1238. lpp.

1.3. Neimaņa statistikas izveduma teorētiskā analīze

Lai izveidotu (pielāgotu) Neimaņa testu atkarīgiem novērojumiem, ir būtiski saprast, kā tieši veidots Neimaņa tests neatkarīgiem novērojumiem. Nodaļā 1.2., izmantojot rakstu [8], dots, ka Neimaņa tests saliktām hipotēzēm neatkarīgiem novērojumiem ir formā:

$$W_k = nY_n(\hat{\beta})\left(I + R(\hat{\beta})\right)Y_n(\hat{\beta}), \tag{1.3.1}$$

kur

$$Y_n(\beta) = (\bar{\phi_1}(\beta), \ldots \bar{\phi_j}(\beta))' = n^{-1}\sum_{i=1}^{n}\left(\phi_1[F(X_i;\beta)], \ldots, \phi_j[F(X_i;\beta)]\right)',$$

$$I_\beta = -E_\beta \frac{\partial}{\partial \beta_t}\phi_j\left[F(X,\beta)\right]_{t=1,\ldots,q \quad j=1,\ldots,k},$$

$$I_{\beta\beta} = -E_\beta \frac{\partial^2}{\partial \beta_t \beta_u}\log f(X,\beta)_{t=1,\ldots,q \quad u=1,\ldots,q},$$

$$R(\beta) = I'_\beta\left(I_{\beta\beta} - I_\beta I'_\beta\right)^{-1} I_\beta.$$

Šī forma ļauj testa statistiku matemātiski aprēķināt, izmantojot atbilstošās datorpaketes. Līdz ar to ir būtiski, modificējot testu atkarīgiem novērojumiem, statistiku iegūt pēc iespējās līdzīgākā formā, tādējādi tās aprēķināšanu padarot vieglāku.

David Roxbee Cox un David Victor Hinkley grāmatā *Theoretical Satistics* [1] parādīts, kā veidojas tests vispārīgā gadījumā. Autores uzdevums ir noskaidrot, kā tieši no grāmatā dotā vispārīgā gadījuma iegūst rakstā *Data Driven Smooth Tests for Composite Hypotheses* [8] doto testa formu W_k 1.3.1.

Ar f_X apzīmēta maksimālās ticamības funkcija un ar log f_X apzīmēta logaritmiskā maksimālās ticamības funkcija vispārīgā gadījumā. Vispirms definēsim vairākus lielumus.

Definīcija 1. Par *Skores* funkciju sauksim

$$U_{\cdot}(\beta) = (U_{\cdot\beta_1}(\beta), U_{\cdot\beta_2}(\beta), \ldots, U_{\cdot\beta_q}(\beta)) = \left(\frac{\partial \log f_X(X;\beta)}{\partial \beta_1}, \frac{\partial \log f_X(X;\beta)}{\partial \beta_2}, \ldots, \frac{\partial \log f_X(X;\beta)}{\partial \beta_q}\right),$$

kur $\beta = (\beta_1, \beta_2, \ldots, \beta_q)$ ir parametru vektors.

Piemēram, normālajā sadalījumā $\beta = (\mu, \sigma)$. Turklāt katra vektora komponente

$$U_{\cdot\beta_k}(\beta) = \sum_{i=1}^{n}(U_{\beta_k})_i(\beta) = \sum_{i=1}^{n}\frac{\partial log f_X(X_i;\beta)}{\partial \beta_k}$$

Definīcija 2. Par Fišera informācijas matricu sauksim

$$i.(\beta) = E(U.^2; \beta) = E_\beta \left\{ -\frac{\partial^2 \log f_X(X;\beta)}{\partial \beta^2} \right\},$$

kur $\beta = (\beta_1, \beta_2, \dots, \beta_q)$ ir parametru vektors.

Ja sadalījumam ir vairāki parametri, tad $i.(\beta)$ kļust par $n \times n$ matricu, kur n ir parametru skaits.

Definīcija 3. Par Fišera informācijas matricas elementu sauksim

$$i._{\beta_l\beta_m}(\beta) = E_\beta \left\{ -\frac{\partial^2 \log f_X(X;\beta)}{\partial \beta_l \partial \beta_m} \right\}.$$

Normālajam sadalījumam $\beta = (\mu, \sigma)$, līdz ar to Fišera informācijas matrica ir formā

$$i.(\beta) = \begin{pmatrix} i._{\mu\mu} & i._{\mu\sigma} \\ i._{\mu\sigma} & i._{\sigma\sigma} \end{pmatrix}$$

Katra matricas komponente

$$i._{\beta_l\beta_m}(\beta) = \sum_{i=1}^{n} (i._{\beta_l\beta_m})_i(\beta) = \sum_{i=1}^{n} E \left\{ -\frac{\partial^2 \log f_X(X_i;\beta)}{\partial \beta_l \partial \beta_m} \right\}$$

Definīcija 4. Par *vidējoto* Fišera informācijas matricu sauksim

$$\mathbf{i.}(\beta) = \frac{i.(\beta)}{n}.$$

Definīcija 5. Par *vidējotās* Fišera informācijas matricas elementu sauksim

$$\mathbf{i.}_{\beta_l\beta_m}(\beta) = \frac{i._{\beta_l\beta_m}(\beta)}{n}.$$

Vispārīgā formā nogludinošais tests saliktām hipotēzēm balstās uz maksimālo ticamības funkciju attiecības.

$$e^{\frac{1}{2}W} = \frac{\sup\limits_{\theta \in \Omega_A} \text{lik}(\theta, X)}{\sup\limits_{\theta \in \Omega_0} \text{lik}(\theta, X)},$$

kur $\text{lik}(\theta, X)$ ir maksimālās ticamības funkcija, pie H_0 pārbaudāmais parametrs $\theta \in \Omega_0$, bet pie H_1 pārbaudāmais parametrs $\theta \in \Omega_A$.

Veicot Teilora izvirzījumu rindā punktā $(\theta = \theta_0, \beta)$, kur θ_0 ir pārbaudāmā parametra vērtība pie H_0, $\hat{\theta}$ ir parametra novērtējums, β ir traucējošā parametra patiesā vērtība, $\hat{\beta}$ ir traucējošā parametra novērtējums, $\hat{\beta}_0$ ir traucējoša parametra novērtētā vērtība pie H_0, iegūst

$$W = 2\{\text{lik}(\hat{\theta}, \hat{\beta}, X) - \text{lik}(\theta_0, \beta, X)\} - 2\{\text{lik}(\theta_0, \hat{\beta}_0, X) - \text{lik}(\theta_0, \beta, X)\} =$$

$$\begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\beta} - \beta \end{bmatrix}' \mathbf{i.}(\theta_0, \beta) \begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\beta} - \beta \end{bmatrix} - \begin{bmatrix} 0 \\ \hat{\beta}_0 - \beta \end{bmatrix}' \mathbf{i.}(\theta_0, \beta) \begin{bmatrix} 0 \\ \hat{\beta}_0 - \beta \end{bmatrix} + o_p(1)$$

Izmantojot [1] pierādītu sakarību

$$\hat{\beta}_0 = \hat{\beta} + \mathbf{i.}_{\beta\beta}^{-1}(\theta_0, \beta)\mathbf{i.}_{\theta\beta}(\theta_0, \beta)(\hat{\theta} - \theta),$$

iegūst, ka statistika ir formā:

$$W = (\hat{\theta} - \theta_0)'(\mathbf{i.}_{\theta\theta}(\theta_0, \beta) - \mathbf{i.}_{\theta\beta}(\theta_0, \beta)\mathbf{i.}_{\beta\beta}{}^{-1}(\theta_0, \beta)\mathbf{i.}_{\theta\beta}{}'(\theta_0, \beta))(\hat{\theta} - \theta_0),$$

Izvirzot $U.(\hat{\theta}, \beta)$ Teilora rindā, iegūst

$$U.(\hat{\theta}, \beta) = U(\theta_0, \beta) + (\hat{\theta} - \theta_0)U.'(\theta_0, \beta) + \frac{1}{2}(\hat{\theta} - \theta_0)^2 U.''(\theta*, \beta), \tag{1.3.2}$$

kur $|\theta * - \theta_0| \leq |\hat{\theta} - \theta_0|$. Pārveidojot

$$(\hat{\theta} - \theta_0)\sqrt{n} = \frac{1}{\sqrt{n}}\mathbf{i.}(\theta_0\beta)^{-1}U.(\theta_0, \beta). \tag{1.3.3}$$

Izmantojot 1.4.4, iegūst testa statistiku formā

$$W_u = U.'_\theta(\theta_0, \hat{\beta}_0)\{\mathbf{i.}_{\theta\theta}(\theta, \beta) - \mathbf{i.}_{\theta\beta}(\theta, \beta)\mathbf{i.}_{\beta\beta}{}^{-1}(\theta, \beta)\mathbf{i.}_{\theta\beta}{}'(\theta, \beta)\}^{-1}U._\theta(\theta_0, \hat{\beta}_0),$$

kur θ_0 ir θ pie H_0 (mūsu gadījumā $\theta_0 = 0$) un $\hat{\beta}_0$ ir novērtētā β pie H_0.

Neimaņa tests pārbauda hipotēzi $H_0 : \theta = \theta_0 \quad$ pret $\quad H_1 : \theta \neq \theta_0$ eksponenciālo blīvuma funkciju saimē, kur blīvuma funkcija dota šādā formā

$$g_k(x; \theta, \beta) = \exp\left(\theta \circ \phi[F(x, \beta)] - \psi_k(\theta)\right) f(x, \beta). \tag{1.3.4}$$

Atbilstošā maksimālās ticamības funkcija ir

$$G = \prod_{i=1}^{n} g_k(x; \theta, \beta) = \exp\left\{\sum_{i=1}^{n} (\theta \circ \phi[F(X_i; \beta)] - \psi_k(\theta))\right\} \prod_{i=1}^{n} f(X_i; \beta).$$

Logaritmiskā maksimālās ticamības funkcija ir

$$\log G = \log \prod_{i=1}^{n} g_k(x; \theta, \beta) = \left\{\sum_{i=1}^{n} (\theta \circ \phi[F(X_i; \beta)] - \psi_k(\theta))\right\} + n \log f(X_i; \beta)$$

Tagad jāparāda, ka [1] uzdotā statistika W_u ir tā pati, kas [8] uzdotā statistika W_k. Vektors

$$U_{\cdot\theta}(\theta_0; \hat{\beta}_0) = \frac{\partial \log G}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} (\theta \circ \phi[F(X_i; \beta)] - \psi_k(\theta)) + 0$$
$$= \sum_{i=1}^{n} (\phi_1[F(X_i; \beta)], \ldots, \phi_j[F(X_i; \beta)])' = Y_n(\beta) \cdot n.$$

Matricā

$$\{\mathbf{i}_{\cdot\theta\theta}(\theta, \beta) - \mathbf{i}_{\cdot\theta\beta}(\theta, \beta)\mathbf{i}_{\cdot\beta\beta}{}^{-1}(\theta, \beta)\mathbf{i}_{\cdot\theta\beta}{}'(\theta, \beta)\}^{-1}$$

vispirms jānoskaidro atsevišķi matricas $\mathbf{i}_{\cdot\theta\theta}(\theta, \beta)$, $\mathbf{i}_{\cdot\theta\beta}(\theta, \beta)$ un $\mathbf{i}_{\cdot\beta\beta}$.

Matrica $\mathbf{i}_{\cdot\beta\beta}(\theta; \beta)$ atšifrējama kā

$$\mathbf{i}_{\cdot\beta\beta}(\theta; \beta) = E\left\{-\frac{\partial^2 \log G}{\partial \beta^2}\right\} = E\left\{-\frac{\partial^2}{\partial \beta^2} n \log f(X_i; \beta)\right\} = n \cdot I_{\beta\beta}$$

Šādi izsakāma matrica

$$\mathbf{i}_{\cdot\theta\beta}(\theta; \beta) = E\left\{-\frac{\partial^2 \log G}{\partial \theta \partial \beta}\right\} = \begin{pmatrix} \frac{\partial^2}{\partial \theta_1 \beta_1} & \cdots & \frac{\partial^2}{\partial \theta_1 \beta_q} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial \theta_k \partial \beta_1} & \cdots & \frac{\partial^2}{\partial \theta_k \partial \beta_q} \end{pmatrix} \log G$$
$$= E\left\{-\sum_{i=1}^{n} \frac{\partial}{\partial \beta_t} \psi_j[F(X_i; \beta)]\right\}_{t=1,\ldots,q \quad j=1,\ldots,k} = n \cdot I_{\beta}$$

Savukārt matrica

$$\begin{aligned}
\mathbf{i}_{\cdot\theta\theta}(\theta;\beta) &= E\left\{-\frac{\partial^2 \log G}{\partial\theta^2}\right\} = E\left\{-\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\psi_k(\theta)\right\} \\
&= E\left\{-n\frac{\partial^2}{\partial\theta^2}\left(\log\int_0^1 \exp\theta\circ\phi(y)dy\right)\right\} \\
&= E\left\{-n\begin{pmatrix} \frac{\partial^2}{\partial{\theta_1}^2} & \cdots & \frac{\partial^2}{\partial\theta_1\theta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial\theta_1\partial\theta_k} & \cdots & \frac{\partial^2}{\partial\theta_k\partial\theta_k}\end{pmatrix}\left(\log\int\limits_0^1 \exp\theta\circ\phi(y)\,\mathrm{d}y\right)\right\} \\
&= E\left\{-n\begin{pmatrix} \frac{\partial^2}{\partial{\theta_1}^2}\left(\log\int\limits_0^1 \exp\theta\circ\phi(y)\,\mathrm{d}y\right) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial^2}{\partial{\theta_k}^2}\left(\log\int\limits_0^1 \exp\theta\circ\phi(y)\,\mathrm{d}y\right)\end{pmatrix}\right\} \\
&= -n\cdot I
\end{aligned}$$

Tātad

$$\mathbf{i}_{\cdot\theta\theta}(\theta,\beta) - \mathbf{i}_{\cdot\theta\beta}(\theta,\beta)\mathbf{i}_{\cdot\beta\beta}{}^{-1}(\theta,\beta)\mathbf{i}_{\cdot\theta\beta}{}'(\theta,\beta) = -n\cdot I + n\cdot I'_\beta I^{-1}_{\beta\beta}I_\beta$$

Izmantosim inversās matricas īpašības [9]. Ja dotas matricas A, B, C un D, tad

$$(A+BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1}+DA^{-1}B)^{-1}DA^{-1}.$$

Mūsu gadījumā matrica $A = I$, $B = I'_\beta$, $C = I^{-1}_{\beta\beta}$, $D = I_\beta$. Tādējādi

$$\begin{aligned}
&(\mathbf{i}_{\cdot\theta\theta}(\theta,\beta) - \mathbf{i}_{\cdot\theta\beta}(\theta,\beta)\mathbf{i}_{\cdot\beta\beta}{}^{-1}(\theta,\beta)\mathbf{i}_{\cdot\theta\beta}{}'(\theta,\beta))^{-1} \\
&= {-n\cdot I + n\cdot I'_\beta I^{-1}_{\beta\beta}I_\beta}^{-1} \\
&= \frac{1}{n}\left(I + II'_\beta(I_{\beta\beta} - I_\beta I'_\beta)^{-1}I_\beta I\right) = \frac{1}{n}R(\beta).
\end{aligned}$$

Esam ieguvuši

$$\begin{aligned}
W_u &= U'_{\cdot\theta}(\theta_0,\hat{\beta}_0)(\mathbf{i}_{\cdot\theta\theta}(\theta,\beta) - \mathbf{i}_{\cdot\theta\beta}(\theta,\beta)\mathbf{i}_{\cdot\beta\beta}{}^{-1}(\theta,\beta)\mathbf{i}_{\cdot\theta\beta}{}'(\theta,\beta))^{-1}U_{\cdot\theta}(\theta_0,\hat{\beta}_0) \\
&= nY_n(\hat{\beta})\left(I + R(\hat{\beta})\right)Y_n(\hat{\beta}) = W_k.
\end{aligned}$$

Esam noskaidrojuši, kā no vispārīgā gadījuma iegūst Neimaņa nogludinošo testu neatkarīgiem un vienādi sadalītiem datiem.

1.4. Neimaņa tests atkarīgiem novērojumiem

Noteikt, vai gadījuma lielumi atbilst kādai blīvuma funkciju kopai, ja gadījuma lielumi ir savstarpēji atkarīgi, ir daudz grūtāk nekā gadījumos, kad gadījuma lielumi ir savstarpēji neatkarīgi. Ir dažādi veidi, kā veidot atkarību vispārējā formā, piemēram, izmantojot jauktos procesus [10]. Šajā darbā analizēm populāros ARMA procesus, kuri tiek izmantoti ekonometrikā [11], inženierzinātnēs [12] u.c. Vispirms definēsim ARMA procesu.

Definīcija 6. Par par autoregresīvo modeli ar kārtu p un apzīmējumu $AR(p)$ sauc procesu [13]

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t,$$

kur $\phi_1, \phi_2, \ldots, \phi_p$ ir koeficienti, $|\phi_i| < 1$, c ir konstante (datu vidējā vērtība) un $\epsilon_t \sim^{iid} N(0, \sigma^2)$ (tā sauktais "baltais troksnis").

Definīcija 7. Par slīdošā vidējā procesu ar kārtu q un apzīmējumu $MA(q)$ sauc procesu [13]

$$X_t = \mu + \sum_{i=1}^{n} \theta_i \epsilon_{t-1} + \epsilon_t,$$

kur $\theta_1, \ldots, \theta_q$ ir konstantes, μ ir X_t sagaidāmā vērtība un $\epsilon_t \sim^{iid} N(0, \sigma^2)$.

Definīcija 8. Par ARMA(p,q) procesu sauc

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t + \mu + \sum_{i=1}^{n} \theta_i \epsilon_{t-1}.$$

Kā jau 1.3. nodaļā minēts, tad Neimaņa tests balstās uz maksimālo ticamības funkciju attiecības. Maksimālās ticamības funkcija ir atkarīga no gadījuma lielumu blīvuma funkcijām. ARMA kopējā procesiem blīvuma funkcija ir

$$f(X_1, X_2, ..., X_n) = f(X_1) \cdot f(X_2|X_1) \cdot f(X_3|X_1, X_2) \cdot \ldots \cdot f(X_n|X_1, X_2, \ldots, X_{n-1}).$$

ARMA procesu var reducēt par AR(1)procesu, tādēļ nakamā nodaļa tiks veltīta AR(1) procesam.

1.4.1. AR(1) process

Definīcija 9. Par stacionāru AR(1)procesu jeb *pirmās kārtas autoregresīvu procesu* [14] sauc procesu

$$X_t = c + \phi X_{t-1} + \epsilon_t,$$

kur $\epsilon_t \sim iid\ N(0, \sigma^2), (t = 1, \ldots, T)$ un $\mid \phi \mid < 1$.

Turpmāk, īsāka pieraksta labad, ar θ apzīmēsim visus šī procesa parametrus $\theta = (c, \phi, \sigma^2)$. AR(1) jeb *pirmās kārtas autoregresīvajā procesā* katrs nākamais gadījuma lielums ir atkarīgs tikai no iepriekšējā gadījuma lieluma, nevis no vairākiem iepriekšējajiem gadījuma lielumiem.

Kā jau 1.3. nodaļā minēts, tad Neimaņa tests balstās uz maksimālo ticamības funkciju attiecības. Maksimālās ticamības funkcija ir atkarīga no gadījuma lielumu blīvuma funkcijām, līdz ar to ir svarīgi uzzināt, kādā formā blīvuma funkcija ir AR(1) procesam.

Doti $X_1, X_2, ..., X_n$ gadījuma lielumi, $\epsilon_t \sim N(0, 1)$, tad AR(1) procesam ir spēkā:

$$X_1 \sim N\left(\frac{c}{1-\phi}, \frac{\sigma^2}{1-\phi^2}\right)$$

$$X_i \mid X_{i-1} \sim N(c + \phi X_{i-1}, \sigma^2) \quad i = 2, \ldots, n$$

Pirmā gadījuma lieluma blīvuma funkcija ir [15]

$$f_{X_1}(X_1; \theta) = \sqrt{\frac{1-\phi^2}{2\pi\sigma^2}} \exp\left(-\frac{1-\phi^2}{2\sigma^2}\left(X_1 - \frac{c}{1-\phi}\right)^2\right) \tag{1.4.1}$$

Pārējās blīvuma funkcijas ir nosacītās blīvuma funkcijas.

$$f_{X_i|X_{i-1}}(X_i; \theta) = \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(X_i - c - \phi X_{i-1})^2\right) \quad i = 2, \ldots, n \tag{1.4.2}$$

Plašāk par AR(1) procesu un tā pielietojumu lasīt [16] [17].

1.4.2. Neimaņa testa modifikācija AR(1) procesam

1.3. nodaļā tika parādīts, kā no vispārīgā gadījuma nogludinošā testa statistikas tiek iegūts Neimaņa tests neatkarīgiem novērojumiem. Šajā nodaļā, izmantojot iepriekšējā nodaļā doto izvedumu, tiks parādīts, kādā formā ir Neimaņa tests AR(1) procesam.

Idejiski (no Neimaņa testa uzbūves viedokļa) galvenā atšķirība starp neatkarīgiem gadījuma lielumiem un AR(1) ir tā, ka neatkarīgiem gadījuma lielumiem maksimālās ticamībās funkcijas iegūšanai tiek sareizinātas vienādas blīvuma funkcijas, taču AR(1) maksimālās ticamības funkcijas iegūšanai pirmā blīvuma funkcija ir atšķirīga no pārējām nosacītajām blīvuma funkcijām, ar ko tiek reizināta.

Neimaņa testā blīvuma funkcija ir eksponenciālo blīvuma funkciju saimē

$$g_k(x;\theta,\beta) = \exp\left(\theta \circ \phi[F(x,\beta)] - \psi_k(\theta)\right) f(x,\beta), \tag{1.4.3}$$

kur pirmajam gadījuma lielumam šī blīvuma funkcija ir $F_{X_1}(x_1,\beta) = \int\limits_{-\infty}^{x_1} f_{X_1}(x_1,\beta)$ un $f_{X_1}(x_1,\beta)$ ir uzdota (1.4.1), savukārt pārējo gadījuma lielumu blīvuma funkcijas ir $F_{X_i|X_{i-1}}(X_i,\beta) = \int\limits_{-\infty}^{X_i} f_{X_i|X_{i-1}}(X_i,\beta)$ un $f_{X_i|X_{i-1}}(X_i;\theta)$ ir uzdota ar (1.4.2).

Maksimālās ticamības funkcija AR(1) procesam:

$$G = \prod_{i=1}^{n} g_k(X_i;\theta,\beta) = g_k(X_1;\theta,\beta)\prod_{i=2}^{n} g_k(X_i;\theta,\beta)$$
$$= \exp\left(\theta \circ \phi[F_{X_1}(x_1,\beta)] - \psi_k(\theta)\right) f_{X_1}(x_1,\beta) \prod_{i=2}^{n} \exp\left(\theta \circ \phi[F_{X_i|X_{i-1}}(X_i,\beta)] - \psi_k(\theta)\right) f_{X_i|X_{i-1}}(X_i,\beta)$$

Logaritmiskā maksimālās ticamības funkcija:

$$\log G = \left\{\theta_1 \circ \phi[F_{X_1}(X_1,\beta)] + \sum_{i=2}^{n} \theta \circ \phi[F_{X_i|X_{i-1}}(X_i,\beta)]\right\}$$
$$+ \log f_{X_1}(X_1,\beta) + \sum_{i=2}^{n} \log f_{X_i|X_{i-1}}(X_i,\beta)$$

No nodaļas 1.3. vispārīgā gadījumā Neimaņa testa statistika ir formā

$$W = (\hat{\theta} - \theta_0)'(\mathbf{i}_{\cdot\theta\theta}(\theta,\beta) - \mathbf{i}_{\cdot\theta\beta}(\theta,\beta)\mathbf{i}_{\cdot\beta\beta}{}^{-1}(\theta,\beta)\mathbf{i}_{\cdot\theta\beta}{}'(\theta,\beta))(\hat{\theta} - \theta_0),$$

Nodaļa 1.3. dots, ka neatkarīgiem gadījuma lielumiem

$$(\hat{\theta} - \theta_0)\sqrt{n} = \frac{1}{\sqrt{n}}\mathbf{i}.(\theta_0\beta)^{-1}U.(\theta_0,\beta). \tag{1.4.4}$$

Veicot izvedumu Teilora rindā AR(1) procesa gadījumā:

$$U.(\hat{\theta},\beta) = U(\theta_0,\beta) + (\hat{\theta} - \theta_0)U.'(\theta_0,\beta) + \frac{1}{2}(\hat{\theta} - \theta_0)^2 U.''(\theta*,\beta), \tag{1.4.5}$$

kur $|\theta * - \theta_0| \leq |\hat{\theta} - \theta_0|$. Pārveidojot:

$$(\hat{\theta} - \theta_0)\sqrt{n} = \frac{1}{\sqrt{n}} n(\mathbf{i}_1(\theta_0\beta) + (n-1)\mathbf{i}.(\theta_0\beta))^{-1}(U_1(X_1,\theta_0\beta) + \sum_{i=2}^{n} U_i(X_i \mid X_{i-1},\theta_0,\beta)), \tag{1.4.6}$$

kur var redzēt, ka informācijas matrica vienam gadījuma lielumam neatkarīgu lielumu gadījumā

$$\mathbf{i.}(\theta_0, \beta) \tag{1.4.7}$$

1.4.4 AR(1) tiek aizstāta ar visu gadījuma lielumu informācijas matricu vidējo

$$\frac{\mathbf{i}_1(\theta_0\beta) + (n-1)\mathbf{i}.(\theta_0\beta)}{n}.$$

Analoģiski arī

$$U.(\theta_0, \beta)$$

vietā iegūst

$$U_1(X_1, \beta) + \sum_{j=2}^{n} U_j(X_j \mid X_{j-1}, \beta).$$

Tagad, izmantojot (1.4.4) un (1.4.6), jāparāda, kāda izskatās Neimaņa testa statistika AR(1) procesam, t.i., jānosaka atbilstošās matricas un vektori.

Vektors

$$\begin{aligned} U_{\cdot\theta}(\theta_0; \hat{\beta}_0) &= \frac{\partial \log G}{\partial \theta} = \frac{\partial}{\partial \theta}\left\{\theta_1 \circ \phi[F_{X_1}(X_1, \beta)] + \sum_{i=2}^{n} \theta \circ \phi[F_{X_i|X_{i-1}}(X_i, \beta)]\right\} + 0 \\ &= \sum_{i=1}^{n}(\phi_1[F_{X_1}(X_i; \beta)], \phi_2[F_{X_i|X_{i-1}}(X_i; \beta)], \ldots, \phi_j[F_{X_i|X_{i-1}}(X_i; \beta)])' = Y_n(\beta) \cdot n \end{aligned}$$

saglabājas tāds pats, tikai sadalījuma funkcija X_i ir atšķirīga no pārējo gadījuma lielumu nosacītajām sadalījuma funkcijām.

Matricā

$$(\mathbf{i}_{\cdot\theta\theta}(\theta, \beta) - \mathbf{i}_{\cdot\theta\beta}(\theta, \beta)\mathbf{i}_{\cdot\beta\beta}{}^{-1}(\theta, \beta)\mathbf{i}_{\cdot\theta\beta}{}'(\theta, \beta))^{-1}$$

vispirms jānoskaidro atsevišķi matricas $\mathbf{i}_{\cdot\theta\theta}(\theta, \beta)$, $\mathbf{i}_{\cdot\theta\beta}(\theta, \beta)$ un $\mathbf{i}_{\cdot\beta\beta}$.

Matrica

$$\begin{aligned} \mathbf{i}_{\cdot\theta\theta}(\theta; \beta) &= E\left\{-\frac{\partial^2 \log G}{\partial \theta^2}\right\} = E\left\{-\sum_{i=1}^{n}\frac{\partial^2}{\partial \theta^2}\psi_k(\theta)\right\} \\ &= E\left\{-n\frac{\partial^2}{\partial \theta^2}\left(\log \int_0^1 \exp \theta \circ \phi(y)\, \mathrm{d}y\right)\right\} = -n \cdot I \end{aligned}$$

ir tāda pati kā neatkarīgiem gadījuma lielumiem. Matrica

$$\begin{aligned}
\mathbf{i}_{\cdot\theta\beta}(\theta;\beta) &= E\left\{-\frac{\partial^2 \log G}{\partial\theta\partial\beta}\right\} = \begin{pmatrix} \frac{\partial^2}{\partial\theta_1\beta_1} & \cdots & \frac{\partial^2}{\partial\theta_1\beta_q} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial\theta_k\partial\beta_1} & \cdots & \frac{\partial^2}{\partial\theta_k\partial\beta_q} \end{pmatrix} \log G = \\
&= E\left\{-\left(\frac{\partial}{\partial\beta_t}\phi_j[F_{X_1}(X_1;\beta)] + \sum_{i=2}^{n}\frac{\partial}{\partial\beta_t}\phi_j[F_{X_i|X_{i-1}}(X_i;\beta)]\right) + 0\right\} \\
&= E\left\{-\left(\frac{\partial}{\partial\beta_t}\phi_j[F_{X_1}(X_1;\beta)] + \sum_{i=2}^{n}\frac{\partial}{\partial\beta_t}\phi_j[F_{X_i|X_{i-1}}(X_i;\beta)]\right)\right\}_{t=1,\ldots,q \quad j=1,\ldots,k} \\
&= n\cdot I_\beta,
\end{aligned}$$

kur tagad

$$I_\beta = E\left\{-n^{-1}\left(\frac{\partial}{\partial\beta_t}\phi_j[F_{X_1}(X_1;\beta)] + \sum_{i=2}^{n}\frac{\partial}{\partial\beta_t}\phi_j[F_{X_i}\mid X_{i-1}(X_i;\beta)]\right)\right\}_{t=1,\ldots,q \quad j=1,\ldots,k}$$

Matrica

$$\begin{aligned}
\mathbf{i}_{\cdot\beta\beta}(\theta;\beta) = E\left\{-\frac{\partial^2 \log G}{\partial\beta_u\beta_t}\right\} &= E\left\{0 - \frac{\partial^2}{\partial\beta_u\beta_t}\left(\log f_{X_1}(X_1,\beta) + \sum_{i=2}^{n}\log f_{X_i|X_{i-1}}(X_i,\beta)\right)\right\} \\
&= E\left\{-\frac{\partial^2}{\partial\beta_u\beta_t}\log f_{X_1}(X_1,\beta) + \frac{\partial^2}{\partial\beta_u\beta_t}\sum_{i=2}^{n}\log f_{X_i|X_{i-1}}(X_i,\beta)\right\} = n\cdot I_{\beta\beta},
\end{aligned}$$

kur tagad

$$I_{\beta\beta} = E\left\{-n^{-1}\left(\frac{\partial^2}{\partial\beta_u\beta_t}(log f_{X_1}(X_1,\beta) + \frac{\partial^2}{\partial\beta_u\beta_t}\sum_{i=2}^{n} log f_{X_i|X_{i-1}}(X_i,\beta)\right)\right\}_{t=1,\ldots,q \quad u=1,\ldots,q}$$

Teorēma 10. *Neimaņa testa statistika AR(1) procesam ir formā*

$$W_k = nY_n(\hat{\beta})\left(I + R(\hat{\beta})\right)Y_n(\hat{\beta}),$$

kur

$$Y_n(\hat{\beta}) = n^{-1}\sum_{i=1}^{n}(\phi_1[F_{X_1}(X_i;\beta)], \phi_2[F_{X_i|X_{i-1}}(X_i;\beta)] \ldots, \phi_j[F_{X_i|X_{i-1}}(X_i;\beta)])',$$

$$R(\beta) = I_\beta'\left(I_{\beta\beta} - I_\beta I_\beta'\right)^{-1} I_\beta,$$

kur

$$I_\beta = E\left\{-n^{-1}\left(\frac{\partial}{\partial\beta_t}\phi_j[F_{X_1}(X_1;\beta)] + \sum_{i=2}^{n}\frac{\partial}{\partial\beta_t}\phi_j[F_{X_i|X_{i-1}}(X_i;\beta)]\right)\right\}_{t=1,\ldots,q \quad j=1,\ldots,k}$$

$$I_{\beta\beta} = E\left\{-n^{-1}\left(\frac{\partial^2}{\partial\beta_u\beta_t}(log f_{X_1}(X_1,\beta) + \frac{\partial^2}{\partial\beta_u\beta_t}\sum_{i=2}^{n} log f_{X_i|X_{i-1}}(X_i,\beta)\right)\right\}_{t=1,\ldots,q \quad u=1,\ldots,q}$$

Pie teorēmas 6 nosacījumiem

$$W_k \longrightarrow_d \chi_1^2.$$

Pie teorēmas 8 nosacījumiem

$$W_k \longrightarrow_d 0.$$

2. Praktiskā daļa

2.1. Neimaņa testa implementācija programmēšanas valodā R

Salīdzinoši ilgu laiku prasīja Neimaņa testa saliktām hipotēzēm neatkarīgiem datiem implementācija programmēšanas valodā **R**. Pie tās izstrādēs tika pavadīti vairāki mēneši. Nodaļā 1.2. dots, ka Neimaņa tests saliktām hipotēzēm ir formā:

$$W_k = nY_n(\hat{\beta})\left(I + R(\hat{\beta})\right)Y_n(\hat{\beta}),$$

kur

$$Y_n(\beta) = (\bar{\phi_1}(\beta), \ldots \bar{\phi_j}(\beta))' = n^{-1}\sum_{i=1}^{n}\left(\phi_1[F(X_i;\beta)], \ldots, \phi_j[F(X_i;\beta)]\right)',$$

$$I_\beta = -E_\beta \frac{\partial}{\partial \beta_t}\phi_j\left[F(X,\beta)\right]_{t=1,\ldots,q \quad j=1,\ldots,k'},$$

$$I_{\beta\beta} = -E_\beta \frac{\partial^2}{\partial \beta_t \beta_u}\log f(X,\beta)_{t=1,\ldots,q \quad u=1,\ldots,q},$$

$$R(\beta) = I_\beta'\left(I_{\beta\beta} - I_\beta I_\beta'\right)^{-1} I_\beta.$$

Ja vektora Y_n realizācija programmēšanas valodā **R** ir salīdzinoši vienkārša, tad lielākās grūtības sagādāja matricu I_β un $I_{\beta\beta}$ realizācija. Matricā I_β vispirms ir jāiegūst sadalījuma funkcija, kas integrē blīvuma funkciju līdz X, tad šī sadalījuma funkcija jāpadod kā arguments Ležandra polinomam, un pēc tam šis polinoms ir jāatvasina pēc atbilstošās β (vienas vai vairākām atkarībā no pārbaudāmā sadalījuma (eksponenciālajam sadalījumam tikai β, bet normālajam sadalījumam $\beta = (\mu, \sigma)$).

Tā kā programmēšanas valodā **R** nav iespējas funkciju analītiski atvasināt un integrēt, tad papildus tika izmantota programma *Maple*. Tiek apskatīta *regulāra* problēma, kas nozīmē, ka atvasināšanu un integrēšanu var mainīt vietām. Tādējādi katrs polinoms vispirms tika atvasināts pēc konkrētā parametra β, tad, apmainot atvasināšanu un integrēšanu vietām, blīvuma funkcija tika atvasināta pēc β un pēc tam nointegrēta līdz X. Un beigās tika paņemta matemātiskā cerība no šī lieluma. Ieskatam dots piemērs, kā tas izskatījās pirmā Ležandra polinoma $\phi_1 = \sqrt{12}(x - 0.5)$ gadījumā eksponeciālajam sadalījumam, kurā ir tikai viens β. Tātad

$$I_\beta = -E_\beta \frac{\partial}{\partial \beta}\phi_1[F(X,\beta)] = -E_\beta \left(\sqrt{12}\frac{\partial}{\partial \beta}[F(X,\beta)]\right) = -E_\beta \left(\sqrt{12}\int_0^X \frac{\partial}{\partial \beta} f(X,\beta)\right)$$

Eksponenciālā sadalījuma blīvuma funkcijas atvasinājums pēc β ir

$$\frac{\partial}{\partial \beta} f(X,\beta) = e^{-\beta x} - \beta x e^{-\beta x}.$$

Līdz ar to

$$I_\beta = \{-E_\beta \left(\sqrt{12}\int_0^X \frac{\partial}{\partial \beta} f(X,\beta)\right)\} = \{-E_\beta \left(\sqrt{12}\int_0^X e^{-\beta x} - \beta x e^{-\beta x} dx\right)\}.$$

Kad blīvuma funkcija ir atvasināta pēc atbilstošā parametra β, tad, visas pārējās matemātiskās darbības var un tiek veiktas programmēšanas valodā **R**.

Matricā $I_{\beta\beta}$ ir jāņem naturāllogaritms no blīvuma funkcijas un pēc tam iegūtais rezultāts jāatvasina pēc atbilstošās β. Arī šīs darbībās ir jāveic analītiski un tas tika darīts programmā *Maple*. Eksponenciālajam sadalījumam ir tikai viens parametrs β, līdz ar to matricai $I_{\beta\beta}$ ir tikai viens elements, savukārt normālajam sadalījumam ir 2 parametri, jo $\beta = (\mu, \sigma)$, līdz ar to matrica $I_{\beta\beta}$ ir 2×2 matrica. Ieskatam tiks dots piemērs, kāds izskatās matricas $I_{\beta\beta}$ pirmais elements, t.i.,

$$-E_\beta \frac{\partial^2}{\partial\mu\partial\mu} \log f(X,\beta) = -E_\beta \frac{\partial^2}{\partial\mu\partial\mu} \ln \left((\sqrt{2\pi}\sigma)^{-1} \exp - \frac{(x-\mu)^2}{2\sigma^2} \right) = -E_\beta \left(-\frac{1}{\sigma^2}\right).$$

Matemātiskā cerībā šim lielumam tiek aprēķināta programmēšanas valodā **R**. Arī visas nepieciešamās darbības ar matricām tiek veiktas **R**. Visi nepieciešamie atvasinājumi, kas veikti ar programmu *Maple*, ir doti pielikumā. Arī visi programmas **R** kodi, lai realizētu Neimaņa testu saliktām hipotēzēm, ir doti pielikumā.

Tā kā Neimaņa testa saliktām hipotēzēm uzbūve ir gana sarežģīta un tās realizācijā jāveic daudzi soļi, tad autore Neimaņa testu saliktām hipotēzēm eksponenciālājam sadalījumam ir veikusi ar 6 Ležandra polinomiem, savukārt normālajam sadalījumam ar 3 Ležandra polinomiem. Jāpiebilst, ka katra nākamā polinoma pievienošana, palielina programmas darbības ilgumu par vismaz 30%.

Neimaņa testa saliktām hipotēzēm AR(1) procesam implementācija programmēšanas valodā **R** tika veikta līdzīgā veidā kā Neimaņa testa saliktām hipotēzēm neatkarīgiem datiem implementācija. Galvenā atšķirība, ka pirmajam datam ir citāda blīvuma funkcija nekā pārējiem. Kad tika veikti atbilstošie atvasinājumi nosacītajām blīvuma funkcijām, tad, bez parametriem c (mūsu gadījumā ir 0) σ (un ϕ), arī iepriekšējais dats, kas ieiet nosacītajā blīvuma funkcijā, tika uzskatīts par parametru, t.i., atvasinot "pret to attiecās" kā pret parametru. Ieskatam dots nosacītās blīvuma funkcijas atvasinājums pēc σ (tas jau ir maksimāli vienkāršots programmā *Maple*).

$$\frac{\partial}{\partial \sigma}\left(f_{X_i|X_{i-1}}(x_i;\theta)\right) = -\frac{\sqrt{2}e^{-\frac{1}{2}\frac{(-x_i+\phi x_{i-1})^2}{\sigma^2}}(\sigma^2 - x_i^2 + 2x\phi x_{i-1} - \phi^2 x_{i-1}^2)}{\sigma^3\sqrt{\pi\sigma^2}}$$

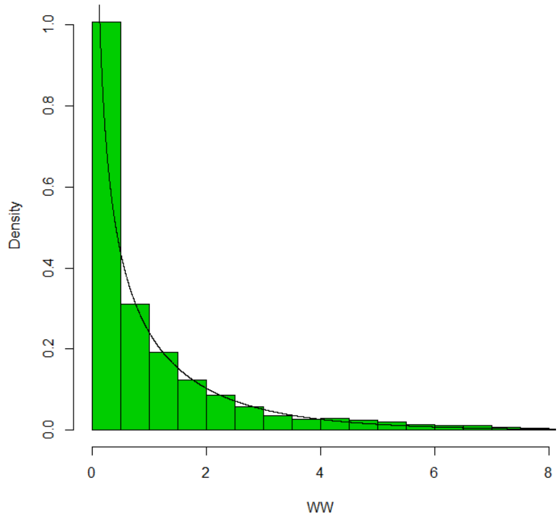
Visi nepieciešami atvasinājumi, kas veikti ar programmu *Maple*, doti pielikumā. Tā kā AR(1) procesa gadījumā nav vairs viena matemātiskā cerība, bet gan katram datam atšķirīga, tad salīdzinot ar neatkarīgiem datiem realizēto Neimaņa testu, darbību skaits pieaug tik reizes, cik datu izlase tiek apskatīta. Piemēram, ar vienu polinomu realizēts Neimaņa tests AR(1) procesam 1000 atkārtojumus 50 datu izlasei uz salīdzinoši ātra datora veic apmēram diennakti. Tas arī ir viens no iemesliem, kādēļ šobrīd vēl nav realizēts Neimaņa tests ar vairākiem polinomiem AR(1) procesam.

2.2. Neimaņa tests saliktām hipotēzēm neatkarīgiem datiem

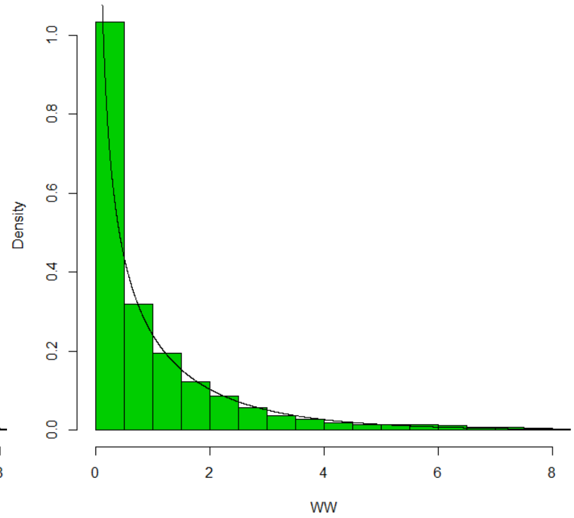
2.2.1. Robežsadalījums

Vispirms tika izveidota programma, kas programmēšanas valodā **R** realizē Neimaņa testu saliktām hipotēzēm eksponenciālajam sadalījumam. Šajā programmā ar Švarca selekcijas kritēriju tiek izvēlēts atbilstošais polinomu skaits no maksimums sešiem polinomiem. Lai pārbaudītu programmas darbību, tika veiktas simulācijas pie H_0. Veicot atkārtojumus 10 000 reižu, tika iegūti robežsadalījumi pie dažādām β vērtībām un dažādiem izlases apjomiem n.

Tika izveidota programma, kas programmēšanas valodā **R** realizē Neimaņa testu saliktām hipotēzēm normālajam sadalījumam. Šajā programmā ar Švarca selekcijas kritēriju tiek izvēlēts



(a) N=10000, n=20, $\beta = 1$



(b) N=10000, n=50, $\beta = 1$

2.1. Att.: Robežsadalījuma pārbaude, nosakot ekponencialitāti.

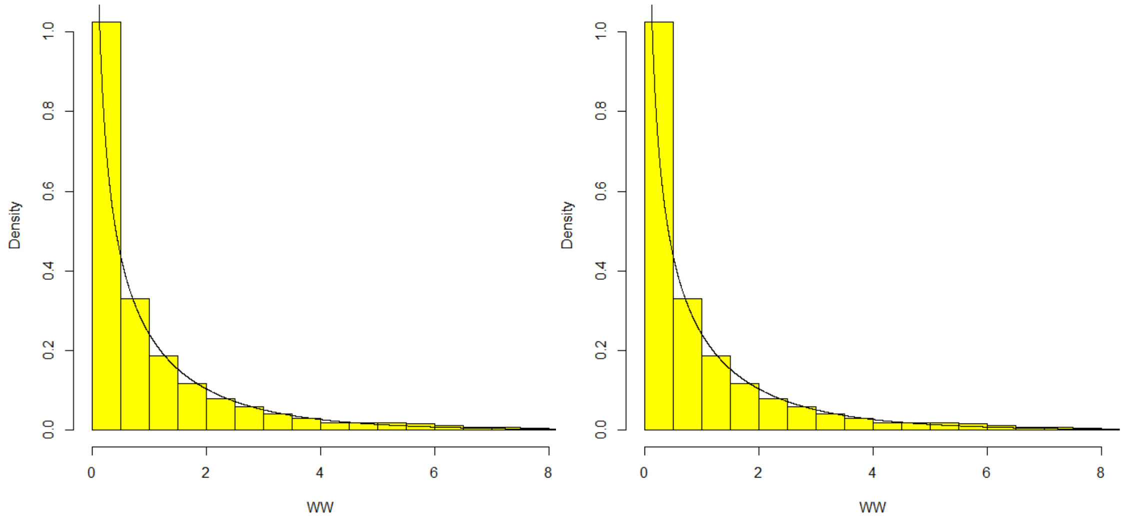
atbilstošais polinomu skaits no maksimums trim Ležandra polinomiem. Lai pārbaudītu programmas darbību, tika veiktas simulācijas pie H_0. Veicot atkārtojumus 10 000 reižu, tika iegūti robežsadalījumi pie dažādām $\beta = (\mu, \sigma)$ vērtībām un dažādiem izlases apjomiem n. 2.1. attēlā redzams, ka, pārbaudot ekponencialitāti, tiek aproksimēts robežsadalījums gan pie $n = 20$, gan $n = 50$.

Arī, pārbaudot normalitāti, tiek aproksimēts robežsadalījums pie $n = 20$, gan $n = 50$, ko var redzēt 2.2. attēlā.

2.2.2. Kritiskās vērtības

2.1. tabulā dotas kritiskās vērtības, pārbaudot datu atbilstību ekponenciālajam sadalījumam izlasēm pie apjoma $n = 20$ un $n = 50$ un pie nozīmīguma līmeņa $\alpha = 0.05$ un $\alpha = 0.10$. Var redzēt, ka kritiskās vērtības ir salīdzinoši tuvas, ja $d(n)$ ir no 2 līdz 6, savukārt, starp 1 un 2 ir liela atstarpe (lēciens). Līdz ar to, simulējot empīriskās jaudas, ir būtiski, lai tiktu izmantoti vismaz 2 polinomi un pielietots Švarca selekcijas kritērijs.

2.2. tabulā dotas kritiskās vērtības, pārbaudot datu atbilstību normālajam sadalījumam izlasēm pie apjoma $n = 20$ un $n = 50$ un pie nozīmīguma līmeņa $\alpha = 0.05$ un $\alpha = 0.10$. Var redzēt, ka kritiskās vērtības ir salīdzinoši tuvas pie visām ģenerētajām $d(n)$ vērtībām.



(a) N=10000, n=20, $\beta = (\mu = 0, \sigma = 1)$

(b) N=10000, n=50, $\beta = (\mu = 0, \sigma = 1)$

2.2. Att.: Robežsadalījuma pārbaude, nosakot normalitāti.

2.1. Tabula: Testa $W_{S(\hat{\beta})}$ 5% un 10% kritiskās vērtības, pārbaudot eksponenciālo sadalījumu

		d(n)					
n	α	1	2	3	4	5	6
20	0.05	3.521	4.894	5.873	6.214	6.361	6.662
	0.10	2.589	3.341	3.651	3.933	4.001	4.056
50	0.05	3.828	4.718	4.819	4.891	5.161	5.422
	0.10	2.597	2.780	2.768	2.890	2.964	3.056

2.2.3. Alternatīvie sadalījumi un simulētās jaudas eksponenciālajam sadalījumam

Testam efektivitāti nosaka tas, cik labi tas spēj noraidīti alternatīvas. Tādēļ tiek veikta empīriskā jaudas analīze. Šajā darbā, lai noteiktu eksponencialitāti, tika veikta jaudas analīze šādiem alternatīvajiem sadalījumiem (to blīvuma funkcijas dotas 2.3. tabulā).

Jaudas salīdzināsim ar Gini statistiku, kas dota formā [18]

$$G = \left[\sum_{i=1}^{n-1} \{ i(n-i)(X_{(i+1)} - X_{(i)}) \} \right] / \{ (n-1) \sum_{i=1}^{n} X_i \},$$

kur $X_{(1)} \leq \ldots \leq X_{(n)}$. Šī statistika ir ”visjaudīgākā pret alternatīvām” [18]. Empīriskā jaudas analīze tika veikta pie datu izlasēm $n = 20$ un $n = 50$ un pie 1000 atkārtojumiem.

2.2. Tabula: Testa $W_{S(\hat{\beta})}$ 5% un 10% kritiskās vērtības, pārbaudot normālo sadalījumu

		d(n)		
n	α	1	2	3
20	0.05	3.721	3.782	3.841
	0.10	2.667	2.672	2.734
50	0.05	3.914	3.993	4.106
	0.10	2.823	2.843	2.851

Salīdzinot empīriskās jaudas var redzēt, ka pie maza izlases apjoma $n = 20$ labāka ir Gini statistika, bet pie $n = 50$ Neimaņa tests ir līdzvērtīgs Gini statistikai (dažbrīd pat labāks par to).

2.3. Tabula: Alternatīvie sadalījumi eksponenciālajam sadalījumam

Alternatīvais sadalījums	Alternatīvā sadalījuma blīvuma funkcija
Weibull(b;k)	$bk(bx)^{k-1}\exp\{-(bk)^k\}, \quad x > 0$
Gamma(p;q)	$q^{-p}\{\Gamma(p)\}^{-1}x^{p-1}\exp(-x/q), \quad x > 0$
Beta (p;q)	$x^{p-1}(1-x)^{q-1}\{B(p,q)\}^{-1}, \quad 0 \leq x \leq 1$
Vienmērīgais (a;b)	$(b-a)^{-1}, \quad a \leq x \leq b$

2.2.4. Alternatīvie sadalījumi un simulētās jaudas normālajam sadalījumam

Šajā darbā, lai noteiktu normalitāti, tika veikta jaudas analīze šādiem alternatīvajiem sadalījumiem (to blīvuma funkcijas dotas 2.5. tabulā).

Empīriskās jaudas salīdzinātas ar Šapiro-Vilksa testu W, kas tiek uzskaitīts par labas precizitātes [18] testu normalitātes noteikšanai. Empīriskā jaudas analīze tikai veikta pie datu izlasēm $n = 20$ un $n = 50$ un pie 1000 atkārtojumiem.

Kā var redzēt, tad gan Neimaņa testam, gan Šapiro-Vilkes testam simulētās jaudas ir līdzīgas gan pie $n = 20$, gan $n = 50$, līdz ar to var teikt, ka Neimaņa tests var tikt lietots, nosakot datu normalitāti.

2.4. Tabula: Simulētās jaudas (%) G un $W_{S(\hat{\beta})}$, pārbaudot eksponencialitāti, $\alpha = 0.05$, d(20)=6 un d(50)=6

Alternatīvais sadalījums	G		$W_{S(\hat{\beta})}$	
	n=20	n=50	n=20	n=50
Weibull(1;0.7)	26	51	23	52
Weinbull(1; 1.5)	49	90	23	89
Weinbull(1; 1.6)	61	97	33	96
Gamma (1.6;1)	19	45	7	45
Gamma (2.6;1)	72	99	35	99
Gamma (0.5;1)	34	69	42	75
Beta(1;2)	91	99	89	99
Vienmērīgais (0;3)	69	98	52	99

2.5. Tabula: Alternatīvie sadalījumi normālajam sadalījumam

Alternatīvais sadalījums	Alternatīvā sadalījuma blīvuma funkcija
Logistic	$e^x(1+e^x)^{-2}$
Weibull(b;k)	$bk(bx)^{k-1}\exp\{-(bk)^k\}, \quad x > 0$
SC(p;d)	$(2\pi)^{-\frac{1}{2}}\left[(p/d)\exp\left(-\frac{1}{2}x^2/d^2\right) + (1-p)\exp\left(-\frac{1}{2}x^2\right)\right]$

2.3. Neimaņa tests saliktām hipotēzēm AR(1) procesam

Programmēšanas valodā **R** tika izveidota programma, kas realizē Neimaņa testu AR(1) procesam. Tā kā Neimaņa testa modifikācija AR(1) procesam ir pēc matemātiskās struktūras sarežģīta, tad programma darbojas salīdzinoši ilgi (ap diennakti uz laba stacionārā datora), līdz ar to šobrīd ir iegūti tikai robežsadalījumi pie H_0, kas apstiprina 1.4.2. nodaļā veikto modifikāciju Neimaņa testam AR(1) procesam. 2.3. attēlā redzams, kā tiek aproksimēts robežsadalījums pie $\phi = 0.1$ un $\phi = 0.3$.

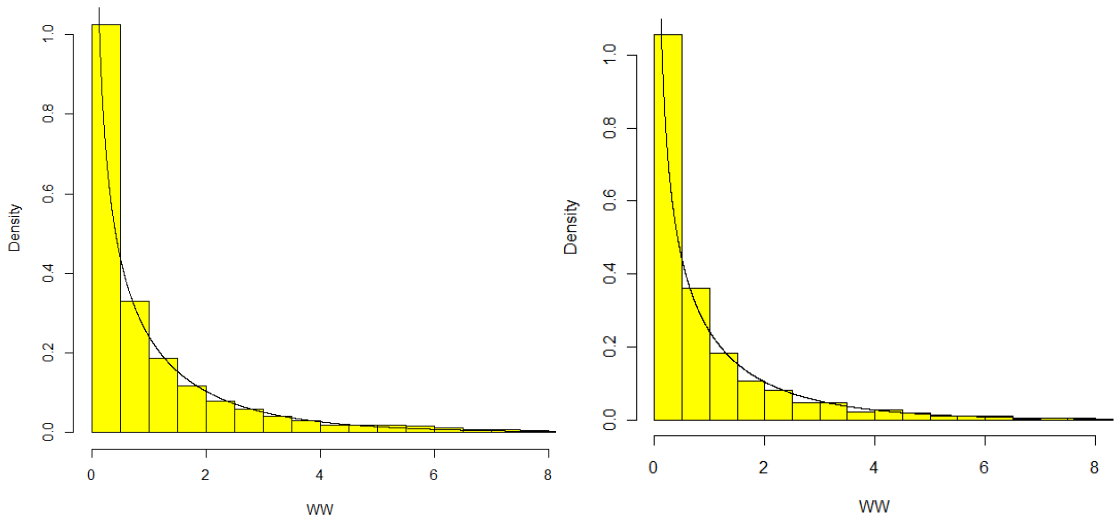
2.4. attēlā redzams, kā tiek aproksimēts robežsadalījums pie $\phi = 0.6$ un $\phi = 0.9$.

ϕ palielinoties jeb datu atkarībai pieaugot, robežsadalījums pie viena un tā paša atkārtojumu skaita $N = 1000$ vairs netiek tik labi aproksimēts. Kritiskās vētības pie būtiskuma līmeņa $\alpha = 0.05$ pie $\phi = 0.1$ ir 3.85 un pie $\phi = 0.3$ ir 3.84, kas praktiski sakrīt ar robežsadalījuma χ^2_1 kritisko vērtību 3.84, taču pie $\phi = 0.6$ kritiskā vērtība ir 2.87 un pie $\phi = 0.9$ vairs tikai 1.97. Tā kā atkārtojumu skaits ir tikai 1000, tad, iespējams, ka pie lielākām ϕ vērtībām statistika

2.6. Tabula: Simulētās jauda (%) W un $W_{S(\hat{\beta})}$, pārbaudot normalitāti, $\alpha = 0.05$, d(20)=3 un d(50)=3

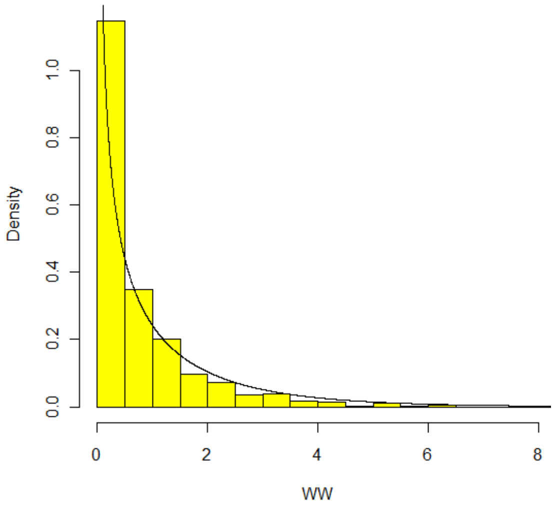
Alternatīvais sadalījums	W		$W_{S(\hat{\beta})}$	
	n=20	n=50	n=20	n=50
Logistic(1)	11	13	11	12
Weinbull(1; 2)	13	41	12	41
SC(0.1;5)	60	93	55	92
SC(0.05;7)	42	73	42	70

joprojām aproksimē robežsadalījumu, taču konverģence notiek lēnāk. Lai to varētu droši teikt, ir nepieciešamas papildus simulācijas, kas šobrīd vēl nav veiktas.

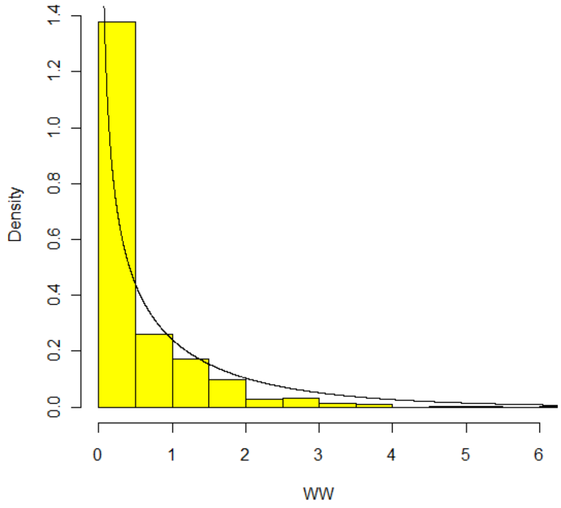


(a) N=1000, $\phi = 0.1$ (b) N=1000, $\phi = 0.3$

2.3. Att.: Robežsadalījuma pārbaude AR(1) procesam [1]



(a) N=1000, $\phi = 0.6$



(b) N=1000, $\phi = 0.9$

2.4. Att.: Robežsadalījuma pārbaude AR(1) procesam [2]

Secinājumi

Ir izveidota programma, kas implementē Neimaņa testu saliktām hipotēzēm neatkarīgiem datiem gan eksponecialitātes, gan normalitātes pārbaudei programmēšanas valodā R, kā arī dots apraksts, kā šī programma veidota. Ar izveidoto programmu varēja pārliecināties, ka gan eksponencialitātes, gan normalitātes gadījumā pie H_0 tiek aproksimēts Hī-kvadrāta sadalījums. Empīriski simulētās jaudas parāda, ka pie vidēja izlašu apjoma $n = 50$ Neimaņa tests saliktām hipotēzēm neatkarīgiem datiem ir konkurētspējīgs ar Gini statistiku eksponencialitātes pārbaudei un Šapiro-Vilksa testu normalitātes pārbaudei.

Izmantojot statistisko literatūru, tika modificēts Neimaņa testa saliktām hipotēzēm AR(1) procesam, kas ir pirmais solis, lai Neimaņa testu saliktām hipotēzēm modificētu atkarīgiem datiem. Ir izveidota programma, kas realizē Neimaņa testu saliktām hipotēzēm AR(1) procesam ar vienu polinomu programmēšanas valodā R. Šīs programmas kods dots darba pielikumā. Tā kā Neimaņa tests ir sarežģīts, tad programmai aprēķinu veikšana prasa ilgu laiku. No šobrīd iegūtajiem robežsadalījumiem var secināt, ka modifikācija ir pareiza, tomēr bažas rada fakts, ka ϕ tuvojoties 1 ("datiem paliekot atkarīgākiem"), pie atkārtojumu skaita $N = 1000$ robežsadalījums tomēr netiek līdz galam aproksimēts. Atbildi uz jautājumu, vai tas ir saistīts tikai ar to, ka pie tik lielām ϕ vērtībām konverģence uz sadalījumu notiek straujāk, var dot tikai simulācijas ar lielāku atkārtojumu skaitu.

Tādējādi var teikt, ka darba novitāte ir Neimaņa testa saliktām hipotēzēm modifikācija AR(1) procesam. Tomēr ir ieteikumi tālākai zinātniskajai darbībai—veikt plašākas simulācijas ar izstrādāto programmu AR(1) procesam, pilnveidot šo programmu, pievienot otro, trešo utt. ortonormālos Ležandra polinomus un Švarca selekcijas likumu, kas atkarībā no datiem izvēlās atbilstošo polinomu skaitu, kā arī turpināt attīstīt Neimaņa testu saliktām hipotēzēm atkarīgiem datiem, ne tikai AR(1) procesam.

Izmantotā literatūra un avoti

[1] D.V. Hinkley D.R. Cox. *Theoretical Statistics*. Chapman & Hall/CRC, 2000.

[2] D.J. Best J.C.W. Rayner. *Smooth Tests of Godness of Fit*. Oxford University Press, 1989.

[3] T.Ledwina W.C.M Kallenberg, T. Inglot. Power approximations and power comparison of certain godness-of-fit tests. *Journal of Statistics*, 21:131–145, 1994.

[4] T. Ledwina W. C.M. Kallenberg. Data driven version of neyman's smooth test of fit. *Journal of the American Statistical Association*, 89:1000–1005, 1994.

[5] T.Ledwina T. Inglot. Asymptotic optimality of data driven Neyman's tests for uniformity. *The Annals of Statistics*, 24(5):342–355, 1988.

[6] G. E.Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[7] T. Ledwina W. C.M. Kallenberg. Consistency and Monte Carlo Simulation of a Data Driven Version of Smooth Goodness-of-Fit Tests. *Annals of Statistics*, 23(5):1594–1608, 1995.

[8] T. Ledwina T. Inglot, W.C.M. Kallenberg. Data Driven Smooth Tests for Composite Hypotheses. *The Annals of Statistics*, 25(3):1222 – 1250, 1997.

[9] Matrix Inversion Identities. `http://www0.cs.ucl.ac.uk/staff/g.ridgway/mil/mil.pdf`.

[10] J. Rone. Jauktu procesu analīze un to pielietojums statistikā. Maģistra darbs, 2009.

[11] Minimum Dispersion and Unbiasedness: "Best" Linear Predictors for Stacionary ARMA α - Stable Processes. `http://www.colorado.edu/econ/papers/papers00/wp00-6.pdf`.

[12] Parameter estimation in low order fractionally differenced ARMA processes. `http://link.springer.com/article/10.1007%2FBF01544075`.

[13] Time Series Analysis. `http://www.etsii.upm.es/ingor/estadistica/Carol/TSAtema4petten.pdf`.

[14] A. T. Walden D. B. Percival. *Time Series and System Analysis with Applications*. John Wiley, 1983.

[15] Estimation of ARMA models. `http://faculty.washington.edu/ezivot/econ584/notes/armaestimation.pdf`.

[16] AR(1) Time Series Process. `http://www.math.utah.edu/~zhorvath/ar1.pdf`.

[17] AR(1) Time Series Models. `https://files.nyu.edu/mrg217/public/timeseries.pdf`.

[18] T. Ledwina W. C.M.Kallenberg. Data driven smooth tests for composite hypotheses: Comparison of powers. *Journal of Statistical Computation and Simulation*, 59(2):101–121, 1997.

A Neimaņa tests saliktām hipotēzēm neatkarīgiem datiem

Eksponencialitātes pārbaude

```
N<-1000
rob<-20
WW<-numeric()
n<-20
W.k<-numeric()
S<-numeric()
S.vert<-numeric()
P1<-function(x) (x-0.5)*sqrt(12)
P2<-function(x) sqrt(5)*(6*(x-0.5)^2-0.5)
P3<-function(x) sqrt(7)*(20*(x-0.5)^3-3*(x-0.5))
P4<-function(x) 210*(x-0.5)^4-45*(x-0.5)^2+9/8
P5<-function(x) sqrt(11)*(63*(2*x-1)^5-70*(2*x-1)^3+15*(2*x-1))/8
P6<-function(x) sqrt(13)*(231*(2*x-1)^6-315*(2*x-1)^4+105*(2*x-1)^2-5)/16
blivuma.fja<-function(x,l) l*exp(-l*x)
sadalijuma.fja<-function(z,l) integrate(function(x) blivuma.fja(x,l),0,z)$value
sadalijuma.fja<-Vectorize(sadalijuma.fja,vectorize.arg="z")
fl<-function(x,l) exp(-l*x)-l*x*exp(-l*x) #atvasin?jums p?c l
int.fl <-function(z,l) integrate(function(x) fl(x,l),0,z)$value
int.fl<-Vectorize(int.fl, vectorize.args="z")
ceriba.fl<-function(l) integrate(function(z) sqrt(12)*int.fl(z,l)*blivuma.fja(z,l),0,rob)$value
ceriba.fl.2<-function(l) integrate(function(z) sqrt(5)*6*2*(sadalijuma.fja(z,l)-0.5)*int.fl(z,l)*blivuma.fja(z,l),0,rob)$value
ceriba.fl.3<-function(l) integrate(function(z) sqrt(7)*(20*3*(sadalijuma.fja(z,l)-0.5)^2
*int.fl(z,l)-3*int.fl(z,l))*blivuma.fja(z,l),0,rob)$value
ceriba.fl.4<-function(l) integrate(function(z) (210*4*
(sadalijuma.fja(z,l)-0.5)^3*int.fl(z,l)-45*2*
(sadalijuma.fja(z,l)-0.5)*int.fl(z,l))*blivuma.fja(z,l),0,rob)$value
ceriba.fl.5<-function(l) integrate(function(z) (sqrt(11)*(63*5*(2*sadalijuma.fja(z,l)-1)^4
*2*int.fl(z,l)-70*3*(2*sadalijuma.fja(z,l)-1)^2*2*int.fl(z,l)
+15*2*int.fl(z,l))/8)*blivuma.fja(z,l),0,rob)$value
ceriba.fl.6<-function(l) integrate(function(z) sqrt(13)*(231*6*(2*sadalijuma.fja(z,l)-1)^5*
2*int.fl(z,l)-315*4*(2*sadalijuma.fja(z,l)-1)^3*2*int.fl(z,l)+105*2*(2*sadalijuma.fja(z,l)
-1)*2*int.fl(z,l))/16*blivuma.fja(z,l),0,rob)$value
log.fl<-function(l) -1/l^2 # log un tad dubultais atvasin?jums p?c l
ceriba.log.fl<-function(l) -1/l^2
for(k in 1:N)
{
dati<-rexp(n,1)
l<-(mean(dati))^(-1)
I.beta<-c(-ceriba.fl(l))
I.beta.2<-matrix(c(-ceriba.fl(l),-ceriba.fl.2(l)),ncol=2)
I.beta.3<-matrix(c(-ceriba.fl(l),-ceriba.fl.2(l),-ceriba.fl.3(l)),ncol=3)
I.beta.4<-matrix(c(-ceriba.fl(l),-ceriba.fl.2(l),-ceriba.fl.3(l),-ceriba.fl.4(l)),ncol=4)
I.beta.5<-matrix(c(-ceriba.fl(l),-ceriba.fl.2(l),-ceriba.fl.3(l),-ceriba.fl.4(l),-ceriba.fl.5(l)),ncol=5)
I.beta.6<-matrix(c(-ceriba.fl(l),-ceriba.fl.2(l),-ceriba.fl.3(l),-ceriba.fl.4(l),-ceriba.fl.5(l),-ceriba.fl.6(l)),ncol=6)
Y.n<-sum(P1(sadalijuma.fja(dati,l)))/n
Y.n.2<-matrix(c(sum(P1(sadalijuma.fja(dati,l)))/n,sum(P2(sadalijuma.fja(dati,l)))/n),ncol=2)
```

```
Y.n.3<-matrix(c(sum(P1(sadalijuma.fja(dati,l)))/n,sum(P2(sadalijuma.fja(dati,l)))/n,
sum(P3(sadalijuma.fja(dati,l)))/n),ncol=3)
Y.n.4<-matrix(c(sum(P1(sadalijuma.fja(dati,l)))/n,
sum(P2(sadalijuma.fja(dati,l)))/n, sum(P3(sadalijuma.fja(dati,l)))/n,
sum(P4(sadalijuma.fja(dati,l)))/n),ncol=4)
Y.n.5<-matrix(c(sum(P1(sadalijuma.fja(dati,l)))/n,
sum(P2(sadalijuma.fja(dati,l)))/n, sum(P3(sadalijuma.fja(dati,l)))/n,
sum(P4(sadalijuma.fja(dati,l)))/n,sum(P5(sadalijuma.fja(dati,l)))/n),ncol=5)
Y.n.6<-matrix(c(sum(P1(sadalijuma.fja(dati,l)))/n,
sum(P2(sadalijuma.fja(dati,l)))/n, sum(P3(sadalijuma.fja(dati,l)))/n,sum(P4(sadalijuma.fja(dati,l)))/n,sum(P5(sadalijuma.fja(dati,l)))/n,
sum(P6(sadalijuma.fja(dati,l)))/n),ncol=6)
I.beta.beta<-c(-ceriba.log.fl(l))
I<-matrix(1)
II<-matrix(c(1,0,0,1),ncol=2)
III<-matrix(c(1,1,1,1,1,1,1,1,1),ncol=3)
IV<-matrix(c(1,0,0,0,0,1,0,0,0,0,1,0,0,0,0,1),ncol=4)
V<-matrix(c(1,0,0,0,0,0,1,0,0,0,0,0,1,0,0,0,0,0,1,0,0,0,0,0,1),ncol=5)
VI<-matrix(c(1,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,1),ncol=6)
R.beta<-I.beta%*%(I.beta.beta-(I.beta%*%t(I.beta)))^(-1)%*%I.beta
R.beta.2<-t(I.beta.2)%*%(I.beta.beta-(I.beta.2%*%t(I.beta.2)))^(-1)%*%I.beta.2
R.beta.3<-t(I.beta.3)%*%(I.beta.beta-(I.beta.3%*%t(I.beta.3)))^(-1)%*%I.beta.3
R.beta.4<-t(I.beta.4)%*%(I.beta.beta-(I.beta.4%*%t(I.beta.4)))^(-1)%*%I.beta.4
R.beta.5<-t(I.beta.5)%*%(I.beta.beta-(I.beta.5%*%t(I.beta.5)))^(-1)%*%I.beta.5
R.beta.6<-t(I.beta.6)%*%(I.beta.beta-(I.beta.6%*%t(I.beta.6)))^(-1)%*%I.beta.6
W.k[1]<-n%*%Y.n%*%(I+R.beta)%*%Y.n
W.k[2]<-n%*%Y.n.2%*%(II+R.beta.2)%*%t(Y.n.2)
W.k[3]<-n%*%Y.n.3%*%(III+R.beta.3)%*%t(Y.n.3)
W.k[4]<-n%*%Y.n.4%*%(IV+R.beta.4)%*%t(Y.n.4)
W.k[5]<-n%*%Y.n.5%*%(V+R.beta.5)%*%t(Y.n.5)
W.k[6]<-n%*%Y.n.6%*%(VI+R.beta.6)%*%t(Y.n.6)
S[1]<-n*Y.n^2-log(n)
S[2]<-n*(Y.n.2[1]*Y.n.2[1]+Y.n.2[2]*Y.n.2[2])-2*log(n)
S[3]<-n*(Y.n.3[1]*Y.n.3[1]+Y.n.3[2]*Y.n.3[2]+Y.n.3[3]*Y.n.3[3])-3*log(n)
S[4]<-n*(Y.n.4[1]*Y.n.4[1]+Y.n.4[2]*Y.n.4[2]+Y.n.4[3]*Y.n.4[3]+Y.n.4[4]*Y.n.4[4])-4*log(n)
S[5]<-n*(Y.n.5[1]*Y.n.5[1]+Y.n.5[2]*Y.n.5[2]+Y.n.5[3]*Y.n.5[3]+Y.n.5[4]*Y.n.5[4]
+Y.n.5[5]*Y.n.5[5])-5*log(n)
S[6]<-n*(Y.n.6[1]*Y.n.6[1]+Y.n.6[2]*Y.n.6[2]+Y.n.6[3]*Y.n.6[3]+Y.n.6[4]*Y.n.6[4]
+Y.n.6[5]*Y.n.6[5]+Y.n.6[6]*Y.n.6[6])-6*log(n)
K<-S*(-1)
K.jauns<-order(K)
S.vertiba<-K.jauns[1]
S.vert[k]<-S.vertiba
if (S.vertiba==1){W<-W.k[1]}
if (S.vertiba==2){W<-W.k[2]}
if (S.vertiba==3){W<-W.k[3]}
if (S.vertiba==4){W<-W.k[4]}
if (S.vertiba==5){W<-W.k[5]}
if (S.vertiba==6){W<-W.k[6]}
WW[k]<-W
}
S.vert
WWW<-sort(WW)
Krit<-WWW[N*0.95]
Krit
rez<-hist(WW,prob=T)
#rez<-hist(WW,prob=T,200,xlim=range(0:30),xlab='', ylab='', main='theta=-0.9')
xx<-seq(0,10,by=0.001)
lines(xx,dchisq(xx,1))
```

Normalitātes pārbaude

```
N<-10000
rob<-20
WW<-numeric()
W.k.<-numeric()
S<-numeric()
S.vert<-numeric()
```

```
blivuma.fja<-function(z,m,s) 1/sqrt(2*pi*s)*exp(-(z-m)^2/2/s)
sadalijuma.fja<-function(z,m,s) integrate(function(x) blivuma.fja(x,m,s),-rob,z)$value
sadalijuma.fja<-Vectorize(sadalijuma.fja,vectorize.arg="z")
fm<-function(x,m,s) 1/2*sqrt(2)*(x-m)*exp(-(x-m)^2/2/s)/sqrt(pi*s)/s
fs<-function(x,m,s) 1/4*sqrt(2)*exp(-1/2*(-x+m)^2/s)*(-s+x^2-2*x*m+m^2)/sqrt(pi*s)/s^2
int.fm <-function(z,m,s) integrate(function(x) fm(x,m,s),-rob,z)$value
ceriba.fm<-function(m,s) integrate(function(x) sqrt(12)*int.fm(x,m,s)*blivuma.fja(x,m,s),-rob,rob)$value
ceriba.fm.2<-function (m,s)integrate(function(x) sqrt(5)*6*2*(sadalijuma.fja(x,m,s)-0.5)*int.fm(x,m,s)*blivuma.fja(x,m,s),-rob,rob)$value
int.fs <-function(z,m,s) integrate(function(x) fs(x,m,s),-rob,z)$value
ceriba.fs<-function(m,s) integrate(function(x) sqrt(12)*int.fs(x,m,s)*blivuma.fja(x,m,s),-rob,rob)$value
ceriba.fs.2<-function (m,s)integrate(function(x) sqrt(5)*6*2*(sadalijuma.fja(x,m,s)-0.5)*int.fs(x,m,s)*blivuma.fja(x,m,s),-rob,rob)$value
log.fmm<-function(x,m,s) -1/s
ceriba.log.fmm<-function(m,s) integrate(function(x) log.fmm(x,m,s)*blivuma.fja(x,m,s),-rob,rob)$value
log.fss<-function(x,m,s) -1/2*(-s+2*x^2-4*x*m+2*m^2)/s^3
ceriba.log.fss<-function(m,s) integrate(function(x) log.fss(x,m,s)*blivuma.fja(x,m,s),-rob,rob)$value
log.fms<-function(x,m,s) (m-x)/s^2
ceriba.log.fms<-function(m,s) integrate(function(x) -sqrt(2)*(x-m)*exp(-1/2*(x-m)^2/s^2)/sqrt(pi)/s^4,-rob,rob)$value
P1<-function(x) (x-0.5)*sqrt(12)
P2<-function(x) sqrt(5)*(6*(x-0.5)^2-0.5
for(l in 1:N)
{
n<-50
dati<-rnorm(n,0,1)
m<-mean(dati)
s<-(sd(dati))^2
Y.n<-sum(P1(sadalijuma.fja(dati,m,s)))/n
Y.n.2<-matrix(c(sum(P1(sadalijuma.fja(dati,m,s)))/n,sum(P2(sadalijuma.fja(dati,m,s)))/n), ncol=2)
I.beta<-matrix(c(-ceriba.fm(m,s),-ceriba.fs(m,s)),nrow=2,ncol=1)
I.beta.2<-matrix(c(-ceriba.fm(m,s),-ceriba.fs(m,s),-ceriba.fm.2(m,s),-ceriba.fs.2(m,s)),ncol=2)
I.beta.beta<-matrix(c(-ceriba.log.fmm(m,s),-ceriba.log.fms(m,s),-ceriba.log.fms(m,s),-ceriba.log.fss(m,s)),nrow=2,ncol=2)
R.beta<-t(I.beta)%*%solve(I.beta.beta-(I.beta%*%t(I.beta)))%*%I.beta
R.beta.2<-t(I.beta.2)%*%(I.beta.beta-(I.beta.2%*%t(I.beta.2)))^(-1)%*%I.beta.2
I<-matrix(1)
II<-matrix(c(1,0,0,1),ncol=2)
W.k[1]<-n*Y.n*(I+R.beta)*Y.n
W.k[2]<-n%*%Y.n.2%*%(II+R.beta.2)%*%t(Y.n.2)
S[1]<-n*Y.n^2-log(n)
S[2]<-n*(Y.n.2[1]*Y.n.2[1]+Y.n.2[2]*Y.n.2[2])-2*log(n)
print(S)
K<-S*(-1)
K.jauns<-order(K)
print(K.jauns)
S.vertiba<-K.jauns[1]
S.vert[l]<-S.vertiba
print(S.vert[l])
if (S.vertiba==1){W<-W.k[1]}
if (S.vertiba==2){W<-W.k[2]}
WW[l]<-W
}
WW
WWW<-sort(WW)
Krit<-WWW[950]
rez<-hist(WW,prob=T)
xx<-seq(0,20,by=0.001)
lines(xx,dchisq(xx,2))
cat(file="neatkarigie.txt",WWW)
```

B Neimaņa tests saliktām hipotēzēm AR(1) procesam

```
N<-500
rob<-20
theta<- 0.9
WW<-numeric()
blivuma.fja.pirmais<-function(z,t,m,s) 1*sqrt(1-t^2)/sqrt(2*pi)/s*exp(-(z-m)^2*(1-t^2)/2/s^2)
blivuma.fja.parejie<-function(z,z1,t,m,s) 1/sqrt(2*pi)/s*exp(-(z-t*z1)^2/2/s^2)
sadalijuma.fja.pirmais<-function(z,t,m,s) integrate(function(x) blivuma.fja.pirmais(x,t,m,s),-rob,z)$value
sadalijuma.fja.pirmais<-Vectorize(sadalijuma.fja.pirmais,vectorize.arg="z")
sadalijuma.fja.parejie<-function(z,z1,t,m,s) integrate(function(x) blivuma.fja.parejie(x,z1,t,m,s),-rob,z)$value
sadalijuma.fja.parejie<-Vectorize(sadalijuma.fja.parejie,vectorize.arg="z")
fm.pirmais<-function(x,t,m,s) 0
fm.parejie<- function (x,z1,t,m,s) 0
fs.pirmais<-function(x,t,m,s) -1/2*sqrt(1-t^2)*sqrt(2)*exp(1/2*(t-1)*(t+1)*x^2)/s^2*(s^2-x^2+x^2*t^2)/s^3/sqrt(pi*s^2)
fs.pirmais<-Vectorize(fs.pirmais,vectorize.args="x")
fs.parejie<-function(x,z1,t,m,s)-1/2*sqrt(2)*exp(-1/2*(-x+t*z1)^2/s^2)*(s^2-x^2+2*x*t*z1-t^2*z1^2)/s^3/sqrt(pi*s^2)
fs.parejie<-Vectorize(fs.parejie,vectorize.args="x")
int.fm.pirmais <-function(z,t,m,s) integrate(function(x) fm.pirmais(x,t,m,s),-rob,z)$value
int.fm.pirmais<-Vectorize(int.fm.pirmais, vectorize.args="z")
int.fm.parejie <-function(z,z1,t,m,s) integrate(function(x) fm.parejie(x,z1,t,m,s),-rob,z)$value
int.fm.parejie<-Vectorize(int.fm.parejie, vectorize.args="z")
int.fs.pirmais <-function(z,t,m,s) integrate(function(x) fs.pirmais(x,t,m,s),-rob,z)$value
int.fs.pirmais<-Vectorize(int.fs.pirmais, vectorize.args="z")
int.fs.parejie <-function(z,z1,t,m,s) integrate(function(x) fs.parejie(x,z1,t,m,s),-rob,z)$value
int.fs.parejie<-Vectorize(int.fs.parejie, vectorize.args="z")
ceriba.fm.pirmais<-function(t,m,s) integrate(function(z) sqrt(12)*int.fm.pirmais(z,t,m,s)*blivuma.fja.pirmais(z,t,m,s),-rob,rob)$value
ceriba.fm.parejie<-function(z1,t,m,s) integrate(function(z) sqrt(12)*int.fm.parejie(z,z1,t,m,s)*blivuma.fja.parejie(z,z1,t,m,s),-rob,rob)$value
ceriba.fs.pirmais<-function(t,m,s) integrate(function(z) sqrt(12)*int.fs.pirmais(z,t,m,s)*blivuma.fja.pirmais(z,t,m,s),-rob,rob)$value
ceriba.fs.parejie<-function(z1,t,m,s) integrate(function(z) sqrt(12)*int.fs.parejie(z,z1,t,m,s)*blivuma.fja.parejie(z,z1,t,m,s),-rob,rob)$value
ceriba.fs.parejie<-Vectorize(ceriba.fs.parejie,vectorize.args="z1")
log.fmm<-0
ceriba.log.fmm<-0
log.fms<-0
ceriba.log.fms<-0
log.fss.pirmais<-function(x,t,m,s) (s^2-3*x^2+3*x^2*t^2)/s^4
log.fss.parejie<-function(x,z1,t,m,s) (s^2-3*t^2*z1^2+6*x*t*z1-3*x^2)/s^4
ceriba.log.fss.pirmais<-function(t,m,s) integrate(function(x) log.fss.pirmais(x,t,m,s)*blivuma.fja.pirmais(x,t,m,s),-rob,rob)$value
ceriba.log.fss.parejie<-function(z1,t,m,s) integrate(function(x) log.fss.parejie(x,z1,t,m,s)*blivuma.fja.parejie(x,z1,t,m,s),-rob,rob)$value
ceriba.log.fss.parejie<-Vectorize(ceriba.log.fss.parejie,vectorize.args="z1")
P1<-function(x) (x-0.5)*sqrt(12)
n<-100
for(l in 1:N)
{
dati<-arima.sim(n=100,list(ar=c(theta)),
               rand.gen=function(n, ...) sqrt(1-theta^2)*rnorm(n))
dati3<-numeric()
for (i in 2:n)
{
dati3[i-1]<-dati[i]
```

```
i<-i+1
}
dati4<-numeric()
for(i in 1:(n-1))
{
dati4[i]<-dati[i]
i<-i+1
}
m<-0
s<-sd(dati)
t<-theta
summa<-0
for (k in 1:(n-1))
{
ss<-P1(sadalijuma.fja.parejie(dati3[k],dati4[k],t,m,s))
summa<-summa+ss
ss<-ss+1
}
Y.n<-(summa+P1(sadalijuma.fja.pirmais(dati[1],t,m,s)))/n
I.beta<-matrix(-(ceriba.fs.pirmais(t,m,s)+sum(ceriba.fs.parejie(dati4,t,m,s)))/n,nrow=1,ncol=1)
I.beta.beta<-matrix(-(ceriba.log.fss.pirmais(t,m,s)+sum(ceriba.log.fss.parejie(dati4,t,m,s)))/n,nrow=1,ncol=1)
R.beta<-t(I.beta)%*%solve(I.beta.beta-(I.beta%*%t(I.beta)))%*%I.beta
I<-matrix(1)
W.k<-n*Y.n*(I+R.beta)*Y.n
WW[l]<-W.k
}
WW
WWW<-sort(WW)
Krit<-WWW[950]
rez<-hist(WW,prob=T)
xx<-seq(0,20,by=0.001)
lines(xx,dchisq(xx,1))
```

C Programma *Maple* veiktie aprēķini

```
> f := proc (x, l) options operator, arrow; l*exp(-l*x) end proc;
(x, l) -> l exp(-l x)
> f(x, l);
                                 l exp(-l x)
> ff := proc (x, l) options operator, arrow; diff(f(x, l), l) end proc;
           d
(x, l) -> --- f(x, l)
           dl
> ff(x, l);
                          exp(-l x) - l x exp(-l x)
> f3 := proc (x, l) options operator, arrow; ln(f(x, l)) end proc;
(x, l) -> ln(f(x, l))
> f3(x, l);
                               ln(l exp(-l x))
> simplify(f3(x, l));
                               ln(l exp(-l x))
> f4 := proc (x, l) options operator, arrow; diff(f3(x, l), l) end proc;
           d
(x, l) -> --- f3(x, l)
           dl
> f4(x, l);
                          exp(-l x) - l x exp(-l x)
                          -------------------------
                                 l exp(-l x)
> f5 := proc (x, l) options operator, arrow; diff(f4(x, l), l) end proc;
           d
(x, l) -> --- f4(x, l)
           dl
> f5(x, l);
                             2
         -2 x exp(-l x) + l x  exp(-l x)   exp(-l x) - l x exp(-l x)
         ------------------------------- - -------------------------
                   l exp(-l x)                    2
                                                 l  exp(-l x)

              (exp(-l x) - l x exp(-l x)) x
            + -----------------------------
                        l exp(-l x)
> simplify(f5(x, l));
                                     1
                                   - --
                                      2
                                     l
> f := proc (x, l) options operator, arrow; l*exp(-l*x) end proc;
(x, l) -> l exp(-l x)
> f(x, l);
                                 l exp(-l x)
> ff := proc (x, l) options operator, arrow; diff(f(x, l), l) end proc;
           d
(x, l) -> --- f(x, l)
           dl
> ff(x, l);
                          exp(-l x) - l x exp(-l x)
```

```
> f3 := proc (x, l) options operator, arrow; ln(f(x, l)) end proc;
(x, l) -> ln(f(x, l))
> f3(x, l);
                               ln(l exp(-l x))
> simplify(f3(x, l));
                               ln(l exp(-l x))
> f4 := proc (x, l) options operator, arrow; diff(f3(x, l), l) end proc;
           d
(x, l) -> --- f3(x, l)
           dl
> f4(x, l);
                          exp(-l x) - l x exp(-l x)
                          -------------------------
                                 l exp(-l x)
> f5 := proc (x, l) options operator, arrow; diff(f4(x, l), l) end proc;
           d
(x, l) -> --- f4(x, l)
           dl
> f5(x, l);
                             2
         -2 x exp(-l x) + l x  exp(-l x)   exp(-l x) - l x exp(-l x)
         ------------------------------- - -------------------------
                   l exp(-l x)                     2
                                                  l  exp(-l x)

              (exp(-l x) - l x exp(-l x)) x
            + -----------------------------
                        l exp(-l x)
> simplify(f5(x, l));
                                      1
                                    - --
                                       2
                                      l
```

Maģistra darbs “Neimaņa tests saliktām hipotēzēm” 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: Anna Jansone

______________________ ________________

(paraksts) (datums)

Rekomendēju darbu aizstāvēšanai.

Vadītājs: LU docents Dr. Math. Jānis Valeinis

______________________ ________________

(paraksts) (datums)

Recenzents: LU lektors Jānis Smotrovs

______________________ ________________

(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)