

LATVIJAS UNIVERSITĀTE  
FIZIKAS UN MATEMĀTIKAS FAKULTĀTE  
MATEMĀTIKAS NODAĻA

**EMPĪRISKĀS TICAMĪBAS FUNKCIJAS METODES  
ANALĪZE UN TĀS PIELIETOJUMI STATISTIKĀ**

DIMPLOMDARBS

Autors: **Sandra Vucāne**

Stud. apl. MaSt 030013

Darba vadītājs: docents Dr. Math. Jānis Valeinis

Rīga 2008

## Anotācija

Bieži vien praksē sastopamajām datu kopām ir ļoti grūti noteikt sadalījuma likumu un tam var būt visdažādākie iemesli. Viena no iespējām, kā šādos gadījumos novērtēt parametrus, ir lietot neparametriskās metodes un viena no populārākajām šāda veida metodēm ir empīriskās tīcamības metode (EL).

Šī darba mērķis ir noskaidrot, cik labi EL metode strādā, vai tā spēj konkurēt ar citām metodēm. Simulējot dažādu sadalījumu datus, ar dažādām metodēm tika konstruēti tīcamības intervāli vidējai vērtībai un mediānai, kas tika salīdzināti, izmantojot tīcamības intervālu pārklājuma precizitāti.

Galvenie darbā iegūtie rezultāti liecina, ka EL metode strādā labi un ka tā ir nopietns sāncensis parametriskajām metodēm.

## **Anotation**

Very often in practical data analysis it is very hard to determine the real distribution of data set and there can be many reasons for that. One of the possibilities in such cases is to use nonparametric methods to estimate parameters, one of the most popular is empirical likelihood (EL) method.

The main purpose of this work is to investigate, how good EL method is, comparing with other methods. For this purpose, there was simulated data from different distributions and different methods were used to construct confidence intervals for mean and median, these methods were compared using coverage accuracy.

The main results are that EL method works very well and that it is a serious competitor to parametric methods.

# Saturs

<b>1 Ievads</b>	<b>6</b>
<b>2 Pamatjēdzieni un definīcijas</b>	<b>8</b>
<b>3 Vislielākās ticamības attiecības tests</b>	<b>10</b>
3.1 Vislielākās ticamības funkcija un metode . . . . .	10
3.2 Vislielākās ticamības attiecības tests . . . . .	12
<b>4 Neparametriskās ticamības funkcijas attiecības tests</b>	<b>16</b>
4.1 Neparametriskā vislielākās ticamības funkcija . . . . .	16
4.2 Neparametriskās ticamības funkcijas attiecības tests . . . . .	17
4.3 Ticamības intervālu konstruēšana vidējai vērtībai ar EL . . . . .	17
4.4 EL vispārējā gadījumā . . . . .	18
<b>5 Vislielākās ticamības attiecības testu salīdzinājums vidējai vērtībai</b>	<b>22</b>
5.1 Nūtona algoritms . . . . .	22
5.2 Parametriskās un neparametriskās metodes salīdzinājums . . . . .	22
<b>6 Bartleta korekcija vidējai vērtībai</b>	<b>28</b>
6.1 Pārklājuma precizitāte . . . . .	28
6.2 Bartleta korekcija . . . . .	29
<b>7 Ticamības intervālu konstruēšana kvantilēm ar EL</b>	<b>35</b>
7.1 Kodolu blīvuma funkcijas novērtēšana . . . . .	35
7.2 EL metodes ticamības intervālu konstruēšanai kvantilēm . . . . .	36
7.3 EL metožu salīdzinājums . . . . .	41
<b>8 Pielietojumi</b>	<b>46</b>
<b>9 Secinājumi</b>	<b>50</b>
<b>10 Izmantotā literatūra</b>	<b>52</b>
<b>A Pielikums</b>	<b>54</b>
A.1 Programma simulāciju veikšanai vidējai vērtībai . . . . .	54
A.2 Programma simulāciju veikšanai mediānai . . . . .	57

A.3 Programma $\chi^2$ kritērija pārbaudei . . . . .	66
--	----

# 1 Ievads

Vislielākās ticamības metode ir viena no plašāk lietotajām metodēm parametru novērtēšanā parametrisko modeļu gadījumā. Taču bieži vien praksē sastopamajām datu kopām ir ļoti grūti noteikt sadalījuma likumu un tam par iemeslu var būt nepietiekošs izlases apjoms, vai arī tas, ka teorētiskais sadalījums ir tik sarežģīts, ka tā matemātisko izteiksmi ir ļoti grūti vai pat praktiski neiespējami iegūt. Statistikā daudzas procedūras (testi, hipotēžu pārbaudes, ticamības intervālu konstruēšana) balstās uz parametriskiem pieņēmumiem par datu sadalījumu. Nosakot datiem nepareizu sadalījumu un veicot datu analīzi, kas balstīta uz šo pieņēmumu, var iegūt pat ļoti neprecīzus rezultātus. Rodas jautājums, kā veikt datu analīzi, novērtēt parametrus, ja datu sadalījumu nevar noteikt. Viena no iespējām ir parametru novērtēšanā lietot neparametriskās metodes un viena no populārākajām šāda veida metodēm ir empīriskās ticamības metode (turpmāk – EL). Owen ([10], [11], [12]) iepazīstināja ar EL metodi ticamības reģionu konstruēšanai. EL pirmsākumi meklējami jau Thomas, Grunkemeier [19] darbos saistībā ar izdzīvošanas varbūtību novērtēšanu. Galvenā šīs metodes priekšrocība ir tāda, ka EL ticamības reģionu forma nebalstās uz pētāmā sadalījuma neesošu simetriju, bet gan uz pētamo datu kopu, t.i., EL ticamības reģioni nav simetriski. Uz doto brīdi visvairāk izpētītais gadījums ir, kad novērtējamais parametrs var tikt izteikts kā gluda funkcija no vidējās vērtības. Tieši šim gadījumam ir parādīts, ka ar Bartleta korekcijas palīdzību var uzlabot pārklājuma klūdu no  $O(n^{-1})$  uz  $O(n^{-2})$ . Pat visparastākajām uz normālā sadalījuma aproksimāciju balstītajām metodēm, kas konstruē divpusējos ticamības intervālus vidējai vērtībai, pārklājuma klūda ir ar kārtu  $n^{-1}$ .

Šī darba mērķis ir noskaidrot, cik labi EL metode strādā, vai tā spēj konkurēt ar citām metodēm. Lai sasniegtu darba mērķi tika izvirzīti vairāki darba uzdevumi:

1. iepazīties ar empīriskās ticamības attiecības testu;
2. teorētiski, veicot simulācijas, pētīt EL metodes efektivitāti;
3. pielietot EL metodi reālai datu problēmai.

Visas darbā atspoguļotās simulācijas tika veiktas, izmantojot programmā R uzrakstītas datorprogrammas. Darbs sastāv no 9 nodaļām un 1 pielikuma. 2. nodaļā definēti daži pamatjēdzieni un sadalījumi, kas izmantoti darbā, 3. nodaļā ir aprakstīts vislielākās ticamības attiecības tests, kas ir viens no populārākajiem testiem, ko lieto parametru novērtēšanā un ticamības intervālu konstruēšanā, 4. nodaļā ir apskatīts neparametriskās ticamības attiecības tests, EL metode ticamības intervālu konstruēšanai vidējai vērtībai un EL metode vispārīgajā gadījumā, 5. nodaļā, izmantojot veiktās simulācijas, ir salīdzinātas dažas parametriskās metodes ar EL

metodi vidējās vērtības gadījumā, 6. nodaļā ir aprakstīta Bartleta korekcija vidējās vērtības gadījumā un ar simulāciju palīdzību parādīts kādu uzlabojumu šī korekcija dod. Nākamajā nodaļā ir aprakstītas 3 EL metodes ticamības intervālu konstruēšanai kvantilēm un ar simulāciju palīdzību šīs metodes ir salīdzinātas mediānas gadījumā. 8. nodaļā EL metode ir pielietota reālai datu problēmai-Latvijas pensiju vidējo vērtību izpētei, 9. nodaļā ir aprakstīti galvenie darbā iegūtie rezultāti un secinājumi. Pēdējā nodaļā ir ievietots izmantotās literatūras saraksts. Lai analizētu un pētītu EL metodes efektivitāti, programmā R tika uzrakstītas vairākas datorprogrammas, dažas no tām ir ievietotas pielikumā.

## 2 Pamatjēdzieni un definīcijas

Šajā nodaļā apskatīsim dažus darbā izmantotos pamatjēdzienus, un tā kā darbā (simulācijās un piemēros) tiek izmantoti dažādi sadalījumi: normālais, dubultais eksponenciālais, Stjūdenta,  $\chi^2$ , Košī sadalījums, tad apskatīsim dažus mazāk pazīstamos sadalījumus. Jēdzienu un sadalījumu definēšanā izmantoti ([1], [2]).

**Definīcija 1.** Gadījuma lielums  $X$  ir nepārtraukts, ja eksistē funkcija  $f$  tāda, ka  $f(x) \geq 0$   $\forall x \in R$ ,  $\int_{-\infty}^{+\infty} f(x)dx = 1$  un  $\forall a \leq b$ ,

$$P(a < X < b) = \int_a^b f(x)dx,$$

kur funkciju  $f$  sauc par varbūtību blīvuma funkciju.

$$F(x) = \int_{-\infty}^x f(t)dt$$

un  $f(x) = F'(x)$  visos punktos  $x$ , kuros  $F$  ir diferencējama.

**Definīcija 2.** Ja  $X$  ir diskrēts gadījuma lielums, kas pieņem vērtības  $\{x_1, \dots, x_n\}$ , tad var definēt varbūtību funkciju (varbūtību masas funkciju)  $f(x) = P(X = x)$ .

**Definīcija 3.** Gadījuma lieluma  $X$  kumulatīvā sadalījuma funkcija (turpmāk – sadalījuma funkcija) ir  $F(x) = P(X \leq x)$ , kur  $-\infty < x < \infty$ .

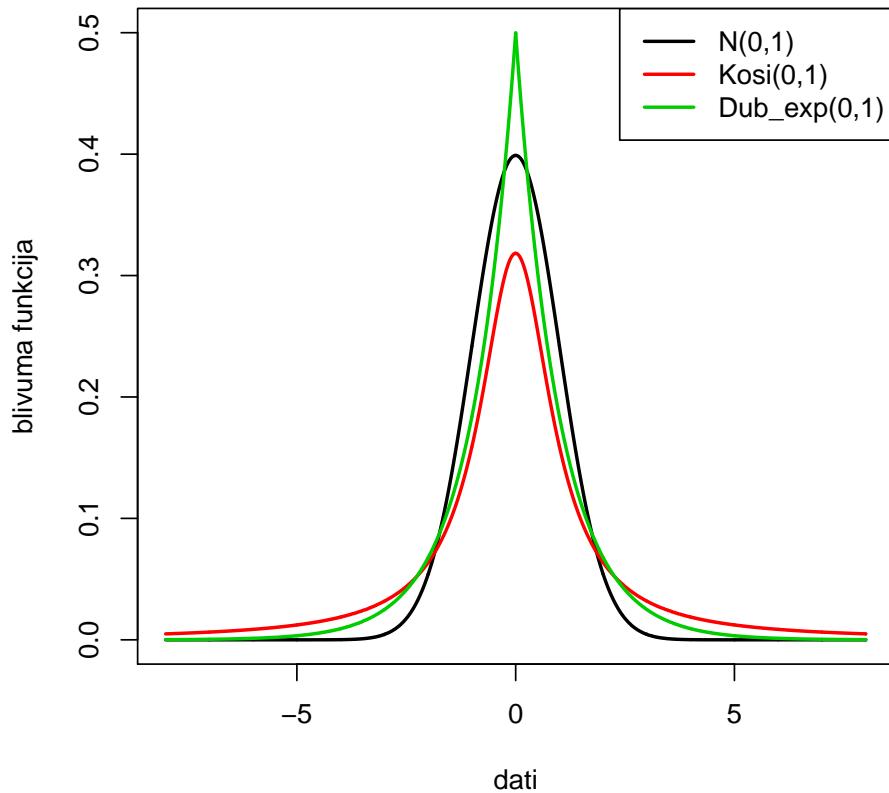
Ja definēsim  $F(x-) = P(X < x)$ , tad  $P(X = x) = F(x) - F(x-)$ .

### Dubultais eksponenciālais sadalījums.

Dubultā eksponenciālā sadalījuma blīvuma funkcija ir  $f(x|\mu, \beta) = (2\beta)^{-1} \exp\{-|x - \mu|/\beta\}$ , ja  $\mu = 0$  un  $\beta = 1$ , tad to sauc par standarta dubulto eksponenciālo sadalījumu un  $f(x|0, 1) = 2^{-1} \exp\{-|x|\}$ . Dubultais eksponenciālais sadalījums ir simetrisks sadalījums, tā vidējā vērtība sakrīt ar mediānu un modu.

### Košī sadalījums.

Košī sadalījuma blīvuma funkcija ir  $f(x|\mu, \gamma) = \pi^{-1} [\gamma / ((x - \mu)^2 + \gamma^2)]$ , ja  $\mu = 0$  un  $\gamma = 1$ , tad to sauc par standarta Košī sadalījumu un  $f(x|0, 1) = (\pi(1 + x^2))^{-1}$ . Košī sadalījums ir ļoti interesants sadalījums, jo tam nav definēta vidējā vērtība, dispersija un augstākas kārtas momenti, taču tā moda un mediāna sakrīt ar  $\mu$ .



**Attēls 1:** Standartnormālā, standarta Košī un standarta dubultā eksponenciālā sadalījuma blīvuma funkcijas

Attēlā 1 ir redzamas standartnormālā  $N(0, 1)$ , standarta Košī (0, 1) un standarta dubultā eksponenciālā (0, 1) sadalījuma blīvuma funkcijas, lai arī visi trīs sadalījumi ir simetriski, taču tie ir ļoti dažādi.

### 3 Vislielākās ticamības attiecības tests

1922. gadā ģenētiķis un statistiķis R. A. Fišers ieviesa vislielākās ticamības metodi, kas bija viens no nozīmīgākajiem notikumiem matemātiskajā statistikā 20. gadsimtā, jo tā ir viena no visplašāk pielietojamajām metodēm parametriskajos modeļos. To pielieto lineārajos modeļos, faktoranalīzē, ekonometrikā, hipotēžu pārbaudēs, ticamības intervālu konstruēšanā u.c.

#### 3.1 Vislielākās ticamības funkcija un metode

Pieņemsim, ka  $X_1, \dots, X_n$  ir neatkarīgi, vienādi sadalīti gadījuma lielumi ar varbūtību blīvuma vai varbūtību masas funkciju  $f(x|\theta)$ , kur  $\theta \in \Theta \subset \mathbb{R}^k$ . Piemēram, ja  $X_1, \dots, X_n$  ir neatkarīgi, normāli sadalīti gadījuma lielumi, tad  $f(x|\theta) = f(x|\mu, \sigma^2) = (\sqrt{2\pi}\sigma)^{-1} \exp\{-(2\sigma^2)^{-1}(x - \mu)^2\}$ . Vislielākās ticamības funkcija tiek definēta kā

$$L_n(\theta) = \prod_{i=1}^n f(X_i|\theta).$$

Vislielākās ticamības metode novērtē parametru  $\theta$ , atrodot tādu  $\hat{\theta}$ , kas maksimizē  $L_n(\theta)$ .  $\hat{\theta}$  sauc par vislielākās ticamības novērtējumu parametram  $\theta$ , t.i.,

$$\hat{\theta} = \arg \max_{\theta} L_n(\theta).$$

Ja  $L_n(\theta)$  ir diferencējama pēc  $\theta$ , tad visbiežāk vislielākās ticamības funkcijas parametra  $\theta$  novērtējumu var iegūt ar vislielākās ticamības metodes palīdzību, atrisinot sekojošu vienādojumu:

$$\frac{\partial L_n(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}} = 0.$$

Iegūtā parametra  $\theta$  vērtība  $\hat{\theta}$  ir vislielākās ticamības funkcijas parametra novērtējums. Jāatzīmē, ka jebkura  $\theta$  vērtība, kas maksimizē  $L_n(\theta)$ , maksimizē arī  $\ln(L_n(\theta))$ , tādēļ bieži vislielākās ticamības funkcijas vietā izmanto logaritmisko vislielākās ticamības funkciju  $\ln(L_n(\theta))$ .

**Definīcija 4.** [1] Ja  $X_1, \dots, X_n$  ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar  $f(x|\theta)$ , tad  $I_n(\theta) = nI(\theta)$  sauc par Fišera informāciju, kur  $I(\theta) = -E_{\theta}(\partial^2 \ln f(X|\theta)/\partial \theta^2)$ .

**Regularitātes nosacījumi maksimālās ticamības metodei:**

1. Gadījuma lielumi  $X_i, i = 1, 2, \dots, n$  ir neatkarīgi un vienādi sadalīti ar  $f(x|\theta)$ ;
2. Parametru telpa  $\Theta$  ir kompakta un īstā parametra  $\theta$  vērtība  $\theta_0$  ir šīs telpas iekšējs punkts;
3. Eksistē  $E_{\theta_0} \ln f(X_i|\theta)$  un īstā parametra vērtība  $\theta_0$  ir  $\theta_0 = \arg \max_{\theta \in \Theta} E_{\theta_0} \ln f(X_i|\theta)$ ;

4. Logaritmiskā ticamības funkcija  $\ln(L_n(\theta)) = \sum_{i=1}^n \ln f(x_i|\theta)$  ir nepārtraukta pēc  $\theta$ ;
5.  $\exists \delta > 0$  tāds, ka  $\sup_{\theta \in \Theta} |n^{-1} \ln(L_n(\theta)) - E_{\theta_0} \ln f(X_i|\theta)| < \delta$  gandrīz droši;
6.  $\ln(L_n(\theta))$  ir divreiz nepārtraukti diferencējama īstās parametra vērtības  $\theta_0$  apkārtne;
7. Intergrēšanas un diferencēšanas operatorus var mainīt vietām;
8. Eksistē matrica  $I(\theta_0)$  un tā nav vienības matrica. [13]

Izpildoties regularitātes nosacījumiem, vislielākās ticamības metodes novērtējumi  $\hat{\theta}$  ir :

1. būtiski, t.i.,  $\forall \epsilon > 0 \lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| \geq \epsilon) = 0$ ;
2. funkcionāli invarianti, t.i., ja  $\hat{\theta}$  ir parametra  $\theta$  vislielākās ticamības metodes novērtējums, tad, ja  $g$  ir funkcija, kas atkarīga no parametra  $\theta$ , tad  $g(\hat{\theta})$  vislielākās ticamības metodes novērtējums ir  $g(\hat{\theta})$ ;
3. asimptotiski nenovirzīti, t.i., to novirze  $E(\hat{\theta} - \theta)$  tiecas uz nulli, ja  $n \rightarrow \infty$ ;
4. asimptotiski normāli, t.i.,  $(\hat{\theta} - \theta)/(\sqrt{1/I_n(\theta)}) \rightarrow N(0, 1)$ , ja  $n \rightarrow \infty$ ;
5. asimptotiski optimāli vai efektīvi, tas nozīmē, ka starp citiem parametra  $\theta$  novērtējumiem, vislielākās ticamības novērtējumiem ir mazākā dispersija. [1]

**Piemērs 1.** Pienemsim, ka  $X_1, \dots, X_n$  ir neatkarīgi un vienādi sadalīti un  $X_i \sim N(\mu, \sigma^2)$ . Tad normālā sadalījuma vislielākās ticamības funkcija ir

$$L_n(\theta) = L_n(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2\right)$$

un vislielākās ticamības metodes parametru novērtējumi ir

$$\hat{\mu} = \bar{X} \text{ un } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

$\hat{\mu}$  ir nenovirzīts parametra  $\mu$  novērtējums, jo  $E\hat{\mu} = E\bar{X} = E(n^{-1} \sum_{i=1}^n X_i) = \mu$ , bet  $\hat{\sigma}^2$  ir novirzīts parametra  $\sigma^2$  novērtējums, taču abi šie novērtējumi asimptotiski ir nenovirzīti. Normālais sadalījums ir simetrisks sadalījums, tam vidējā vērtība sakrīt ar mediānu.

Mediāna  $m$  tiek definēta kā  $P(X \leq m) \geq 1/2 \wedge P(X \geq m) \geq 1/2$  diskrētajā gadījumā un  $\int_{-\infty}^m dF(x) \geq 1/2 \wedge \int_m^{+\infty} dF(x) \geq 1/2$  nepārtrauktajā gadījumā.

Salīdzināsim divus  $\mu$  novērtējumus:  $\hat{\mu}$  un  $\tilde{\mu}$ , kur  $\hat{\mu}$  ir izlases vidējā vērtība (maksimālās ticamības metodes novērtējums), bet  $\tilde{\mu}$  ir izlases mediāna. Maksimālās ticamības metodes novērtējumam

ir spēkā, ka  $\sqrt{n}(\hat{\mu} - \mu) \rightarrow N(0, \sigma^2)$ . Var pierādīt [1], ka  $\sqrt{n}(\tilde{\mu} - \mu) \rightarrow N(0, \frac{\pi}{2}\sigma^2)$ , tas nozīmē, ka mediāna tāpat kā vidējā vērtība tiecas uz īsto vērtību, taču mediānai ir lielāka dispersija kā vidējai vērtībai. Tas arī uzskatāmi parāda maksimālās ticamības metodes parametru optimitāti.

### 3.2 Vislielākās ticamības attiecības tests

Vislielākās ticamības attiecības tests ir metode, kā pārbaudīt nulles hipotēzi  $H_0: \theta \in \Theta_0$  pret  $H_1: \theta \in \Theta_1$ , kur  $\Theta_0 \cup \Theta_1 = \Theta \subset R^k$  un  $\Theta_0 \cap \Theta_1 = \emptyset$ . Tests noraida  $H_0$  pie mazām vislielākās ticamības attiecības testa statistikas

$$\lambda_n = \frac{L_n(\hat{\theta}_0)}{L_n(\hat{\theta}_1)}$$

vērtībām, kur  $L_n(\hat{\theta}_0) = \sup_{\theta \in \Theta_0} L_n(\theta)$  un  $L_n(\hat{\theta}_1) = \sup_{\theta \in \Theta_0 \cup \Theta_1} L_n(\theta)$ . Vislielākās ticamības funkcija  $L_n(\theta)$  tiek maksimizēta divreiz: vispirms pār  $\Theta_0$  un pēc tam pār  $\Theta_0 \cup \Theta_1$ .

**Definīcija 5.** [1] Testa ar noraidīšanas reģionu  $R$ ,  $R = \{x : T(x) > c\}$ , jaudas funkcija tiek definēta ar  $\beta(\theta) = P_\theta(X \in R)$ , kur  $T$  ir testa statistika un  $c$  kritiskā vērtība.  $\alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$  ir testa lielums. Saka, ka testam ir nozīmības līmenis  $\alpha$ , ja tā lielums ir mazāks vai vienāds par  $\alpha$ .

**Lemma 1.** (Neimaņa-Pīrsona lemma, [1]) Aplūkosim vienkāršu hipotēžu pārbaudi  $H_0 : \theta = \theta_0$  pret  $H_1 : \theta = \theta_1$  un definēsim

$$T = \frac{L_n(\theta_0)}{L_n(\theta_1)} = \frac{\prod_{i=1}^n f(x_i | \theta_0)}{\prod_{i=1}^n f(x_i | \theta_1)}.$$

$H_0$  tiek noraidīta, ja  $T < k$  un ja  $k$  izvēlas tādu, ka  $P_{\theta_0}(T < k) = \alpha$ , tad šis ir jaudīgākais tests ar lielumu  $\alpha$ . Tas nozīmē, ka starp visiem testiem ar lielumu  $\alpha$ , šis tests maksimizē jaudu  $\beta(\theta_1)$ .

Tātad, pārbaudot vienkāršu hipotēzi  $H_0 : \theta = \theta_0$  pret  $H_1 : \theta = \theta_1$ , speciālgadījumā vislielākās ticamības attiecības tests ir jaudīgākais tests.

**Teorēma 2.** (Vilkса teorēma, [14]) Pieņemsim, ka  $X_1, \dots, X_n$  ir neatkarīgi un vienādi sadalīti ar zināmu  $f(x|\theta)$ , kur parametrs  $\theta \subset \Theta \subset R^k$  nav zināms, un pieņemsim, ka  $f(x|\theta)$  apmierina regularitātes nosacījumus (sk. 10. lpp.) Tad, pārbaudot  $H_0 : \theta = \theta_0 \subset \Theta \subset R^r$ , kur  $0 < r < k$ ,

$$-2 \ln \lambda_n \sim \chi_r^2.$$

Vilkса teorēma nosaka, ka  $-2 \ln \lambda_n \sim \chi_r^2$ . Izmantojot šo rezultātu var konstruēt ticamības intervālus parametram  $\theta$  sekojošā veidā. No kritisko vērtību tabulas var atrast  $c$  tādu, ka

$$P(\chi_r^2 \leq c) = 1 - \alpha, \tag{3.1}$$

kur  $1 - \alpha$  ir reģiona

$$\mathcal{R}_c = \left\{ \theta : -2 \ln(L_n(\theta)/L_n(\hat{\theta}_1)) \leq c \right\} \quad (3.2)$$

ticamība. No Vilkса teorēmas seko, ka  $\mathcal{R}_c$  konvergēce ir vienāda ar  $1 - \alpha$ :

$$P(\theta_0 \in \mathcal{R}_c) = P(-2 \ln(L_n(\theta_0)/L_n(\hat{\theta}_1)) \leq c) \rightarrow 1 - \alpha, \quad (3.3)$$

kad  $n \rightarrow \infty$ .

**Piemērs 2.** Pieņemsim, ka  $X_1, \dots, X_n$  ir neatkarīgi un normāli sadalīti ( $N(\mu, \sigma^2)$ ), un veiksim sekojošu hipotēžu pārbaudi:  $H_0 : \mu = \mu_0$  pret  $H_1 : \mu \neq \mu_0$ . Atradīsim vislielākās ticamības attiecības testa statistiku divos gadījumos, kad  $\sigma^2$  ir zināma un kad  $\sigma^2$  nav zināma.

1.  $\sigma^2$  ir zināma.

$$\begin{aligned} \lambda_n &= \frac{[-\sqrt{2\pi}\sigma]^{-n} \exp\left(-(2\sigma^2)^{-1} \sum_{i=1}^n (X_i - \mu_0)^2\right)}{[-\sqrt{2\pi}\sigma]^{-n} \exp\left(-(2\sigma^2)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2\right)} = \frac{\exp\left(-(2\sigma^2)^{-1} \sum_{i=1}^n (X_i - \mu_0)^2\right)}{\exp\left(-(2\sigma^2)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2\right)} \\ &= \exp\left(-(2\sigma^2)^{-1} \sum_{i=1}^n (X_i - \mu_0)^2 + (2\sigma^2)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) \\ &= \exp\left(-(2\sigma^2)^{-1} \sum_{i=1}^n [(X_i - \mu_0)^2 - (X_i - \bar{X})^2]\right) \\ &= \exp\left(-(2\sigma^2)^{-1} \sum_{i=1}^n [(X_i - \bar{X} + \bar{X} - \mu_0)^2 - (X_i - \bar{X})^2]\right) \\ &= \exp\left(-(2\sigma^2)^{-1} \left[ 2 \sum_{i=1}^n [(X_i - \bar{X})(\bar{X} - \mu_0)] + n(\bar{X} - \mu_0)^2 \right] \right) \\ &= \exp(-(2\sigma^2)^{-1} n(\bar{X} - \mu_0)^2), \end{aligned}$$

jo

$$\begin{aligned} \sum_{i=1}^n [(X_i - \bar{X})(\bar{X} - \mu_0)] &= (\bar{X} - \mu_0) \sum_{i=1}^n (X_i - \bar{X}) = (\bar{X} - \mu_0) \left( \sum_{i=1}^n X_i - n\bar{X} \right) \\ &= (\bar{X} - \mu_0) \left( \sum_{i=1}^n X_i - n \frac{1}{n} \sum_{i=1}^n X_i \right) = 0. \end{aligned}$$

Līdz ar to

$$-2 \ln(\lambda_n) = \frac{n}{\sigma^2} (\bar{X} - \mu_0)^2 \sim \chi_1^2.$$

Ievērojam, ka  $-2 \ln(\lambda_n) = Z^2$ , kur  $Z = \sqrt{n}(\bar{X} - \mu_0)/\sigma \sim N(0, 1)$  ir  $Z$ -testa statistika.

2.  $\sigma^2$  nav zināma.

**Lemma 3.** [15] Ja  $a$  ir konstante vai mainīgais, kas nav atkarīgs no  $\sigma^2$  un  $n$  ir pozitīva konstante, tad funkcija  $h(\sigma^2; a) = (\sigma^2)^{-(n/2)} e^{-(a/2\sigma^2)}$  savu maksimumu sasniedz pie  $\hat{\sigma}^2 = a/n$  un  $\max_{\sigma^2} h(a/n; a) = (a/n)^{-(n/2)} e^{-(n/2)}$ .

□ Atvasinām  $\ln(h) = -a(2\sigma^2)^{-1} - (n \ln(\sigma^2))/2$  pēc  $\sigma^2$  un pielīdzinām 0:

$$\frac{\partial}{\partial \sigma^2} \ln(h) = \frac{1}{2\sigma^4} a - \frac{n}{2\sigma^2}, \quad \frac{\partial}{\partial \sigma^2} \ln(h) = 0 \implies \frac{a}{\sigma^2} - n = 0 \implies \sigma^2 = \frac{a}{n}.$$

Iegūtais atrisinājums ir maksimums, jo  $\ln(h)$  otrs atvasinājums punktā  $\sigma^2 = a/n$  ir negatīvs, proti,

$$\begin{aligned} \frac{\partial^2}{(\partial \sigma^2)^2} \ln(h) &= \frac{\partial^2}{\partial \sigma^2} \left( \frac{a}{2\sigma^4} - \frac{n}{2\sigma^2} \right) = -\frac{a}{\sigma^6} + \frac{n}{2\sigma^4}, \\ \frac{\partial^2}{(\partial \sigma^2)^2} \ln(h) \Big|_{\sigma^2=a/n} &= -\frac{a}{(a/n)^3} + \frac{n}{2(a/n)^2} = -\frac{n^3}{a^2} < 0. \end{aligned}$$

Lai iegūtu maksimālo izteiksmes  $h(\sigma^2; a)$  vērtību  $\sigma^2$  jāaizvieto ar  $a/n$ . ■

$L_n(\mu, \sigma^2)$  pie  $H_0$  ir vienāda ar

$$L_n(\mu_0, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu_0)^2 \right).$$

Izmantojot Lemmu 3,  $a$  vietā ņemot  $a = \sum_{i=1}^n (X_i - \mu_0)^2$ , iegūstam, ka  $\hat{\sigma}_0^2 = n^{-1} \sum_{i=1}^n (X_i - \mu_0)^2$  un

$$\max_{\sigma^2} L_n(\mu_0, \sigma^2) = L_n(\mu_0, \hat{\sigma}_0^2) = e^{-n/2} \left( n^{-1} \sum_{i=1}^n (X_i - \mu_0)^2 \right)^{-n/2} (2\pi)^{-n/2}.$$

Pie  $H_0 \cup H_1$  vislielākās ticamības funkciju maksimizē tās vislielākās ticamības metodes parametru novērtējumi  $\hat{\mu}$  un  $\hat{\sigma}^2$ , tātad

$$\max_{\sigma^2, \mu} L_n(\mu, \sigma^2) = L_n(\bar{X}, \hat{\sigma}^2) = e^{-n/2} \left( n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2 \right)^{-n/2} (2\pi)^{-n/2}.$$

Vislielākās ticamības attiecības testa statistikas vērtība ir

$$\begin{aligned}
\lambda_n &= L_n(\mu_0, \hat{\sigma}_0^2) / L_n(\bar{X}, \hat{\sigma}^2) = \left( \sum_{i=1}^n (X_i - \bar{X})^2 / \sum_{i=1}^n (X_i - \mu_0)^2 \right)^{n/2} \\
&= \left( \sum_{i=1}^n (X_i - \bar{X})^2 / \sum_{i=1}^n (X_i - \bar{X} + \bar{X} - \mu_0)^2 \right)^{n/2} \\
&= \left( \sum_{i=1}^n (X_i - \bar{X})^2 / \sum_{i=1}^n [(X_i - \bar{X})^2 + 2(X_i - \bar{X})(\bar{X} - \mu_0) + (\bar{X} - \mu_0)^2] \right)^{n/2} \\
&= \left( \sum_{i=1}^n (X_i - \bar{X})^2 / \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu_0)^2 \right)^{n/2}.
\end{aligned}$$

Tātad

$$-2 \ln(\lambda_n) = n \ln \left( 1 + (n(\bar{X} - \mu_0)^2) / \sum_{i=1}^n (X_i - \bar{X})^2 \right) \sim \chi_1^2.$$

Ievērosim, ka šajā gadījumā

$$\left( \lambda_n^{-2/n} - 1 \right) (n - 1) = \frac{n(\bar{X} - \mu_0)^2}{S^2} = T^2,$$

kur  $T$  ir Stjūdenta  $t$ -statistika.

## 4 Neparametriskās ticamības funkcijas attiecības tests

### 4.1 Neparametriskā vislielākās ticamības funkcija

**Definīcija 6.** [2] Ja  $X_1, \dots, X_n$  ir neatkarīgi un vienādi sadalīti, tad empīriskā kumulatīvā sadalījuma funkcija (turpmāk – empīriskā sadalījuma funkcija) tiek definēta kā

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}}, \quad (4.1)$$

kur  $-\infty < x < \infty$ .

**Definīcija 7.** [2] Ja  $X_1, \dots, X_n$  ir neatkarīgi un vienādi sadalīti ar kādu sadalījuma funkciju  $F_0$ , tad funkcijas  $F$  neparametriskā ticamības funkcija ir

$$L(F) = \prod_{i=1}^n (F(X_i) - F(X_i-)). \quad (4.2)$$

Būtiska neparametriskās ticamības funkcijas īpašība ir, ka  $L(F) = 0$ , ja  $F$  ir nepārtraukts sadalījums. Lai neparametriskā ticamības funkcija būtu pozitīva, sadalījuma funkcijai  $F$  jāuzliek pozitīva varbūtību katram no pētāmās datu kopas punktiem.

**Teorēma 4.** (Owen, [2]) Pieņemsim, ka  $X_1, \dots, X_n \in \mathbb{R}$  ir neatkarīgi gadījuma lielumi ar sadalījuma funkciju  $F_0$  un  $F_n$  ir to empīriskā sadalījuma funkcija. Ja  $F \neq F_n$ , tad  $L(F) < L(F_n)$ .

□ Pieņemsim, ka  $z_1 < z_2 < \dots < z_m$  ir atšķirīgas  $X_1, \dots, X_n$  vērtības un  $n_j \geq 1$  ir  $X_i$  skaits, kad  $X_i$  sakrīt ar  $z_j$ ,  $p_j = F(z_j) - F(z_j-) = P(z_j = X_i)$  un  $\hat{p}_j = n_j/n$ .

Ja  $p_j = 0 \forall j = 1, \dots, m$ , tad  $L(F) = 0 < L(F_n)$ . Tāpēc pieņemsim, ka visi  $p_j > 0$  un ka vismaz vienam  $j$ :  $p_j \neq \hat{p}_j$ . Tā kā  $\log(x) \leq x - 1 \forall x > 0$  ar vienādību tikai, kad  $x = 1$ , tad

$$\log \left( \frac{L(F)}{L(F_n)} \right) = \sum_{j=1}^m n_j \log \left( \frac{p_j}{\hat{p}_j} \right) = n \sum_{j=1}^m \hat{p}_j \log \left( \frac{p_j}{\hat{p}_j} \right) \leq n \sum_{j=1}^m \hat{p}_j \left( \frac{p_j}{\hat{p}_j} - 1 \right) < 0.$$

Tātad  $L(F) < L(F_n)$ . ■

Teorēma 4 pierāda to, ka neparametisko ticamības funkciju maksimizē empīriskā sadalījuma funkcija. Tātad empīriskā sadalījuma funkcija  $F_n$  ir funkcijas  $F$  neparametriskās vislielākās ticamības funkcijas novērtējums.

## 4.2 Neparametriskās ticamības funkcijas attiecības tests

Līdzīgi kā parametriskajā gadījumā, arī neparametriskās ticamības funkcijas attiecību var izmantot hipotēžu pārbaudei un ticamības intervālu konstruēšanai. Sadalījuma funkcijai  $F$  definēsim neparametriskās ticamības funkcijas attiecību:

$$R(F) = \frac{L(F)}{L(F_n)} = \prod_{i=1}^n np_i, \quad (4.3)$$

kur  $L(F)$  ir definēts ar (4.2). Lai neparametriskajā gadījumā novērtētu parametrus un veiktu hipotēžu pārbaudi, nezināmo parametru  $\theta$  izsaka funkcionāli no sadalījuma funkcijas  $F$ , t.i.,  $\theta = T(F)$  un parametra  $\theta$  novērtējums ir  $\hat{\theta} = T(F_n)$ .

**Piemērs 3.** Ja  $X_1, \dots, X_n$  ir neatkarīgi un vienādi sadalīti

1. vidējā vērtība:  $\theta = EX = \int_{-\infty}^{+\infty} xdF(x)$  un  $\hat{\theta}$  iegūstam sekojoši  $\hat{\theta} = \int_{-\infty}^{+\infty} xdF_n(x) = n^{-1} \sum X_i = \bar{X}$ , tas nozīmē, ka neparametriskais vidējās vērtības novērtējums vienmēr būs izlases vidējā vērtība, kas sakrīt ar paramerisko gadījumu, kad izlase ir normāli sadalīta;
2. mediāna  $\theta = m : \int_{-\infty}^m dF(x) - \int_m^{+\infty} dF(x) = 0$ .

Uzskatīsim, ka  $F$  pieder kopai  $\mathcal{F}$ ,  $\mathcal{F}$  var sakrist ar visu  $\mathbb{R}$ , taču biežāk tiek izmantota  $\mathbb{R}$  apakškopa. Definēsim tā saukto profila ticamības attiecības funkciju:

$$R(\theta) = \sup \{R(F)|T(F) = \theta, F \in \mathcal{F}\}.$$

Empīriskās ticamības (EL) hipotēžu pārbaude noraida  $H_0 : T(F_0) = \theta_0$ , kad  $R(\theta_0) < r_0$ , kur  $r_0$  ir konstante. EL ticamības reģioni ir sekojošā formā:  $\{\theta|R(\theta) \geq r_0\}$ ,  $r_0$  var noteikt izmantojot Vilksa teorēmas neparametrisko versiju (sk. Teorēmu 8).

Viendimensionāla parametra gadījumā zīmes-saknes empīriskās logaritmiskās ticamības attiecības statistikai  $R_0 = \text{sign}(\hat{\theta} - \theta_0) \sqrt{-2 \ln R(\theta_0)}$  ir asimptotisks standartnormālais sadalījums  $N(0, 1)$ , bet, ja  $\theta$  ir  $p$ -dimensionāls parametrs, tad  $R'_0 R_0 = -2/n \ln R(\theta_0)$  un vektoram  $n^{1/2} R_0$  ir  $N(0, I_p)$  sadalījums ar kārtu  $O(n^{-1/2})$ . [2]

## 4.3 Ticamības intevālu konstruēšana vidējai vērtībai ar EL

Kā jau iepriekš tika minēts par parametru  $\theta$ , mēs interesēsimies kā par funkcionāli no sadalījuma funkcijas  $F$ , t.i.,  $\theta = T(F)$ . Vispirms, lai labāk izprastu problemātiku, apskatīsim vienkāršāko gadījumu, kad  $T(F) = \mu$ .

Pieņemsim, ka mums ir doti neatkarīgi un vienādi sadalīti gadījuma lielumi  $X_1, \dots, X_n$  ar

nezināmu sadalījuma funkciju  $F$ . Ja empīriskā ticamības funkcija ir definēta ar (4.2) un empīriskās ticamības attiecība ar (4.3), tad, lai konstruētu ticamības reģionus parametram  $\mu$ , izmantosim tā saukto profila empīriskās ticamības attiecības funkciju

$$R_E(\mu) = \sup \left\{ \prod_{i=1}^n np_i \middle| p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i X_i = \mu \right\}. \quad (4.4)$$

Owen ([10],[11]) parādīja, ka eksistē viens vienīgs izteiksmes (4.4) labās putas atrisinājums, ja  $\mu$  atrodas izliektajā čaulā, ko veido  $X_1, \dots, X_n$ . Maksimizācijas problēmu (4.4) var atrisināt ar Lagranža reizinātāju palīdzību. Izteiksme  $\prod_{i=1}^n np_i$ , pie ierobežojumiem

$$p_i \geq 0, \sum_{i=1}^n p_i = 1 \text{ un } \sum_{i=1}^n p_i X_i = \mu$$

savu maksimumu sasniedz, kad

$$p_i = p_i(\mu) = n^{-1} \{1 + \lambda(X_i - \mu)\}^{-1}, \quad (4.5)$$

kur  $\lambda = \lambda(\mu)$  var iegūt no izteiksmes

$$\sum_{i=1}^n \{1 + \lambda(X_i - \mu)\}^{-1} (X_i - \mu) = 0. \quad (4.6)$$

Tātad

$$R_E(\mu) = \prod_{i=1}^n \{1 + \lambda(X_i - \mu)\}^{-1}. \quad (4.7)$$

Logaritmiskā empīriskās ticamības attiecības statistika ir  $W_E(\mu) = -2 \ln R_E(\mu)$ , t.i,

$$W_E(\mu) = 2 \sum_{i=1}^n \ln \{1 + \lambda(X_i - \mu)\}. \quad (4.8)$$

**Teorēma 5.** (Owen, [2]) Ja  $X_1, \dots, X_n$  ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar sadalījuma funkciju  $F_0$  un  $\mu_0 = E(X_i)$ , un  $0 < D(X_i) < \infty$ , tad  $-2 \ln(R(\mu_0))$  tiecas uz  $\chi_1^2$ , kad  $n \rightarrow \infty$ .

Teorēma 5 parāda, ka, ja  $\mu = \mu_0$ , tad  $W_E(\mu)$  tiecas uz  $\chi_1^2$ , kad  $n \rightarrow \infty$ . Ticamības intervālus parametram  $\mu$  pie nozīmības līmeņa  $\alpha$  var iegūt kā tādu punktu kopu, kam  $W_E(\mu) \leq c$ , kur  $c$  ir noteikts kā  $P(\chi_1^2 \leq c) = 1 - \alpha$ .

## 4.4 EL vispārējā gadījumā

EL vispārējais gadījums izklāstīts [3]. Pieņemsim, ka mums ir doti neatkarīgi un vienādi sadalīti  $d$ -dimensionāli gadījuma lielumi  $X_1, \dots, X_n$  ar nezināmu sadalījuma funkciju  $F$ , kurai ir  $p$ -dimensionāls parametrs  $\theta$ . Pieņemsim, ka informācija par  $F$  un  $\theta$  ir dota  $r \geq p$  funkcionāli

neatkarīgu nenovirzītu funkciju formā, t.i., kā funkcijas  $g_j(X, \theta)$ ,  $j = 1, 2, \dots, r$ , tādas, ka  $E_F \{g_j(X, \theta)\} = 0$ .

Šajā gadījumā funkciju  $L(F) = \prod_{i=1}^n p_i$  maksimizēsim pie sekojošiem ierobežojumiem

$$p_i \geq 0, \quad \sum_i p_i = 1, \quad \sum_i p_i g(X_i, \theta) = 0. \quad (4.9)$$

Dotam  $\theta$  eksistē viens vienīgs maksimums, ja 0 ir izliektajā čaulā, kura sastāv no punktiem  $g(X_1, \theta), \dots, g(X_n, \theta)$ . Tāpat kā iepriekš maksimumu var atrast ar Lagranža reizinātāju palīdzību.

$$H = \sum_i \ln(p_i) + \lambda_0 \left( 1 - \sum_i p_i \right) - n\lambda^\tau \sum_i p_i g(X_i, \theta), \quad (4.10)$$

kur  $\lambda_0$  un  $\lambda = (\lambda_1, \dots, \lambda_r)^\tau$  ir Lagranža reizinātāji. Atvasinot izteiksmi (4.10) pēc  $p_i$ , iegūstam

$$\frac{\partial H}{\partial p_i} = \frac{1}{p_i} - \lambda_0 - n\lambda^\tau g(X_i, \theta) = 0, \quad \sum_i p_i \frac{\partial H}{\partial p_i} = n - \lambda_0 = 0 \Rightarrow \lambda_0 = n$$

un

$$p_i = \left( \frac{1}{n} \right) \frac{1}{1 + \lambda^\tau g(X_i, \theta)}. \quad (4.11)$$

Nemot vērā trešo ierobežojumu (4.9), iegūstam, ka

$$0 = \sum_i p_i g(X_i, \theta) = \frac{1}{n} \sum_i \frac{1}{1 + \lambda^\tau g(X_i, \theta)} g(X_i, \theta). \quad (4.12)$$

Lagranža reizinātājs  $\lambda$  var tikt noteikts kā funkcija no  $\theta$ . Tā kā  $0 \leq p_i \leq 1$ , tad  $\lambda$  un  $\theta$  jāizpildās  $1 + \lambda^\tau g(X_i, \theta) \geq 1/n$  katram  $i$ . Fiksētam  $\theta$ ,  $D_\theta = \{\lambda : 1 + \lambda^\tau g(X_i, \theta) \geq 1/n\}$ .  $D_\theta$  ir izliekts, slēgts un ierobežots, ja 0 ir izliektajā čaulā, ko veido  $g(X_i, \theta)$ . Bez tam

$$\frac{\partial}{\partial \lambda} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \lambda^\tau g(X_i, \theta)} g(X_i, \theta) \right\} = -\frac{1}{n} \sum_{i=1}^n \frac{g(X_i, \theta) g^\tau(X_i, \theta)}{(1 + \lambda^\tau g(X_i, \theta))^2}$$

ir negatīvi definita argumentam  $\lambda$ , nodrošinot to, ka  $\sum_{i=1}^n g(X_i, \theta) g^\tau(X_i, \theta)$  ir pozitīvi definita.  $\lambda = \lambda(\theta)$  ir nepārtraukti diferencējama funkcija pēc  $\theta$ . Profila empīriskā ticamības funkcija parametram  $\theta$  tagad tiek definēta kā

$$L_E(\theta) = \prod_{i=1}^n \left\{ \frac{1}{n} \frac{1}{1 + \lambda^\tau(\theta) g(X_i, \theta)} \right\}.$$

Logaritmiskā empīriskās ticamības funkcijas attiecība tiek definēta sekojoši:

$$l_E(\theta) = \sum_{i=1}^n \ln [1 + \lambda^\tau(\theta) g(X_i, \theta)]. \quad (4.13)$$

Acīmredzami, gadījumā, ja  $\theta = \mu$ , tad (4.5) un (4.6) ir (4.11) un (4.12) speciālgadījumi, kad  $g(X_i, \mu) = X_i - \mu$ , un  $2^{-1}W_E(\mu)$  no (4.8) ir (4.13) speciālgadījums.

Minimizējot  $l_E(\theta)$ , var iegūt parametra  $\theta$  novērtējumu  $\tilde{\theta}$ , t.i.,  $\tilde{\theta} = \arg \min_{\theta} l_E(\theta)$ , kur  $\tilde{\theta}$  ir empīriskās vislielākās ticamības novērtējums. Turklat ar  $\tilde{\theta}$  palīdzību var noteikt  $p_i$  novērtējumus  $\tilde{p}_i$  no (4.11) un sadalījuma funkcijas  $F$  novērtējumu

$$\tilde{F}_n(x) = \sum_{i=1}^n \tilde{p}_i 1(X_i \leq x). \quad (4.14)$$

Ja  $r = p$ , tad  $\tilde{\theta} = \hat{\theta}$  minimizē  $l_E(\theta)$ , kur  $\hat{\theta}$  ir vienādojumu  $\sum_{i=1}^n g(X_i, \theta) = 0$  atrisinājums. Turklat  $\tilde{p}_i = n^{-1}$  un (4.14) ir empīriskā sadalījuma funkcija. Gadījums, kad  $r > p$  sīkāk ar ilustratīvu piemēru palīdzību aprakstīts [3].

**Lemma 6.** (Qin, Lawless, [3]) Pienemsim, ka  $E[g(X, \theta_0)g^\tau(X, \theta_0)]$  ir pozitīva, atvasinājums  $\partial g(X, \theta)/\partial \theta$  ir nepārtraukts īstās vērtības  $\theta_0$  apkārtnē,  $\|\partial g(X, \theta_0)/\partial \theta\|$  un  $\|g(X, \theta)\|^3$  šajā apkārtnē ierobežo kāda integrējama funkcija  $G(X)$  un  $E[\partial g(X, \theta_0)/\partial \theta]$  rangs ir  $p$ . Tad, ja  $n \rightarrow \infty$ , ar varbūtību 1  $l_E(\theta)$  sasniedz savu minimālo vērtību punktā  $\tilde{\theta}$  lodes  $\|\theta - \theta_0\| \leq n^{-1/3}$  iekšienē, un  $\tilde{\theta}$  un  $\tilde{\lambda} = \lambda(\tilde{\theta})$  apmierina

$$Q_{1n}(\tilde{\theta}, \tilde{\lambda}) = 0, \quad Q_{2n}(\tilde{\theta}, \tilde{\lambda}) = 0, \quad (4.15)$$

kur

$$Q_{1n}(\theta, \lambda) = \frac{1}{n} \sum_i \frac{1}{1 + \lambda^\tau g(X_i, \theta)} g(X_i, \theta), \quad (4.16)$$

$$Q_{2n}(\theta, \lambda) = \frac{1}{n} \sum_i \frac{1}{1 + \lambda^\tau g(X_i, \theta)} \left( \frac{\partial g(X_i, \theta)}{\partial \theta} \right)^\tau \lambda. \quad (4.17)$$

**Teorēma 7.** (Qin, Lawless, [3]) Izpildoties Lemmas 6 nosacījumiem un pieņemot, ka  $\partial^2 g(X, \theta)/\partial \theta \partial \theta^\tau$  ir nepārtraukts pēc  $\theta$  īstās vērtības  $\theta_0$  apkārtnē, un ja  $\|\partial^2 g(X, \theta)/\partial \theta \partial \theta^\tau\|$  šajā apkārtnē var ierobežot ar kādu integrējamu funkciju  $G(X)$ , tad

$$\sqrt{n}(\tilde{\theta} - \theta_0) \rightarrow N(0, V), \quad \sqrt{n}(\tilde{\lambda} - 0) \rightarrow N(0, U), \quad \sqrt{n}(\tilde{F}_n(x) - F(x)) \rightarrow N(0, W(x)),$$

kur

$$\tilde{F}_n(x) = \sum_{i=1}^n \tilde{p}_i 1(X_i < x), \quad \tilde{p}_i = \left( \frac{1}{n} \right) \frac{1}{1 + \tilde{\lambda}^\tau g(X_i, \tilde{\theta})}, \quad V = \left[ E \left( \frac{\partial g}{\partial \theta} \right)^\tau (Egg^\tau)^{-1} E \left( \frac{\partial g}{\partial \theta} \right) \right]^{-1},$$

$$W(x) = F(x)(1 - F(x)) - B(x)UB^\tau(x), \quad B(x) = E \{ g(X_i, \theta_0)1(X_i < x) \},$$

$$U = [E(gg^\tau)]^{-1} \left\{ I - E \left( \frac{\partial g}{\partial \theta} \right) V E \left( \frac{\partial g}{\partial \theta} \right)^\tau [E(gg^\tau)]^{-1} \right\}$$

un  $\tilde{\theta}$  un  $\tilde{\lambda}$  ir asimptotiski nekorelēti.

**Teorēma 8.** (Qin, Lawless, [3]) Empīriskās ticamības attiecības statistika hipotēzei  $H_0 : \theta = \theta_0$ , ir

$$W_E(\theta_0) = 2l_E(\theta_0) - 2l_E(\tilde{\theta}), \quad (4.18)$$

kur  $l_E(\theta)$  ir uzdots ar (4.13). Izpildoties Lemmas 6 un Teorēmas 7 nosacījumiem,  $W_E(\theta_0) \rightarrow \chi_p^2$ , kad  $n \rightarrow \infty$  un  $H_0$  ir spēkā.

Lemmas 6, Teorēmu 7 un 8 rezultāti dod iespēju novērtēt parametrus un konstruēt ticamības intervālus. Tāpat kā parametriskajā gadījumā, arī neparametriskajā gadījumā logaritmiskajai attiecības testa statistikai ir  $\chi^2$  sadalījums un empīriskās vislielākās ticamības novērtējums  $\tilde{\theta}$  ir asimptotiski normāls.

## 5 Vislielākās ticamības attiecības testu salīdzinājums vidējai vērtībai

Lai noskaidrotu, kā darbojas empīriskās ticamības tests, kas nebalstās uz pieņēmumiem par datu sadalījumu, salīdzinājumā ar parametriskajiem testiem, konstruēsim ticamības intervālus vidējai vērtībai  $\mu$ , izmantojot dažādas metodes, un analizēsim iegūtos rezultātus.

Lai konstruētu ticamības intervālu vidējai vērtībai ar EL palīdzību, ir nepieciešams noteikt parametru  $\lambda$  no izteiksmes (4.6), tā kā šī izteiksme ir polinoms un tai var būt vairākas saknes, tad jānoskaidro intervāls, kurā meklēt  $\lambda$  un, izmantojot zināmos nosacījumus, ka  $\forall i p_i > 0$  un  $p_i < 1$ , kur  $p_i$  definēts ar (4.5), iegūstam, ka

$$\frac{1 - n^{-1}}{\mu - X_{(n)}} < \lambda < \frac{1 - n^{-1}}{\mu - X_{(1)}}, \quad (5.1)$$

kur  $X_{(1)}, \dots, X_{(n)}$  ir augošā secībā sakārtota sākotnējā izlase. Apgabals, kurā meklēt vienādojuma sakni  $\lambda$ , ir noteikts un tālāk ar Nūtona, Brenta vai kāda cita algoritma palīdzību jānosaka  $\lambda$  vērtība.

### 5.1 Nūtona algoritms

Ja funkcija  $f(w)$  ir diferencējama un  $w_1$  ir vienādojuma  $f(w) = 0$  saknes  $w_0$  tuvinātā vērtība, tad par saknes nākošo tuvinājumu var ņemt skaitli

$$w_2 = w_1 - \frac{f(w_1)}{f'(w_1)}, \quad f'(w_1) \neq 0.$$

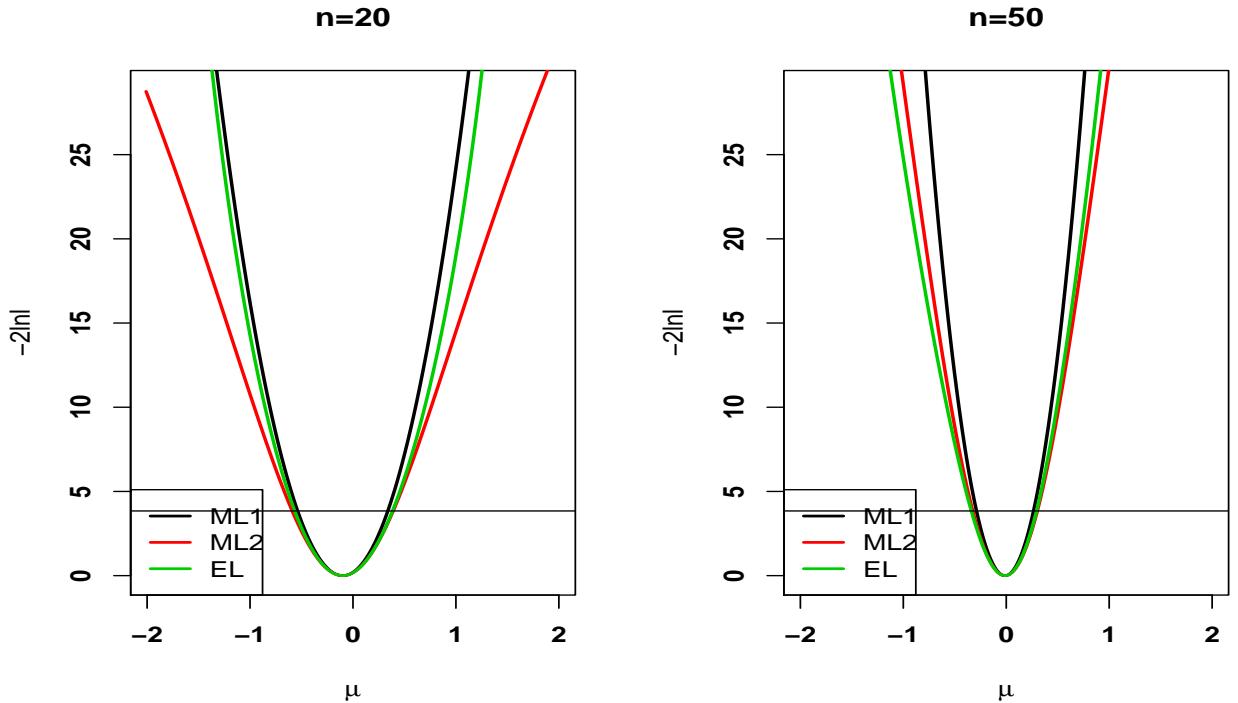
Vispārīgajā gadījumā  $w_{n+1} = w_n - f(w_n)/f'(w_n)$ . Metodi var lietot, ja

$$\left| \frac{f(w)f''(w)}{(f'(w))^2} \right| \leq m < 1. [20]$$

### 5.2 Parametriskās un neparametriskās metodes salīdzinājums

Parametriskās vislielākās ticamības novērtējumiem jābūt labākiem, ja īstais sadalījums ir tuvs parametriskajam sadalījumam, taču ja pētāmās datu kopas sadalījums nav tuvs parametriskajam sadalījumam, tad neparametriskās vislielākās ticamības novērtējumiem būtu jābūt labākiem vismaz pie pietiekami lieliem  $n$ . Lai pārbaudītu augstāk minēto faktu praksē, tika uzgenerēti divu dažādu sadalījumu dati un salīdzināti savā starpā parametriskās un neparametriskās vislielākās ticamības novērtējumi parametram  $\mu$ . ML1 un ML2 ir maksimālās ticamības metodes,

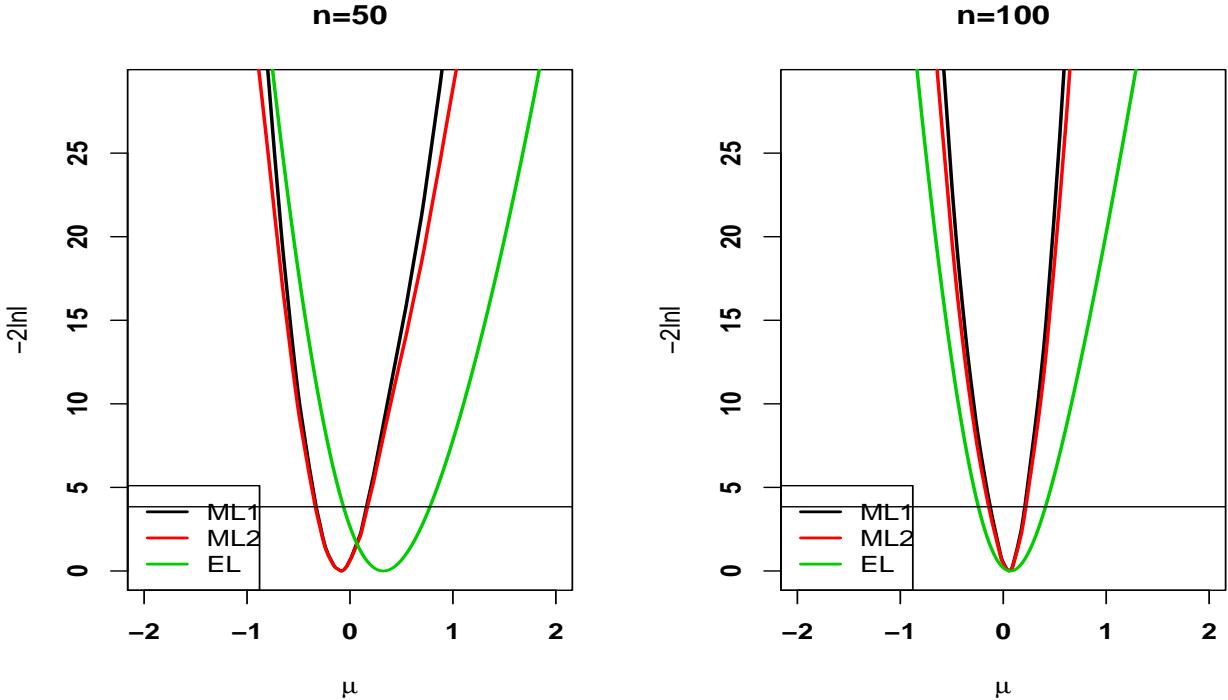
kur pirmajā gadījumā izkliedes rādītāja (dispersijas) īstā vērtība ir zināma, bet otrajā – nav zināma, EL ir empīriskās ticamības metode.



**Attēls 2:** Parametriskās (ML1, ML2) un neparametriskā (EL)  $-2 \times \ln$  ticamības liknes standartnormālā sadalījuma parametram  $\mu$ ,  $\alpha = 0.05$ .

Kā redzams Attēlā 2, tad standartnormālajam sadalījumam  $N(0, 1)$  gan pie izlases apjoma  $n = 20$ , gan pie  $n = 50$   $\mu$  novērtējumi ir ļoti tuvi  $\mu$  īstajai vērtībai  $\mu_0 = 0$ . Abos gadījumos visas trīs metodes, pie izvēlētās precizitātes, vienādi novērtē parametru  $\mu$ , pirmajā gadījumā  $\hat{\mu} = -0.1$ , bet otrajā –  $\hat{\mu} = -0.01$ . Konkrētajā gadījumā neparametriskās ticamības metodes  $\mu$  novērtējums ir vienāds ar parametriskās ticamības metodes novērtējumu.

Normālajam sadalījumam pie nelieliem izlašu apjomiem EL strādā labi un, palielinoties izlases apjomam, ticamības intervāls parametram  $\mu$  tikai sašaurinās. Taču visos gadījumos parametriskā un empīriskā ticamības metode nedarbosies tik līdzīgi.



**Attēls 3:** Parametriskās (ML1, ML2) un neparametriskā (EL)  $-2 \times \ln$  ticamības līknēs standarta dubultā eksponenciālā sadalījuma parametram  $\mu$ ,  $\alpha = 0.05$ .

Aplūkojot Attēlu 3, redzams, ka pie  $n = 50$  parametriskās un neparametriskās metodes parametra  $\mu$  novērtējumi diezgan atšķiras. Lai arī dubultais eksponenciālais, tāpat kā normālais, ir simetriks sadalījums, taču  $\mu$  maksimālās ticamības metodes novērtējums dubultajam eksponenciālajam sadalījumam ir nevis izlases vidējā vērtība, kā normālajam sadalījumam, bet gan izlases mediāna un tas arī ir galvenais iemesls, kāpēc ir diezgan ievērojamas atšķirības starp empirisko ticamības metodi, kas balstīta uz vidējo vērtību, un parametrisko ticamības metodi. Pie  $n = 50$  ML1, ML2 un EL parametra  $\mu$  novērtējumi ir attiecīgi -0.087, -0.089 un 0.324, bet pie  $n = 100$  atšķirība starp parametrisko un neparametrisko gadījumu ir nenozīmīga, un  $\mu$  novērtējumi attiecīgi ir 0.056, 0.056 un 0.07, taču empiriskajā gadījumā ticamības intervāla garums parametram  $\mu$  ir daudz lielāks. Plašaks empiriskās un parametriskās ticamības salīdzinājums ir pieejams [6], kur ir aprakstīti nosacījumi, kuriem jāizpildās parametru saimei, lai tās empiriskā un parametriskā ticamība sakristu zināmā līmeni.

Apskatītie piemēri vēl nevar kalpot par pamatu nopietniem secinājumiem par to, cik efektīvi ir neparametrisko metožu novērtējumi salīdzinājumā ar parametriskajiem. Lai noskaidrotu, cik labi strādā empiriskās ticamības metode salīdzinājumā ar parametriskajām metodēm, tiks iz-

mantota pārklājuma precizitāte. 10000 reizes tika ġenerēti dažādu sadalījumu dati pie atšķirīgiem izlašu apjomiem un ar dažādām metodēm tika konstruēti ticamības intervāli parametram  $\mu$  pie nozīmības līmeņa  $\alpha = 5\%$  un pārbaudīta to pārklājuma precizitāte, t.i., cik bieži īstā parametra vērtība iekrīt konstruētajā ticamības intervālā. Savā starpā tika salīdzinātas 5 metodes: ML1 (maksimālās ticamības attiecības tests ar zināmu izkliedes rādītāju), ML2 (maksimālās ticamības attiecības tests ar nezināmu izkliedes rādītāju), Z-tests, t-tests un EL (empīriskās ticamības attiecības tests).

### Z-tests [15]

Pieņemsim, ka  $X_1, \dots, X_n$  ir gadījuma izlase no  $N(\mu, \sigma^2)$ , kur īstā  $\sigma^2$  vērtība ir zināma. Tad, lai pārbaudītu  $H_0 : \mu = \mu_0$  pret  $H_1 : \mu \neq \mu_0$ ,  $H_1 : \mu > \mu_0$  vai  $H_1 : \mu < \mu_0$  un lai konstruētu ticamības intervālus, izmanto testa statistiku

$$Z = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}.$$

Ja  $H_0$  ir spēkā, tad no centrālās robežteorēmas seko, ka testa statistikai  $Z$  asimptotiski ir standartnormālais  $N(0, 1)$  sadalījums.

### t-tests [15]

Pieņemsim, ka  $X_1, \dots, X_n$  ir gadījuma izlase no  $N(\mu, \sigma^2)$ , kur īstā  $\sigma^2$  vērtība nav zināma. Tad, lai pārbaudītu  $H_0 : \mu = \mu_0$  pret  $H_1 : \mu \neq \mu_0$ ,  $H_1 : \mu > \mu_0$  vai  $H_1 : \mu < \mu_0$  un lai konstruētu ticamības intervālus izmanto testa statistiku

$$T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S},$$

kur izlases dispersija  $S^2 = (n - 1)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ . Ja  $H_0$  ir spēkā, tad testa statistikai  $T$  asimptotiski ir Stjūdenta  $t$ -sadalījums ar  $n - 1$  brīvības pakāpi.

**Tabula 1:** Pārklājuma precizitāte  $N(0,1)$  sadalījumam,  $\alpha = 0.05$

	ML1	Z-tests	ML2	t-tests	EL
$n = 20$	0.9483	0.9493	0.9396	0.9495	0.9319
$n = 50$	0.9524	0.9537	0.9469	0.9514	0.9463
$n = 100$	0.9483	0.9500	0.9471	0.9510	0.9472
$n = 500$	0.9457	0.9504	0.9442	0.9495	0.9445
$n = 1000$	0.9430	0.9490	0.9430	0.9487	0.9430

Tabulā 1 redzams, ka ML1, Z-tests un t-tests strādā ļoti labi pat pie nelieliem izlašu apjomiem, ML2 un EL arī strādā samērā labi, taču varētu vēlēties, lai pie nelieliem izlašu apjomiem šie testi strādātu labāk. Pie pietiekami lieliem  $n$  empīriskās ticamības tests ir nopietns sāncensis apskatītajiem parametriskajiem testiem. Taču, kas notiek, ja ir nepareizi noteikts datu sadalījuma likums, kura metode, tad strādās labāk? Datu kopai, kurai ir standarta dubultais eksponenciālais sadalījums, kļūdas pēc nosaka standartnormālo sadalījumu.

**Tabula 2:** Pārklājuma precizitāte Dubultajam eksponenciālajam sadalījumam, kuru

pārbauda kā  $N(0,1)$ ,  $\alpha = 0.05$

	ML1	Z-tests	ML2	t-tests	EL
$n = 20$	0.8344	0.8356	0.9390	0.9516	0.9108
$n = 50$	0.8415	0.8426	0.9515	0.9550	0.9425
$n = 100$	0.8359	0.8384	0.9470	0.9494	0.9441
$n = 500$	0.8280	0.8310	0.9505	0.9512	0.9495
$n = 1000$	0.8341	0.8397	0.9511	0.9527	0.9512

Tabulā 2 redzamie rezultāti ir samērā interesanti, tie parāda, ka ML1 un Z-tests šajā gadījumā strādā slikti, kaut gan normālais un dubultais eksponenciālais sadalījums savā ziņā ir diezgan līdzīgi, abi ir simetriiski, tikai dubultajam eksponenciālajam ir "smagākas" astes (sk. Attēlu 1). Pārejās trīs metodes strādā labi, taču nevar teikt, ka EL ir vislabākā, jo pie maziem  $n$ , tā strādā sliktāk par t-testu.

**Tabula 3:** Pārklājuma precizitāte  $\chi^2_1$  sadalījumam, kuru pārbauda kā  $N(0,1)$ ,  $\alpha = 0.05$

	ML1	Z-tests	ML2	t-tests	EL
$n = 20$	0.8445	0.8445	0.8839	0.8929	0.8953
$n = 50$	0.8395	0.8395	0.9197	0.9224	0.9297
$n = 100$	0.8355	0.8355	0.9339	0.9352	0.9404
$n = 500$	0.8378	0.8378	0.9508	0.9512	0.9516
$n = 1000$	0.8360	0.8360	0.9454	0.9456	0.9470

Tabulā 3 redzams, ja  $\chi^2$  sadalījums tiek sajaukts ar normālo sadalījumu, tad vislabāk no apskatītajām metodēm strādā tieši EL, taču, ņemot vērā ticamības līmeni, nevar teikt, ka tā

labi strādā pie nelieliem izlašu apjomiem. Interesanti ir tas, ka arī t-tests šajā gadījumā strādā ļoti līdzīgi EL.

## 6 Bartleta korekcija vidējai vērtībai

Logaritmiskās empīriskās ticamības funkcijas attiecības asimptotiskā sadalījuma konvergences ātrums vidējās vērtības gadījumā ir  $O(n^{-1})$ , kur  $n$  ir izlases apjoms un, apskatot divpusējos ticamības intervālus, ar Bartleta korekcijas palīdzību konvergences ātrumu, var uzlabot līdz  $O(n^{-2})$ . [4]

### 6.1 Pārklājuma precizitāte

Empīriskās ticamības konvergences apgabals asimptotiski ir ar uzdoto līmeni, t.i., ja  $c$  ir izvēlēts saskaņā ar (3.1) un  $\mathcal{R}_c$  ir definēts ar (3.2), tad  $P(\theta_0 \in \mathcal{R}_c) \rightarrow 1 - \alpha$ ,  $n \rightarrow \infty$ . Īstā pārklājuma klūda,  $P(\theta_0 \in \mathcal{R}_c) - (1 - \alpha)$ , ir maza, ar lielumu  $n^{-1}$ . To var pamatot ar zīmes-saknes logaritmiskās empīriskās ticamības funkcijas attiecības testa statistikas palīdzību. Logaritmiskajai empīriskās ticamības attiecībai  $l(\theta_0) = W_E(\theta_0)$  saskaņā ar Teorēmu 8 ir asimptotisks  $\chi_p^2$  sadalījums un ņemot vērā to, ka  $\chi_p^2$  ir  $p$  neatkarīgu, standartnormāli sadalītu ( $N(0, 1)$ ) gadījuma lielumu kvadrātu summa, tad var rakstīt, ka

$$l(\theta_0) = W^T W,$$

kur  $W$  ir vektors ar garumu  $p$ , kurš ir asimptotiski normāls  $N(0, I_{p \times p})$ . DiCiccio, Hall un Romano (sk. [16] un [17]) izstrādāja formulas  $W$  sadalījumam. Viņi parādīja, ka  $W$  atzīst vispārēja tipa Edgeworth izvirzījumus [21] statistiskām funkcijām. Tas ir,  $W$  blīvuma funkcija  $f$  var tikt izteikta ar formulu:

$$f(w) = \phi(w) + \sum_{j=1}^k n^{-j/2} \pi_j(w) \phi(w) + O(n^{-(k+1)/2}), \quad (6.1)$$

kur  $k \geq 1$ ,  $w$  ir vektors ar garumu  $p$ ,  $\phi$  apzīmē  $N(0, I_{p \times p})$  blīvumu un  $\pi_j$  ir  $3j$ . pakāpes polinoms, kurš ir pāra pakāpes polinoms, ja  $j$  ir pāra skaitlis, un nepāra pakāpes polinoms, ja  $j$  ir nepāra skaitlis. Apzīmējot ar  $\mathcal{J}$  apgabalu  $p$ -dimensionālai sfērai ar rādiusu  $c^{1/2}$ , mēs iegūstam, ka

$$\begin{aligned} P(\theta_0 \in \mathcal{R}_c) &= P(l(\theta_0) \leq c) = P(W^T W \leq c) = P(W \in S) = \int_{\mathcal{J}} f(w) dw \\ &= \int_{\mathcal{J}} \phi(w) dw + n^{-\frac{1}{2}} \int_{\mathcal{J}} \pi_1(w) \phi(w) dw + O(n^{-1}). \end{aligned} \quad (6.2)$$

Pēdējā vienādība (6.2) izriet no (6.1). Pirmais integrālis vienādības (6.2) labajā malā ir vienāds ar

$$P(\chi_p^2 \leq c) = 1 - \alpha,$$

$c$  izvēloties no (3.1). Otrs integrālis tiecas uz nulli, jo  $\mathcal{J}$  ir sfēra ar centru koordinātu sākumpunktā un  $\pi_1$  ir nepāra pakāpes polinoms. Tādēļ

$$P(\theta_0 \in \mathcal{R}_c) = 1 - \alpha + O(n^{-1}),$$

kā sākumā tika minēts.

## 6.2 Bartleta korekcija

Hall un La Scala [4] aprakstīja Bartleta korekciju vidējās vērtības gadījumā. Empīriskās ticamības funkcijas ticamības reģiona pārklājuma neprecizitātes iemesls ir  $\chi^2$  aproksimācija. Kā jau iepriekš tika minēts, tad aproksimācijas klūda ir ar kārtu  $n^{-1}$ , kā rezultātā pārklājuma klūda ir ar kārtu  $n^{-1}$ . Bartleta korekcija uzlabo  $\chi^2$  aproksimācijas precizitāti. Daļa no  $l(\theta_0)$  sadalījuma  $\chi^2$  aproksimācijas klūdas var tikt izskaidrota ar faktu, ka divu sadalījumu vidējās vērtības nesakrīt, t.i.,  $E\{l(\theta_0)\} \neq p$ . Šī neatbilstība var tikt novērsta, koriģējot  $l(\theta_0)$ , tā lai tam ir vajadzīgā vidējā vērtība. Citiem vārdiem sakot,  $\chi^2$  aproksimāciju pielietot nevis  $l(\theta_0)$ , bet gan  $pl(\theta_0)/E\{l(\theta_0)\}$ .

Tad var parādīt [4], ka

$$E\{l(\theta_0)\} = p\{1 + n^{-1}a + O(n^{-2})\},$$

kur  $a$  ir konstante. Tātad vidējās vērtības korekcija līdz klūdai ar kārtu  $n^{-2}$  ir vienāda ar  $\chi^2$  aproksimācijas pielietošanu  $l(\theta_0)/(1 + n^{-1}a)$ . Ja  $a$  nav zināma, tad  $a$  aizstāj ar  $a\sqrt{n}$  būtisku novērtējumu  $\hat{a}$ , t.i.,  $\sqrt{n}(a - \hat{a}) \rightarrow N(0, 1)$ . Bartleta korekcija ir noteikt  $c$  no  $\chi^2$  tabulām tādu, lai

$$P(\chi_p^2 \leq c) = 1 - \alpha \tag{6.3}$$

un ticamības reģions būtu

$$\mathcal{R}'_c = \{\theta : l(\theta)/(1 + n^{-1}a) \leq c\} = \{\theta : l(\theta) \leq c(1 + n^{-1}a)\}$$

vai  $\mathcal{R}'_c = \{\theta : l(\theta) \leq c(1 + n^{-1}\hat{a})\}$ , ja izmanto  $a$  novērtējumu  $\hat{a}$ . Bartleta korekcija darbojas, jo pāra – 6. pakāpes polinoms  $\pi_2$  Edgeworth izvirzījumā (6.1) īstenībā ir ar pakāpi 2, visi 4. un 6. pakāpes  $\pi_2$  locekļi tiecas uz nulli. Tā ir būtiska empīriskās ticamības funkcijas īpašība, kuru pierādīja DiCiccio, Hall un Romano [16]. Pierādīsim, ka, ja  $\pi_2$  īstenībā ir 2. pakāpes polinoms, tad Bartleta korekcija samazina pārklājuma klūdu līdz  $O(n^{-2})$ . Vispārīgāk runājot, var parādīt, ka, ja  $W$  blīvuma funkcijai  $f$  ir spēkā Edgeworth izvirzījums

$$f(w) = \phi(w) + n^{-\frac{1}{2}}\pi_1(w)\phi(w) + n^{-1}\pi_2(w)\phi(w) + n^{-\frac{3}{2}}\pi_3(w)\phi(w) + O(n^{-2}), \tag{6.4}$$

kur  $\pi_1, \pi_3$  ir nepāra pakāpes polinomi un  $\pi_2$  ir pāra pakāpes (2. pakāpes) polinoms un ja  $c > 0$  ir uzdot ar (6.3) un  $a$  ir definēts ar  $E(W^T W) = p \{1 + n^{-1}a + O(n^{-2})\}$ , tad

$$P \{W^T W \leq c(1 + n^{-1}a)\} = 1 - \alpha + O(n^{-2}).$$

Empīriskās ticamības funkcijas gadījumā  $W$  var uzskatīt par zīmes-saknes logaritmiskās ticamības funkcijas attiecību. Nointegrējot (6.4) pār sfēru  $\{w : \|w\|^2 \leq c(1 + n^{-1}a)\}$ , un ņemot vērā, ka  $\pi_1$  un  $\pi_3$  ir nepāra pakāpes polinomi, iegūstam

$$\begin{aligned} P \{W^T W \leq c(1 + n^{-1}a)\} = \\ P \{\chi_p^2 \leq c(1 + n^{-1}a)\} + n^{-1} \int_{\|w\|^2 \leq c} \pi_2(w) \phi(w) dw + O(n^{-2}). \end{aligned} \quad (6.5)$$

Ja ar  $g$  apzīmē  $\chi_p^2$  blīvuma funkciju, tad

$$P \{\chi_p^2 \leq c(1 + n^{-1}a)\} = P(\chi_p^2 \leq c) + n^{-1} acg(c) + O(n^{-2}). \quad (6.6)$$

Tā kā  $\pi_2$  ir 2. pakāpes polinoms un  $\int \pi_2 \phi = 0$ , (jo  $f$  ir blīvuma funkcija), tad konstantēm  $a_j$  un  $a_{jk}$ ,

$$\pi_2(w) = \sum_{j=1}^p a_j (w^{(j)2} - 1) + \sum_{j \neq k} a_{jk} w^{(j)} w^{(k)},$$

kur  $w = (w^{(1)}, \dots, w^{(p)})^T$ . Tāpēc

$$\begin{aligned} \int_{\|w\|^2 \leq c} \pi_2(w) \phi(w) dw = \sum_{j=1}^p a_j \int_{\|w\|^2 \leq c} (w^{(j)2} - 1) \phi(w) dw = \\ \left( p^{-1} \sum_{j=1}^p a_j \right) \int_0^c (x - p) g(x) dx = -2 \left( p^{-1} \sum_{j=1}^p a_j \right) cg(c). \end{aligned} \quad (6.7)$$

Tāpat,

$$pa = \int \|w\|^2 \pi_2(w) \phi(w) dw = 2 \sum_{j=1}^p a_j. \quad (6.8)$$

Apvienojot (6.5) – (6.8), varam secināt, ka

$$P \{W^T W \leq c(1 + n^{-1}a)\} = 1 - \alpha + n^{-1} acg(c) - n^{-1} acg(c) + O(n^{-2}) = 1 - \alpha + O(n^{-2}).$$

Augstāk izdarītais aprēķinos mums bija pieņēmums, ka Bartleta korekcija tiek realizēta, izmantojot īsto  $a$  vērtību, nevis  $a$  novērtējumu  $\hat{a}$ . Gadījums ar  $\hat{a}$  ir ļoti līdzīgs. Tā kā  $\hat{a} = a + O_p(n^{-\frac{1}{2}})$  [4], tad  $1 + n^{-1}\hat{a} = 1 + n^{-1}a + O_p(n^{-\frac{3}{2}})$  un tas nozīmē, ka, ja  $a$  vietā izmanto  $\hat{a}$ , tad pēc Bartleta korekcijas pārklājuma kļūda būs vizmaz  $O(n^{-\frac{3}{2}})$ . Faktiski tā ir vienāda ar  $O(n^{-2})$ , ko pierādīja Barndorff-Nielsen un Hall [18]. Tas nozīmē, ka Bartleta korekcija samazina pārklājuma kļūdu

no  $O(n^{-1})$  uz  $O(n^{-2})$  neatkarīgi no tā, vai tiek izmantota  $a$  īstā vērtība, vai tās novērtējums  $\hat{a}$ . Vispārējā gadījumā konstantes  $a$  noteikšana ir ļoti sarežģīts uzdevums, kura atrisināšanai tiek izmantotas butstrapa metodes. Ja  $\theta = \mu$  ir populācijas vidējā vērtība, tad formulas iegūšana priekš  $a$  ir daudz vienkāršāks uzdevums.

Pieņemsim, ka populācijas dispersiju matrica  $\sum$  nav vienības matrica. Nemsim

$$Y = (Y^{(1)}, \dots, Y^{(p)})^T = \sum^{-\frac{1}{2}}(X - EX),$$

$$\alpha^{jkl} = E(Y^{(j)}Y^{(k)}Y^{(l)}), \alpha^{klm} = E(Y^{(j)}Y^{(k)}Y^{(l)}Y^{(m)}).$$

Tad

$$a = \frac{5}{3} \sum_j \sum_k \sum_l (\alpha^{jkl})^2 - 2 \sum_j \sum_k \sum_l \alpha^{jjl} \alpha^{kkl} + \frac{1}{2} \sum_j \sum_k \alpha^{jjkk}.$$

Lai aprēķinātu  $\hat{a}$ , ar  $\bar{X}$  un  $\widehat{\sum}$  apzīmēsim attiecīgi izlases vidējo vērtību un dispersiju. Nemsim  $Y_i = (Y_i^{(1)}, \dots, Y_i^{(p)})^T = \widehat{\sum}^{-\frac{1}{2}}(X_i - \bar{X})$  un definēsim

$$\hat{\alpha}^{jkl} = n^{-1} \sum_{i=1}^n Y_i^{(j)} Y_i^{(k)} Y_i^{(l)}, \hat{\alpha}^{klm} = n^{-1} \sum_{i=1}^n Y_i^{(j)} Y_i^{(k)} Y_i^{(l)} Y_i^{(m)},$$

$$\hat{a} = \frac{5}{3} \sum_j \sum_k \sum_l (\hat{\alpha}^{jkl})^2 - 2 \sum_j \sum_k \sum_l \hat{\alpha}^{jjl} \hat{\alpha}^{kkl} + \frac{1}{2} \sum_j \sum_k \hat{\alpha}^{jjkk}. [4]$$

**Piemērs 4.** (Viendimensionāla videjā vērtība) Pieņemsim, ka  $p = 1$  un ka  $\theta_0 = \mu_0 = EX$  ir populācijas vidējā vērtība. Nemsim  $\sigma^2 = D(X)$ ,  $\mu_3 = E(X - EX)^3/\sigma^3$ ,  $\mu_4 = E(X - EX)^4/\sigma^4$ ,  $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ ,  $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ ,  $\hat{\mu}_3 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^3/\hat{\sigma}^3$ ,  $\hat{\mu}_4 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^4/\hat{\sigma}^4$ . Bartleta korekcijai nepieciešamā konstante  $a$  ir

$$a = \frac{1}{2}\mu_4 - \frac{1}{3}\mu_3^2$$

un  $a$  novērtējums  $\hat{a} = \frac{1}{2}\hat{\mu}_4 - \frac{1}{3}\hat{\mu}_3^2$ . Barleta koriģētais ticamības reģions ir

$$\mathcal{R}'_c = \{\theta : l(\theta \leq c(1 + n^{-1}\hat{a}))\}.$$

Lai noskaidrotu kā Bartleta korekcija uzlabo pārklājuma precizitāti ticamības intervāliem, kas konstruēti videjai vērtībai, izmantojot īsto vērtību  $a$  un tās novērtējumu  $\hat{a}$ , tika veiktas simulācijas. Dažādiem sadalījumiem, ticamības līmeniem un izlašu apjomiem tika uzgenerētas 10000 izlases, lai konstruētu divpusējus ticamības intervālus, izmantojot 3 atšķirīgas metodes. Pirmā metode "Empīriskā ticamība" ir nekoriģēta empīriskās ticamības metode, kas izmanto  $\chi^2$  kritisko vērtību  $c_{1-\alpha}$ . Otrā metode "Teorētiskais Bartlets", konstruējot ticamības intervālus ar empīriskās ticamības metodes palīdzību,  $\chi^2$  kritiskās vērtības  $c_{1-\alpha}$  vietā izmanto

$c_{1-\alpha}/(1 - an^{-1})$ . Šī metode pieprasā zināšanas par populācijas īstajām momentu vērtībām. Pēdējā metode ”Novērtētais Bartlets”  $a$  vietā izmanto  $\hat{a}$ , kas iegūts no datiem.

**Tabula 4:** Bartleta korekcija  $N(0,1)$  sadalījuma datiem

	Ticamības līmenis		
	90%	95%	99%
$n = 10$			
Empīriskā ticamība	0.8429	0.8975	0.9552
Teorētiskais Bartlets	0.8674	0.9162	0.9650
Novērtētais Bartlets	0.8616	0.9118	0.9633
$n = 20$			
Empīriskā ticamība	0.8786	0.9334	0.9787
Teorētiskais Bartlets	0.8924	0.9420	0.9830
Novērtētais Bartlets	0.8900	0.9410	0.9825
$n = 50$			
Empīriskā ticamība	0.8910	0.9456	0.9867
Teorētiskais Bartlets	0.8952	0.9486	0.9879
Novērtētais Bartlets	0.8948	0.9482	0.9877

No iepriekšējās nodaļas simulāciju rezultātiem, tika secināts, ka empīriskās ticamības attiecības tests nav tik efektīvs tieši pie nelielam izlašu apjomam, tāpēc īpaša uzmanība tiks pievērsta tam, kādu uzlabojumu Bartleta korekcija dod neliela apjoma izlasēm ( $n = 10, 20, 50$ ). Tabulā 4 redzamie rezultāti ir iegūti, izmantojot standartnormālā sadalījuma  $N(0,1)$  datus, Tabulā 5 redzamie rezultāti ir iegūti, izmantojot  $\chi_1^2$  sadalījuma datus un Tabulā 6 redzamie rezultāti ir iegūti, izmantojot  $t_5$  sadalījuma datus. Īstās parametra  $a$  vērtības apskatītajiem sadalījumiem ir attiecīgi:  $3/2$ ,  $29/6$  un  $9/2$ . [5]

**Tabula 5:** Bartleta korekcija  $\chi^2_1$  sadalījuma datiem

	Ticamības līmenis		
	90%	95%	99%
<i>n</i> = 10			
Empīriskā ticamība	0.7704	0.8329	0.8975
Teorētiskais Bartlets	0.8389	0.8826	0.9297
Novērtētais Bartlets	0.7906	0.8480	0.9074
<i>n</i> = 20			
Empīriskā ticamība	0.8394	0.8925	0.9523
Teorētiskais Bartlets	0.8741	0.9198	0.9669
Novērtētais Bartlets	0.8540	0.9042	0.9587
<i>n</i> = 50			
Empīriskā ticamība	0.8710	0.9265	0.9776
Teorētiskais Bartlets	0.8876	0.9387	0.9818
Novērtētais Bartlets	0.8797	0.9328	0.9793

**Tabula 6:** Bartleta korekcija  $t_5$  sadalījuma datiem

	Ticamības līmenis		
	90%	95%	99%
<i>n</i> = 10			
Empīriskā ticamība	0.8216	0.8862	0.9492
Teorētiskais Bartlets	0.8890	0.9334	0.9735
Novērtētais Bartlets	0.8442	0.9025	0.9575
<i>n</i> = 20			
Empīriskā ticamība	0.8611	0.9159	0.9753
Teorētiskais Bartlets	0.8957	0.9431	0.9864
Novērtētais Bartlets	0.8741	0.9267	0.9794
<i>n</i> = 50			
Empīriskā ticamība	0.8881	0.9408	0.9834
Teorētiskais Bartlets	0.9040	0.9505	0.9885
Novērtētais Bartlets	0.8951	0.9451	0.9860

Aplūkojot iegūtos rezultātus, jāsecina, ka pie maziem izlašu apjomiem ar nekoriģētās empīriskās ticamības metodes palīdzību konstruēto ticamības intervālu pārklājuma līmenis diezgan ievērojami atšķiras no uzdotā ticamības līmeņa. Teorētiskā un novērtētā Bartleta korekcija ievērojami uzlabo pārklājuma precizitāti, pie tam pie lielākiem izlašu apjomiem "Teorētiskā Bartleta" rezultāti ir tikai nedaudz labāki par "Novērtēto Bartletu". Vissarežģītākā situācija ir ar  $\chi^2$  sadalījumu, ar "Teorētisko Bartletu" iegūtie rezultāti ir diezgan apmierinoši, taču "Novērtētais Bartlets" minimāli uzlabo "Empīrisko ticamību", ko var izskaidrot ar to, ka  $\chi^2$  sadalījums ir asimetrisks.

## 7 Ticamības intervālu konstruēšana kvantilēm ar EL

Kvantiles ir svarīgs populācijas raksturotājs, dažkārt kvantiles ir labākie parametru novērtējumi. Piemēram, ja mēs vēlamies novērtēt Košī sadalījuma, kura blīvuma funkcija, ja  $\gamma = 1$ , ir  $f(x) = \pi^{-1} [1 + (x - \mu)^2]^{-1}$ ,  $-\infty < x < +\infty$  parametru  $\mu$ , izrādās, ka izlases vidējā vērtība  $\bar{X}$  nav būtisks izvietojuma rādītāja  $\mu$  novērtējums, bet izlases mediāna  $\theta_{1/2}$  ir, tā asymptotiski ir normāli sadalīta, t. i.,  $\theta_{1/2} \rightarrow N(\mu, \pi^2/4n)$ .

Parastajiem EL ticamības intervāliem kvantilēm ir relatīvi liela pārklājuma klūda ar kārtu  $n^{-1/2}$  pat divpusējo intervālu gadījumā. Savukārt ar kodolu metožu palīdzību nogludinātās EL ticamības intervālu pārklājuma klūda ir ar kārtu  $n^{-1}$ , pie samērā plašas nogludināšanas parametra izvēles. Vilkša teorēma nosaka, kādiem nosacījumiem jābūt spēkā, lai testa statistikai būtu  $\chi^2$  sadalījums. Darbā tiks apskatītas divas EL gludināšanas metodes, vienu piedāvāja Chen un Hall [7], bet otro Zhou un Jing [8].

### 7.1 Kodolu blīvuma funkcijas novērtēšana

Bieži vien cenšoties noteikt kādas datu kopas sadalījumu, mēs zīmējam histogrammas un ar to palīdzību izvirzām hipotēzi par iespējamo datu sadalījumu, taču pastāv arī citas iespējas, var, piemēram, izmantot kodolu gludināšanas metodi. Kodolu gludināšana ir neparametriska blīvuma funkcijas novērtēšanas metode un ar tās palīdzību iegūtais blīvuma funkcijas novērtējums ir daudz gludāks un uz īsto sadalījumu konverģē daudz ātrāk kā histogrammas.

Pieņemsim, ka  $X_1, \dots, X_n$  ( $X_i$ -neatkarīgi un vienādi sadalīti) ir izlase ar blīvuma funkciju  $f$ . Kodols tiek definēts kā gluda funkcija  $K$ , kurai ir spēkā, ka  $K(x) \geq 0$ ,  $\int K(x)dx = 1$ ,  $\int xK(x)dx = 0$  un  $\int x^2K(x)dx > 0$ .

Daži kodolu piemēri: Epanečnikova kodols

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2/5)/\sqrt{5}, & |x| < \sqrt{5} \\ 0, & \text{pretējā gadījumā} \end{cases},$$

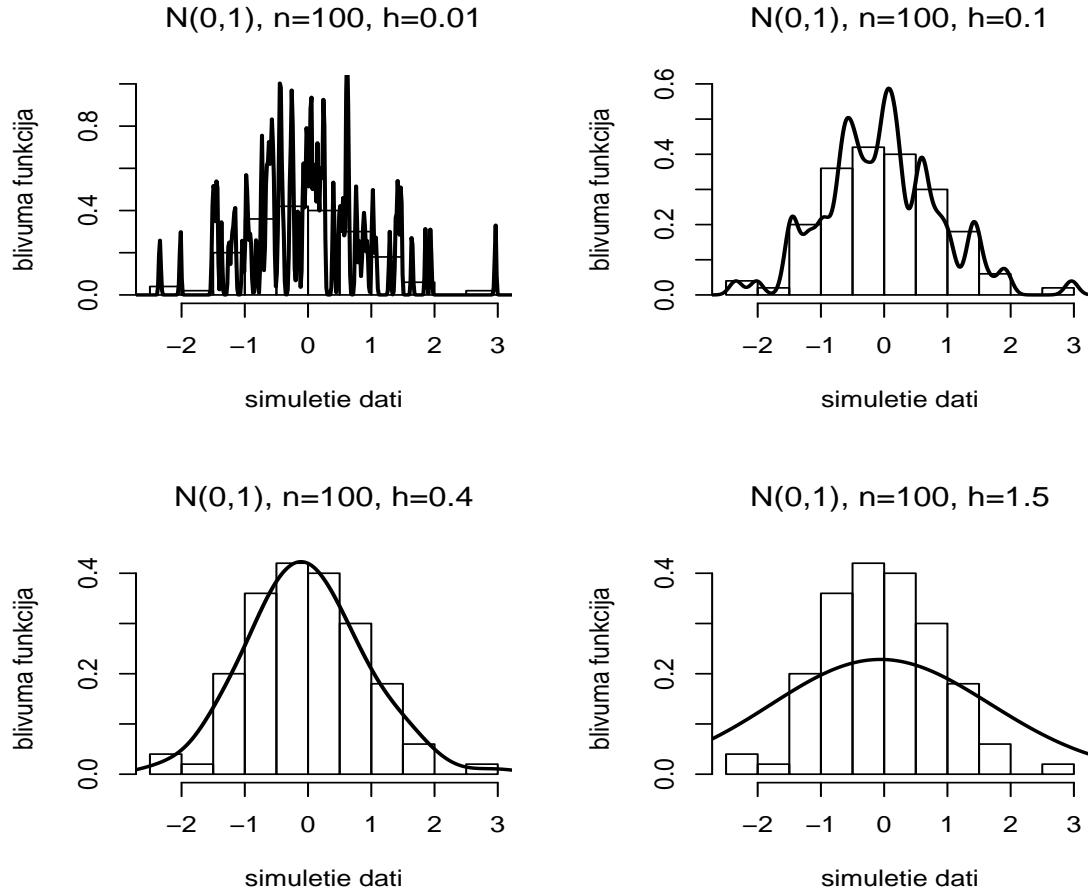
normālais jeb Gausa kodols

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

**Definīcija 8.** [1] Ja  $K$  ir kodols un joslas platums  $h$  ir pozitīvs skaitlis, tad

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - X_i}{h}\right)$$

ir kodolu blīvuma funkcijas novērtējums.



**Attēls 4:**  $N(0,1)$  sadalījuma ģenerēšana, izmantojot kodolu gludināšanas metodi,  $n=100$

Attēlā 4 ir redzams kodolu blīvuma funkcijas novērtējums standartnormālajam sadalījumam  $N(0,1)$  pie dažādiem joslas platumiem  $h$ , kā var redzēt, tad  $h$  nosaka gludināšanas pakāpi. Ja  $h$  ir tuvs 0, tad  $\hat{f}_n$  sastāv no pīķiem, pīķis katrā datu punktā. Pīķu augstums tiecas uz bezgalību, kad  $h \rightarrow 0$ . Kad  $h \rightarrow \infty$ , tad  $\hat{f}_n$  tiecas uz vienmērīgo sadalījumu.

## 7.2 EL metodes ticamības intervālu konstruēšanai kvantilēm

Vispirms apskatīsim parasto, negludināto EL metodi ticamības intervālu konstruēšanai. ([9], [8])

Pienemsim, ka  $X_1, \dots, X_n$  ( $X_i$ -neatkarigi un vienādi sadalīti) ir gadījuma izlase no nezināma sadalījuma  $F(x)$  ar blīvuma funkciju  $f(x)$  un ka  $\theta_q = F^{-1}(q)$  ir  $q$ -tā kvantile,  $0 < q < 1$ .  $X_{(1)}, \dots, X_{(n)}$  ir sakārtotā izlase un  $\theta_q$  novērtējums ir  $\hat{\theta}_q = X_{([nq])}$ , kur  $[nq]$  ir skaitļa  $nq$  veselā

daļa.  $\theta$  var definēt ar tā saukto M-vienādojumu ([10], [9]) palīdzību

$$\int_{-\infty}^{+\infty} \phi(x - \theta) dF(x) = 0, \quad (7.1)$$

kur

$$\phi(z) = \begin{cases} -1, & \text{ja } z \leq 0, \\ q/(1-q), & \text{ja } z > 0. \end{cases} \quad (7.2)$$

Empīriskās ticamības attiecība parametram  $\theta$  ir

$$R(\theta) = \sup \left\{ \prod_{i=1}^n np_i \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \phi(X_i - \theta_0) = 0 \right\} \quad (7.3)$$

un logaritmiskā empīriskās ticamības attiecība ir

$$l(\theta) = -2 \ln(R(\theta)). \quad (7.4)$$

Maksimizācijas problēmu (7.3), līdzīgi kā vidējās vērtības gadījumā, risina ar Lagranža reizinātāju palīdzību. Var pierādīt, ka maksimums tiks sasniegts, ja

$$p_i = \frac{1}{n} \frac{1}{\lambda(\theta)\phi(X_i - \theta)}, \quad i = 1, \dots, n, \quad (7.5)$$

kur  $\lambda(\theta)$  var noteikt no

$$\frac{1}{n} \sum_{i=1}^n \frac{\phi(X_i - \theta)}{1 + \lambda(\theta)\phi(X_i - \theta)} = 0. \quad (7.6)$$

Kad  $\theta \in [X_{(1)}, X_{(n)}]$ , tad (7.3) atrisinājums  $\lambda(\theta)$  ir

$$\lambda(\theta) = (q - F_n(\theta))/q, \quad (7.7)$$

kur  $F_n$  ir empīriskā sadalījuma funkcijas, kas definēta ar (4.1).

□ Dati  $X_1, \dots, X_n$  ir sakārtoti augošā secībā un funkcija  $\phi(z)$  definēta ar (7.2).

$$\frac{1}{n} \sum_{i=1}^m \frac{\phi(X_i - \theta)}{1 + \lambda\phi(X_i - \theta)} + \frac{1}{n} \sum_{i=m+1}^n \frac{\phi(X_i - \theta)}{1 + \lambda\phi(X_i - \theta)} = 0$$

$$\frac{1}{n} \sum_{i=1}^m \frac{-1}{1 + \lambda(-1)} + \frac{1}{n} \sum_{i=m+1}^n \frac{q(1-q)^{-1}}{1 + \lambda q(1-q)^{-1}} = 0$$

$$\frac{m}{n} \frac{-1}{1 - \lambda} + \frac{n-m}{n} \frac{q(1-q)^{-1}}{1 + \lambda q(1-q)^{-1}} = 0$$

$$\frac{m}{n} \left( \frac{-1}{1 - \lambda} - \frac{q(1-q)^{-1}}{1 + \lambda q(1-q)^{-1}} \right) = - \left( \frac{q(1-q)^{-1}}{1 + \lambda q(1-q)^{-1}} \right)$$

$$F_n(\theta) \left( \frac{-1}{1 - \lambda} - \frac{q}{(1-q) + \lambda q} \right) = - \frac{q}{(1-q) + \lambda q}$$

$$F_n(\theta) \left( \frac{-1}{1 - \lambda} \right) = -q \Rightarrow \lambda = \frac{q - F_n(\theta)}{q}. \blacksquare$$

Tādējādi, ievietojot iegūto  $\lambda$  izteiksmē (7.4), iegūstam, ka logaritmiskā empīriskās ticamības attiecība ir

$$l(\theta) = -2 \ln R(\theta) = 2n \left( F_n(\theta) \ln \frac{F_n(\theta)}{q} + (1 - F_n(\theta)) \ln \frac{1 - F_n(\theta)}{1 - q} \right). \quad (7.8)$$

□ Dati  $X_1, \dots, X_n$  ir sakārtoti augošā secībā un funkcija  $\phi(z)$  definēta ar (7.2) un  $\lambda$  ar (7.4).

$$\begin{aligned} & 2 \sum_{i=1}^n \ln \left( 1 + \frac{q - F_n(\theta)}{q} \phi(X_i - \theta) \right) \\ &= 2 \sum_{i=1}^m \ln \left( 1 + \frac{q - F_n(\theta)}{q} (-1) \right) + 2 \sum_{i=m+1}^n \ln \left( 1 + \frac{q - F_n(\theta)}{q} \frac{q}{1 - q} \right) \\ &= 2m \ln \left( \frac{F_n(\theta)}{q} \right) + 2(n-m) \ln \left( 1 + \frac{q - F_n(\theta)}{1 - q} \right) \\ &= 2n F_n(\theta) \left( \ln \left( \frac{F_n(\theta)}{q} \right) - \ln \left( 1 + \frac{q - F_n(\theta)}{1 - q} \right) \right) + \ln \left( 1 + \frac{q - F_n(\theta)}{1 - q} \right) \\ &= 2n \left( F_n(\theta) \ln \left( \frac{F_n(\theta)}{q} \right) + \ln \left( 1 + \frac{q - F_n(\theta)}{1 - q} \right) (1 - F_n(\theta)) \right) \\ &= 2n \left( F_n(\theta) \ln \frac{F_n(\theta)}{q} + (1 - F_n(\theta)) \ln \frac{1 - F_n(\theta)}{1 - q} \right). \blacksquare \end{aligned}$$

Funkcija  $\ln R(\theta)$  no izteiksmes (7.8) ir pakāpienveida funkcija ar lēcieniem datu punktos, tā kā  $\ln R(\theta)$  pieņem galīga skaita vērtības, tad tas  $\chi_1^2$  aproksimāciju padara ne īpaši precīzu. Tālāk tiks apskatītas 2 gludināšanas metodes, kas uzlabo novērtējumu precizitāti.

### Chen un Hall pieeja. [7]

Šo metodi Chen un Hall [7] prezentēja 1993. gadā. Šajā metodē  $\lambda(\theta)$  netiek izteikta formā (7.7) un līdz ar to arī statistika  $l(\theta)$  nav formā (7.8). Viņi izmanto kodolu gludināšanas metodi, lai panāktu, ka ticamības intervālu pārklājuma klūda ir ar kārtu  $n^{-1}$  un parāda, ka pie zināmiem nosacījumiem, izmantojot Bartleta korekciju, var panākt, ka ticamības intervālu pārklājuma klūda ir ar kārtu  $n^{-2}$ . Definēsim

$$l(\theta) = 2 \sum_{i=1}^n \ln \{1 + \lambda(\theta)w_i(\theta)\}$$

un  $\lambda(\theta)$  noteikšanai izmanto izteiksmi

$$\sum_{i=1}^n w_i(\theta) \{1 + \lambda(\theta)w_i(\theta)\}^{-1} = 0,$$

kur  $w_i = G((\theta - X_i)/h) - q$ . Ar  $K$  tiek apzīmēts  $r$ -tās kārtas kodols, kuram pie  $r \geq 2$  un  $C \neq 0$  ir spēkā

$$\int u^j K(u) du = \begin{cases} 1, & j = 0, \\ 0, & 1 \leq j \leq r, \\ C, & j = r, \end{cases}, \quad (7.9)$$

un  $G(t) = \int_{-\infty}^t K(x) dx$ . Gadījums, kad  $r = 2$ , ir visbiežāk sastopamais, tad  $K$  ir simetriska blīvuma funkcija un  $G(t)$  ir sadalījuma funkcija. Ar  $f(x) = F'(x)$  tiek apzīmēts funkcijas  $F$  pirmais atvasinājums.

**Teorēma 9.** (Chen, Hall [7]) Ja  $K$  ir spēkā (7.9), tas ir ierobežots un tam ir kompakts definīcijas apgabals, un ja eksistē  $f$  un  $f^{(r-1)}$  punkta  $\theta$  apkārtnē, kā arī tie ir nepārtraukti punktā  $\theta$  un  $f(\theta) > 0$ , un ja dažiem  $t > 0$   $nh^t \rightarrow 0$ , kad  $n \rightarrow \infty$ . Tad  $\hat{l}(\theta)$  ir asimptotisks  $\chi_1^2$  sadalījums, ja  $nh^r \rightarrow 0$  un  $f^{(r-1)}(\theta) \neq 0$ .

Teorēmas 9 nosacījumi nodrošina, ka  $K$  ir  $r$ -tās kārtas kodols, ka sadalījuma funkcija  $F$  ir pietiekami gluda  $\theta$  apkārtnē. Nosacījums, ka  $f(\theta) > 0$  nodrošina to, ka asimptotiskā izlases kvantiles dispersija ir ar kārtu  $n^{-1}$ . Un tas, ka  $nh^t \rightarrow 0$ , kad  $n \rightarrow \infty$  nodrošina, ka joslas platums  $h$  nekonverģē uz nulli pārāk lēni. Ja  $K$  ir 2. kārtas kodols un  $f'(\theta) \neq 0$ , tad  $l(\theta)$  asimptotiski ir  $\chi_1^2$  tad un tikai tad, ja  $h = o(n^{-1/4})$ . Ja  $f^{(r-1)}(\theta) = 0$ , ir iespējams, ka  $l(\theta)$  būs asimptotisks  $\chi_1^2$ ,

taču  $nh^{2r}$  netieksies uz 0. Ja izpildās Teorēmas 9 nosacījumi un  $nh^{2r} \rightarrow 0$ , tad seko, ka

$$P(l(\theta) \leq c) \rightarrow P(\chi_1^2 \leq c),$$

kur  $c$  ir izvēlēts atbilstoši  $\chi^2$  tabulām un  $P(\chi_1^2 \leq c) = \alpha$ .

**Teorēma 10.** (Chen, Hall [7]) Pieņemsim, ka izpildās Teorēmas 9 nosacījumi un joslas platumam  $h$  ir spēkā

$$nh^{2r} \rightarrow 0 \text{ un } nh/\ln n \rightarrow \infty, \quad (7.10)$$

un  $c = c(\alpha)$ . Tad

$$P(\theta \in I_c) = \alpha + O(n^{-1}) \quad (7.11)$$

pietiekamais nosacījums ir  $nh^r$  ierobežotība, kad  $n \rightarrow \infty$ . Šis nosacījums ir nepieciešams arī, ja  $f'(\theta) \neq 0$ . Neskatoties uz to, ka  $c > 0$ , (7.11) labā puse nevar būt vienāda ar  $\alpha + o(n^{-1})$ , attiecīgi neizvēloties  $h$ .

Teorēmas 9 pierādījumā tiek izmantoti Edgeworth izvirzījumi. Teorēmu 9 un 10 pierādījumi atrodami [7].

### Bartleta korekcija kvantilēm

Var pierādīt, ka, ja  $nh^r \rightarrow 0$ , tad

$$E\{l(\theta)\} = 1 + n^{-1}\beta + o(n^{-1}),$$

kur  $\beta = \frac{1}{6}(3\mu_2^{-2}\mu_4 - 2\mu_2^{-3}\mu_3^2)$  un  $\mu_j = E[G\{(\theta - X_i)/h\} - q]^j$ . Tātad  $l(\theta)$  matemātiskā cerība no aproksimētā  $\chi^2$  sadalījuma atšķiras par  $n^{-1}\beta$ . Tas nozīmē, ka zināms uzlabojums,  $\chi^2$  aproksimāciju padarītu daudz efektīvāku. Ja  $n^3h^{2r}$  ir ierobežots, tad  $l(\theta)/(1 + n^{-1}\beta)$  sadalījums no  $\chi_1^2$  sadalījuma atšķirsies ar kārtu  $n^{-2}$ , nevis ar kārtu  $n^{-1}$ . Praksē  $\beta$  nav zināma un tā ir jānovērtē,

$$\hat{\mu}_j = n^{-1} \sum_{i=1}^n \left[ G\left\{(\hat{\theta} - X_i)/h\right\} - q \right]^j$$

un  $\hat{\beta} = \frac{1}{6}(3\hat{\mu}_2^{-2}\hat{\mu}_4 - 2\hat{\mu}_2^{-3}\hat{\mu}_3^2)$ , kur  $\hat{\theta}$  ir  $\theta$  novērtējums.

**Teorēma 11.** (Chen, Hall [7]) Pieņemsim, ka izpildās Teorēmas 9 nosacījumi un (7.10), un  $c = c(\alpha)$ , tad

$$P(\theta \in I_{d(c,\gamma)}) = \alpha + O(n^{-2}) \quad (7.12)$$

pietiekamais nosacījums ir  $n^3h^{2r}$  ierobežotība, kad  $\gamma = \beta$  vai  $\gamma = \hat{\beta}$ , kur  $d(c, \gamma) = c(1 + n^{-1}\gamma)$ . Ja  $f^{(r-1)}(\theta) \neq 0$ , tad  $n^3h^{2r}$  ierobežotība arī ir nepieciešama, lai būtu spēkā (7.12).

Tas nozīmē, ka, pielietojot Bartleta korekciju, nogludinātās EL ticamības intervāliem, var iegūt mazāku pārklājuma klūdu ar kārtu  $n^{-2}$ . [7]

### Zhou un Jing pieeja [8]

Zhou un Jing savu pieeju EL gludināšanai ar kodolu metodes palīdzību, publicēja 2003. gadā. Šī metode no iepriekšējās atšķiras ar to, ka šie autori izmanto logaritmisko empīriskās ticamības attiecību (7.8), kas iegūta negludinātajā gadījumā. Empīriskā sadalījuma funkcija  $F_n$  tiek aizstāta ar nogludināto empīriskā sadalījuma funkciju

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n G\left(\frac{x - X_i}{h}\right). \quad (7.13)$$

Aizstājot izteiksmē (7.8)  $F_n(x)$  ar  $\hat{F}_n(x)$ , iegūsim koriģēto logaritmisko empīriskās ticamības attiecību

$$\hat{l}(\theta) = 2n \left( \hat{F}_n(\theta) \ln \frac{\hat{F}_n(\theta)}{q} + (1 - \hat{F}_n(\theta)) \ln \frac{1 - \hat{F}_n(\theta)}{1 - q} \right). \quad (7.14)$$

Nogludinātās empīriskās ticamības intervāls parametram  $\theta$  ir  $I_c = \{\theta : \hat{l}(\theta) \leq c\}$ , kur  $c$  ir konstante, ko nosaka, izmantojot uzdoto ticamības līmeni  $\alpha$ ,  $\alpha = P(\theta \in I_c) = P\{\hat{l}(\theta) \leq c\}$ .

**Teorēma 12.** (Zhou, Jing, [8]) Ja  $\hat{l}(\theta)$  ir definēts ar (7.14) un izpildās visi Teorēmas 9 nosacījumi, kā arī kodolam  $K$  ir kompakts definīcijas apgabals  $[a, b]$  un kādai  $[a, b]$  dekompozīcijai,  $a = u_0 < u_1 < \dots < u_m = b$ , kodols  $K$  ir vai nu tikai pozitīvs, vai negatīvs katrā no intervaliem  $(u_{j-1}, u_j)$ , kur  $j = 1, \dots, m$ , un  $\int u K(u) G(u) du = 0$ , tad, ja  $n \rightarrow \infty$ ,

$$P(\hat{l}(\theta) \leq x) - P(\chi_1^2 \leq x) = O(n^{-1})$$

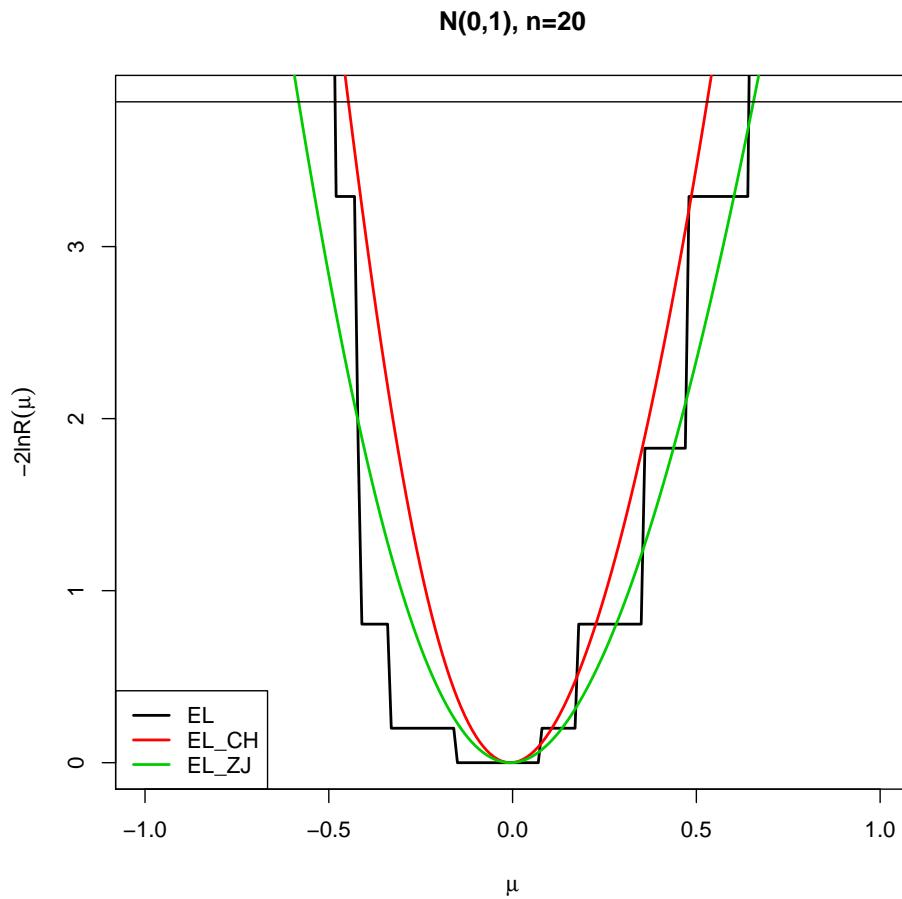
katram fiksētam  $x$ .

No Teorēmas 12 seko, ka  $\lim_{n \rightarrow \infty} P(\theta_0 \in I_{hc}) = P(\chi_1^2 \leq c)$ , kur  $I_{hc} = \{\theta : \hat{l}(\theta) \leq c\}$ . Ja  $c$  ir izvēlēts tā, lai  $P(\chi_1^2 \leq c) = 1 - \alpha$ , tad  $I_{hc}$  pārklājuma varbūtība aproksimēs  $\alpha$  ar klūdu  $O(n^{-1})$ , kad  $n \rightarrow \infty$ .

## 7.3 EL metožu salīdzinājums

Šajā sadalā, izmantojot pārklājuma precizitāti, salīdzināsim visas 3 apskatītās metodes. Tiks apskatīti 2 sadalījumi viens simetrisks (standartnormālais sadalījums) un viens asimetrisks ( $\chi_3^2$  sadalījums). Abiem sadalījumiem 10000 reižu tika generēti dati pie izlašu apjomiem  $n = 20$  un

$n = 50$ , un ar apskatīto metožu palīdzību noteikti ticamības intervāli mediānai pie dažādiem nozīmības līmeniem.



**Attēls 5:** Negludinātā (EL) un gludinātās ( $EL_{CH}$ ,  $EL_{ZJ}$ )  $-2 \times \ln$  ticamības līknes standartnormālā sadalījuma mediānai,  $\alpha = 0.05$ .

Attēlā 5 redzamas logaritmiskās empīriskās ticamības attiecības līknes standartnormālā  $N(0,1)$  sadalījuma mediānai. EL ir parastā EL metode,  $EL_{CH}$  ir Chen un Hall gludinātā EL metode un  $EL_{ZJ}$  ir Zhou un Jing gludinātā EL metode. Konkrētajā gadījumā redzams, ka gludinātās metodes dod labākus rezultātus, jo abos gadījumos neparametriskās vislielākās ticamības parameatra  $\mu$  novērtējums ir  $-0.01$ , kas, ņemot vērā izlases apjomu, ir ļoti tuvu īstajai mediānas vērtībai  $\mu = 0$ . Arī ticamības intervāli gludinātajām metodēm ir šaurāki kā negludinātajai (EL [-0.48;0.64],  $EL_{CH}$  [-0.44;0.52],  $EL_{ZJ}$  [-0.58;0.65]).

**Tabula 7:** Pārklājuma precizitāte  $\chi^2_3$  mediānai

$n = 20$	$h$	EL	Ticamības līmenis		
			90%	95%	99%
$n = 20$	$h$	EL	0.8879	0.9609	0.9888
$n^{-1}$	$EL_{CH}$	0.8901	0.9563	0.9887	
	$EL_{CHB}$	0.8922	0.9574	0.9890	
	$EL_{ZJ}$	0.8952	0.9585	0.9897	
$n^{-3/4}$	$EL_{CH}$	0.8945	0.9524	0.9889	
	$EL_{CHB}$	0.8981	0.9541	0.9897	
	$EL_{ZJ}$	0.9033	0.9568	0.9915	
$n^{-1/2}$	$EL_{CH}$	0.8956	0.9490	0.9895	
	$EL_{CHB}$	0.9000	0.9509	0.9902	
	$EL_{ZJ}$	0.9135	0.9586	0.9932	
$n^{-1/4}$	$EL_{CH}$	0.8933	0.9447	0.9887	
	$EL_{CHB}$	0.8986	0.9476	0.9899	
	$EL_{ZJ}$	0.9302	0.9686	0.9952	
$n = 50$	$h$	EL	0.8798	0.9347	0.993
$n^{-1}$	$EL_{CH}$	0.8859	0.9425	0.9914	
	$EL_{CHB}$	0.8866	0.9437	0.9917	
	$EL_{ZJ}$	0.8874	0.9438	0.9914	
$n^{-3/4}$	$EL_{CH}$	0.8927	0.9464	0.9906	
	$EL_{CHB}$	0.8942	0.9482	0.9912	
	$EL_{ZJ}$	0.8965	0.9495	0.9913	
$n^{-1/2}$	$EL_{CH}$	0.8953	0.9490	0.9906	
	$EL_{CHB}$	0.8971	0.9493	0.9908	
	$EL_{ZJ}$	0.9058	0.9561	0.9923	
$n^{-1/4}$	$EL_{CH}$	0.8926	0.9463	0.9893	
	$EL_{CHB}$	0.8943	0.9483	0.9903	
	$EL_{ZJ}$	0.9223	0.9658	0.9947	

Kā jau iepriekš tika uzsvērts kodolu gludināšanās metodē,  $h$  izvēlei ir ļoti liela nozīme, tāpēc, veicot simulācijas, tika izvēlētas vairākas  $h$  vērtības ( $n^{-1}$ ,  $n^{-3/4}$ ,  $n^{-1/2}$  un  $n^{-1/4}$ ). Tabulā 7 ir redzami simulāciju rezultāti  $\chi^2_3$  sadalījuma mediānai pie nozīmības līmeņiem  $\alpha = 0.05$ , 0.1, un 0.01. EL ir parastā EL metode (negludinātā),  $EL_{CH}$  ir Chen un Hall gludinātā EL metode,  $EL_{CHB}$  ir Chen un Hall gludinātā EL metode ar Bartleta korekciju un  $EL_{ZJ}$  ir Zhou un Jing

gludinātā EL metode. Aplūkojot rezultātus, jāsecina, ka jau pie tik nelieliem izlašu apjomiem, apskatītās metodes strādā ļoti labi un ka gludinātās metodes ir labākas par negludināto. Bartleta korekcijas uzlabojums ir samērā minimāls, taču rezultāti rāda, ka, palielinoties  $n$ , Bartleta korekcijas uzlabojums samazinās un palielinoties  $h$  Bartleta korekcijas uzlabojums palielinās.  $EL_{ZJ}$  pārklājuma precizitāte visos gadījumos ir lielāka kā Chen un Hall metodei. Izvērtējot kā mainās pārklājuma precizitāte, mainoties  $h$ , jāsecina, ka  $EL_{CH}$  un  $EL_{CHB}$  minimāli ietekmē izvēlētā  $h$  amplitūda, savukārt  $EL_{ZJ}$  ir daudz jūtīgāka pret  $h$  izmaiņām.

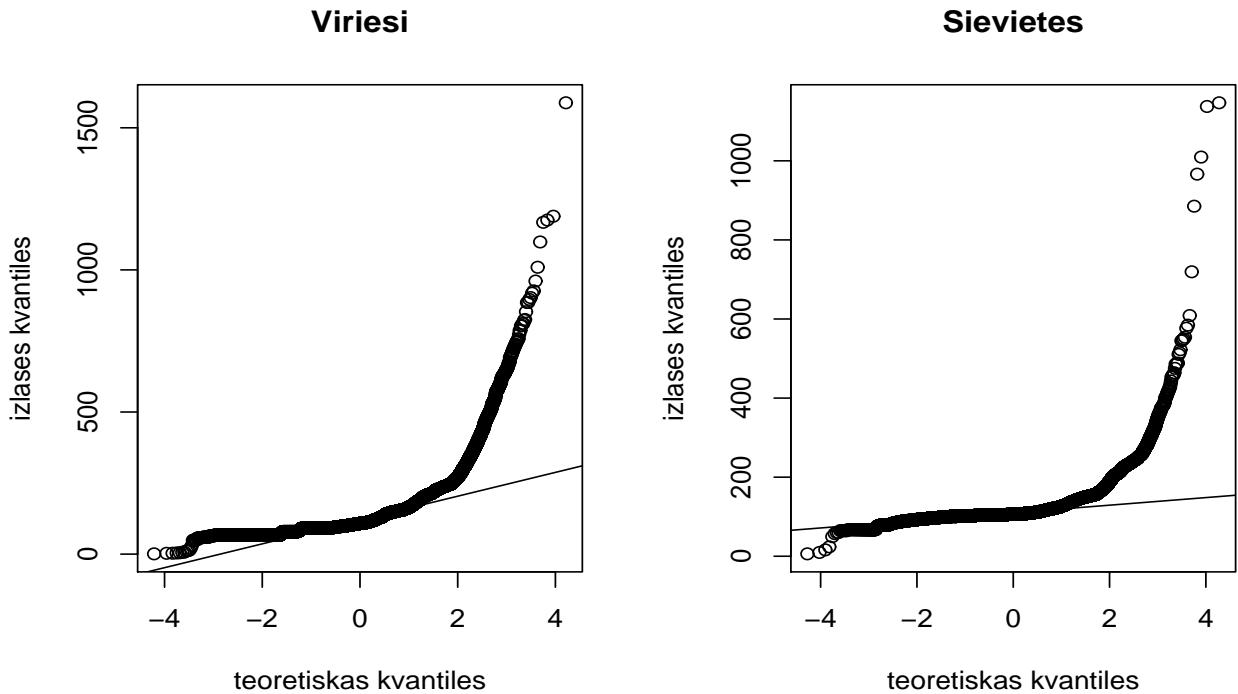
**Tabula 8:** Pārkājuma precizitāte  $N(0,1)$  mediānai

$n = 20$	$h$	EL	Ticamības līmenis		
			90%	95%	99%
$n = 20$	$h$	EL	0.8837	0.9589	0.9884
$n^{-1}$	$EL_{CH}$	0.8898	0.9498	0.9885	
	$EL_{CHB}$	0.8942	0.9509	0.9889	
	$EL_{ZJ}$	0.8988	0.9550	0.9907	
$n^{-3/4}$	$EL_{CH}$	0.8951	0.9468	0.9883	
	$EL_{CHB}$	0.8996	0.9507	0.9897	
	$EL_{ZJ}$	0.9117	0.9598	0.9930	
$n^{-1/2}$	$EL_{CH}$	0.8977	0.9465	0.9881	
	$EL_{CHB}$	0.9026	0.9505	0.9895	
	$EL_{ZJ}$	0.9327	0.9703	0.9959	
$n^{-1/4}$	$EL_{CH}$	0.8940	0.9474	0.9883	
	$EL_{CHB}$	0.9007	0.9504	0.9891	
	$EL_{ZJ}$	0.9622	0.9864	0.9988	
$n = 50$	$h$	EL	0.8848	0.9361	0.9941
$n^{-1}$	$EL_{CH}$	0.8959	0.9461	0.9905	
	$EL_{CHB}$	0.8973	0.9476	0.9910	
	$EL_{ZJ}$	0.8987	0.9486	0.9913	
$n^{-3/4}$	$EL_{CH}$	0.8982	0.9492	0.9902	
	$EL_{CHB}$	0.9002	0.9500	0.9907	
	$EL_{ZJ}$	0.9067	0.9536	0.9920	
$n^{-1/2}$	$EL_{CH}$	0.8988	0.9480	0.9903	
	$EL_{CHB}$	0.9005	0.9496	0.9905	
	$EL_{ZJ}$	0.9208	0.9639	0.9939	
$n^{-1/4}$	$EL_{CH}$	0.8958	0.9484	0.9897	
	$EL_{CHB}$	0.8984	0.9503	0.9901	
	$EL_{ZJ}$	0.9537	0.9829	0.9982	

Tabulā 8 ir redzami simulāciju rezultāti  $N(0, 1)$  sadalījuma mediānai pie nozīmības līmeņiem  $\alpha = 0.05, 0.1,$  un  $0.01.$  Līdzīgi kā gadījumā ar asimetrisko  $\chi_3^2$  sadalījumu, jāsecina, ka arī standartnormālā sadalījuma mediānai visstabilākos rezultātos uzrāda Chen un Hall metode, jo tās rezultātus  $h$  izmaņas ietekmē minimāli, ko nevar teikt par Zhou un Jing metodi.

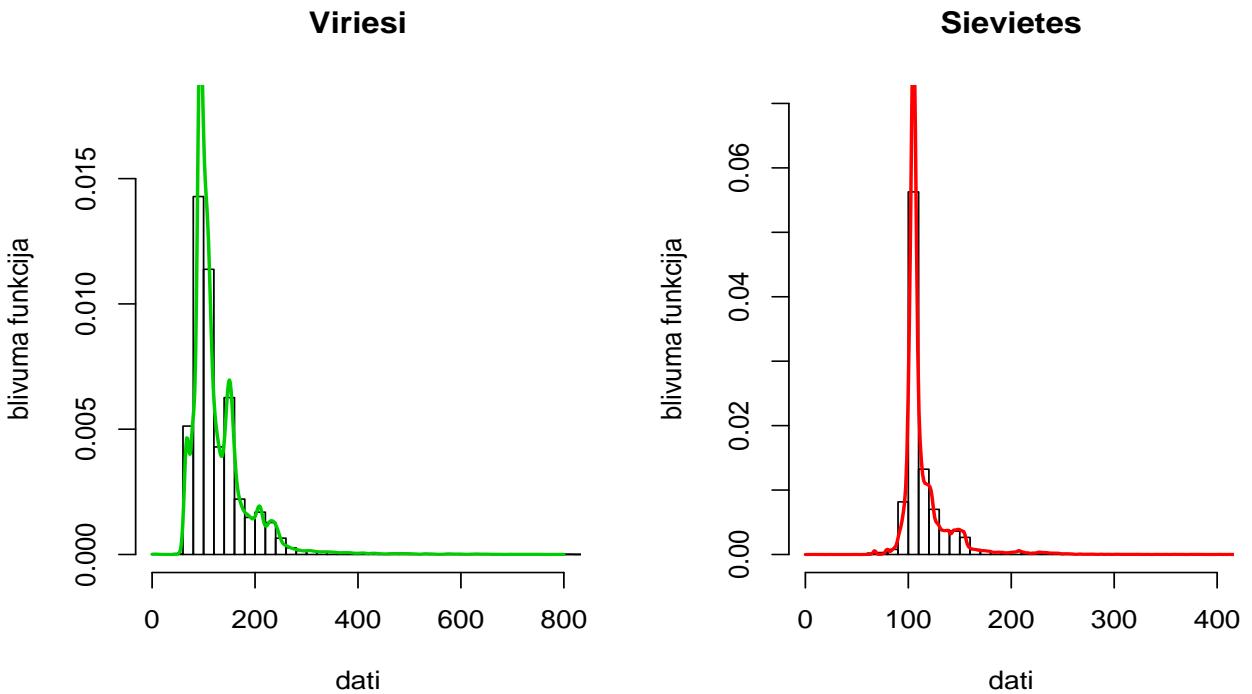
## 8 Pielietojumi

Šajā nodalā pielietosim empiriskās tīcamības metodi reālai datu problēmai—Latvijas pensiju vidējo vērtību izpētei. Tieki apskatīts viens no pensiju veidiem, šajā gadījumā vecuma pensija, kas piešķirta pēc vispārīgiem noteikumiem, t.i., persona ir sasniegusi likumā noteikto vecumu un viņas apdrošināšanas štāzs ir vismaz 10 gadi un nekādu citu nosacījumu, lai piešķirtu vecuma pensiju nav bijis. Šobrīd, lai pieteiktos uz vecuma pensiju nepieciešamais vecums, vīriešiem ir 62 gadi un sievietēm līdz 2008. gada 1. jūlijam 61.5, bet pēc tam 62 gadi. Jānovērtē vidējais pensijas lielums šim pensiju veidam vīriešiem un sievietēm. Dati ir aktuāli uz 2008. gada februāri un datu kopa sastāv no 92201 ierakstiem (52367-sievietes, 39834-vīrieši). Lai noskaidrotu, kādu metodi var pielietot tīcamības intervālu (divpusējo) konstruēšanā, vispirms jācenšas noskaidrot datu sadalījums, jāpārbauda, vai tie ir normāli sadalīti.



**Attēls 6:** QQ-grafiki

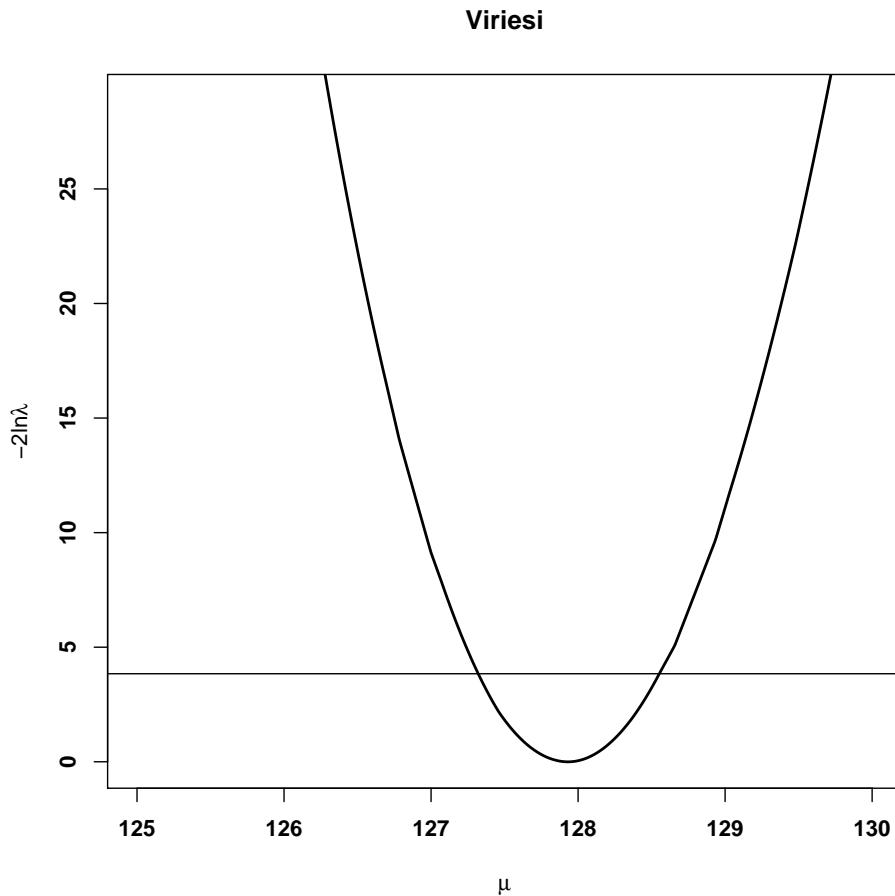
Attēlā 6 redzamie QQ-grafiki, liecina par to, ka analizējamajiem datiem nav normālais sadalījums. Arī, veicot datu normalitātes pārbaudi ar Kolmogorova-Smirnova testa palīdzību, jāsecina, ka, tā kā p-vērtība abos gadījumos ir mazāka par  $2.2e-16$ , tad, pie nozīmības līmeņa  $\alpha = 0.05$ , jānoraida hipotēze par analizējamo datu normalitāti.



**Attēls 7:** Neparametriskā blīvuma funkcijas novērtēšana

Attēlā 7 redzamas abu gadījumu histogrammas ar blīvuma funkciju novērtējumiem, kas iegūti, izmantojot kodolu gludināšanas metodi. Aplūkojot abu gadījumu blīvuma funkciju novērtējumus, jāsecina, ka abos gadījumos datiem ir asimetrisks sadalījums un sadalījumi samērā atšķiras savā starpā. Izvirzu hipotēzi par to, ka datiem varētu būt  $\chi^2$  sadalījums. Hipotēžu pārbaudē, izmantosim  $\chi^2$  kritēriju, programma, ar kurās palīdzību tas tika realizēts, ir pielikumā. Tā kā  $\chi^2$  vidējā vērtība sakrīt ar brīvības pakāpju skaitu, tad, veicot hipotēžu pārbaudi, brīvības pakāpju skaits tiks novērtēts ar neparametisko videjās vērtības novērtējumu. Abos gadījumos pie nozīmības līmeņa  $\alpha = 0.05$ , jānoraida hipotēze par to, ka analizējamajiem datiem ir  $\chi^2$  sadalījums. Tā kā teorētisko sadalījumu šajā gadījumā ir grūti noteikt, tad vidējās vērtības novērtēšanā, izmantosim EL metodi.

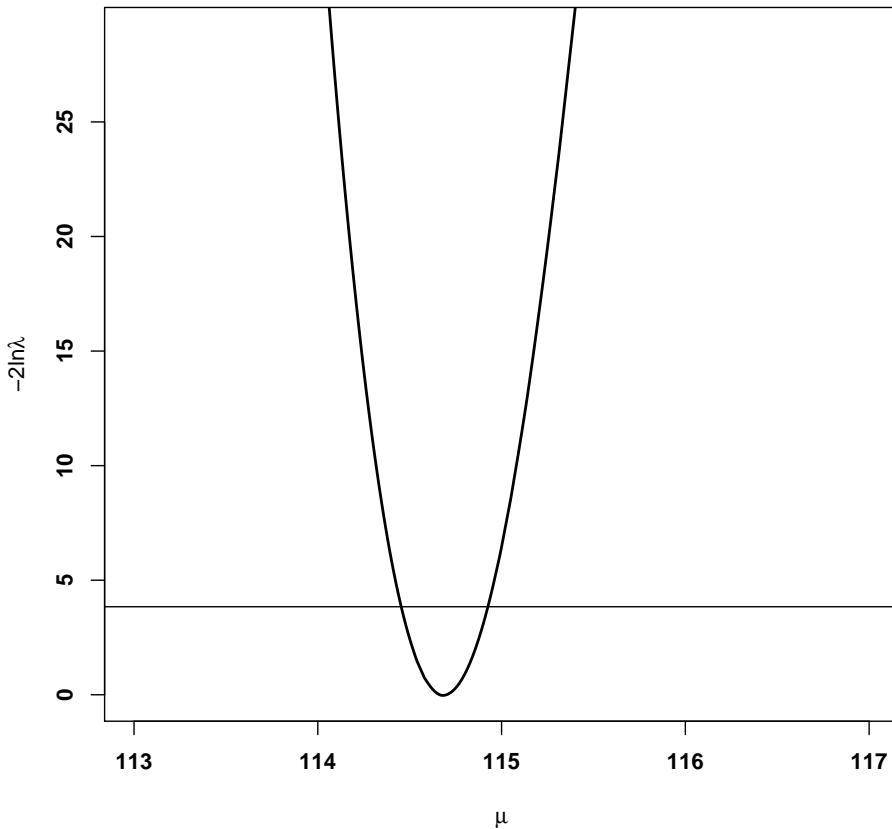
Konstruēsim ticamības intervālu vidējai vērtībai pie nozīmības līmeņa  $\alpha = 0.05$ .



**Attēls 8:**  $-2 \times \ln$  ticamības līkne vidējai vērtībai,  $\alpha = 0.05$ .

Attēlā 8 redzama logaritmiskās empīriskās attiecības statistikas līkne un EL metodes noteiktais ticamības intervāls videjai vērtībai ir  $[127.33; 128.55]$  un visticamākā  $\mu$  vērtība ir 127.93. Konstruējot, ticamības intervālu videjai vērtībai ar t-testa palīdzību, jāsecina, ka iegūtais ticamības intervāls  $[127.3180; 128.5427]$  praktiski sakrīt ar EL, taču ir nedaudz plašāks kā EL ticamības intervāls.

### Sievietes



**Attēls 9:**  $-2 \times \ln$  ticamības līkne vidējai vērtībai,  $\alpha = 0.05$ .

Attēlā 9 redzama logaritmiskās empīriskās attiecības statistikas līkne un EL metodes noteiktais ticamības intervāls videjai vērtībai ir  $[114.46; 114.92]$  un visticamākā  $\mu$  vērtība ir 114.68. Arī šajā gadījumā t-testa ticamības intervāls  $[114.4494; 114.9194]$  ir tikai nedaudz plašāks par EL ticamības intervālu.

## 9 Secinājumi

Darbā tika apskatīta empīriskās ticamības metode, kas tika pielietota parametru novērtēšanā un ticamības intervālu konstruēšanā. Simulējot dažādu sadalījumu datus, vidējai vērtībai un mediānai tika konstruēti ticamības intervāli un, izmantojot pārklājumu precizitāti, tika analizēta EL metodes efektivitāte. Viena no galvenajām EL ticamības intervālu priekšrocībām ir tāda, ka tie nav simetriiski, tie balstās uz analizējamo datu kopu. Simulācijās iegūtie rezultāti liecina, ka pie nelieliem izlašu apjomiem ar EL metodi konstruētie ticamības intervāli vidējai vērtībai nav tik precīzi kā uz parametriskajiem testiem balstītie ticamības intervāli, ja pareizi noteikts datu sadalījums. Taču jau pie  $n > 20$  EL tests strādā gandrīz tikpat labi kā parametriskie testi, pie tam, izpildoties zināmiem nosacījumiem, var pielietot Bartleta korekciju, kas vēl uzlabo EL ticamības intervālu pārklājuma precizitāti. Analizējot, kā EL metode strādā salīdzinājumā ar parametriskajām metodēm, ja datu sadalījums ir noteikts nepareizi, jāsecina, ka EL metode strādā labi un atsevišķos gadījumos pat vislabāk, taču nevar noliegt faktu, ka arī t–tests strādā ļoti labi. Lai arī visplašāk EL metodi izmanto tieši vidējai vērtībai, to var izmantot arī ticamības intervālu konstruēšanā kvantilēm.

Konstruējot ticamības intervālus mediānai pie nelieliem izlašu apjomiem, tika salīdzinātas 3 empīriskās ticamības metodes, no kurām divas izmanto kodolu gludināšanas metodi, kurā ļoti liela nozīme ir joslas platuma  $h$  izvēlei. Salīdzinot pārklājuma precizitāti pie dažādiem  $h$ , jāsecina, ka visstabilākos rezultātus uzrāda Chen un Hall [7] metode, jo gan simetriska, gan asimetriska sadalījuma gadījumā  $h$  izmaiņas minimāli ietekmē pārklājuma precizitāti, pie tam šai metodei ir iespējama arī Bartleta korekcija, kas vēl uzlabo pārklājuma precizitāti.

Viens no darba uzdevumiem bija pielietot EL metodi reālai problemātikai. Šajā gadījumā tika veikta Latvijas pensiju vidējo vērtību izpēte. Vienam pensiju veidam tika konstruēti ticamības intervāli vidējai vērtībai pie nozīmības līmeņa  $\alpha = 0.05$ . Ticamības intervāli tika konstruēti atsevišķi vīriešiem un sievietēm, tā kā teorētisko datu sadalījumu nevarēja noteikt, tad tika izmantota EL metode. Tā kā pētāmās datu kopas apjoms ir liels, tad ticamības intervāli ir šauri un visticamākie empīriskās ticamības vidējās vērtības novērtējumi ir: vīriešiem – 127.93Ls un sievietēm – 114.68Ls. Jāpiepilst, ka ticamības intervāli, kas šajā gadījumā konstruēti, izmantojot t-testu, minimāli atšķiras no EL ticamības intrevāliem.

EL metode salīdzinoši ir ļoti jauna metode un ķemot vērā to, ka neparametriskās metodes klūst arvien populārākas, jo praksē sastopamajām datu kopām ļoti bieži ir grūti noteikt teorētisko sadalījuma likumu, tad šīs metodes attīstība tikai turpināsies.

EL metode nav vienīgā neparametriskā metode, ko var lietot ticamības intervālu konstruēšanā

un parametru novērtēšanā. ļoti populāras tagad ir arī butstrapa metodes, tāpēc šo darbu tālāk varētu attīstīt, izpētot butstrapa metodes un salīdzinot tās ar EL metodi.

## 10 Izmantotā literatūra

1. Wasserman, L. All of statistics A Concise Course in Statistical inference. Springer, 2004. 442 p.
2. Owen, A. B. Empirical likelihood. Chapman & Hall/CRC, 2001. 304 p.
3. Qin, J. and Lawless, J. Empirical likelihood and general estimating equations. *The Annals of Statistics*, 1994, Vol. 22, No. 1, p. 300-325.
4. Hall, P. and La Scala, B. Methodology and algorithms of empirical likelihood. *International Statistical Review*, 1990, Vol. 58, No. 2, p. 109-127.
5. Diciccio, T., Hall, P. and Romano, J. Empirical likelihood is Bartlett-correctable. *The Annals of Statistics*, 1991, Vol. 19, No. 2, p. 1053-1061.
6. Diciccio, T., Hall, P., and Romano, J. Comparison of parametric and empirical likelihood functions. *Biometrika*, 1989, Vol. 76, No. 3, p. 465-476.
7. Chen, S. and Hall, P. Smoothed empirical likelihood confidence intervals for quantiles. *The Annals of Statistics*, 1993, Vol. 21, No. 3, p. 1166-1181.
8. Zhou, W. and Jing, B. Adjusted empirical likelihood method for quantiles. *Ann. Inst. Statist. Math.*, 2003, Vol. 55, No. 4, p. 689-703.
9. Adimari, G. An empirical likelihood statistic for quantiles. *J. Statist. Comput. Simul.*, 1998, Vol. 60, p. 85-95.
10. Owen, A.B. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 1988, Vol. 75, No. 2, p. 237-249.
11. Owen, A.B. Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 1990, Vol. 18, No. 1, p. 90-120.
12. Owen, A.B. Empirical likelihood for linear models. *The Annals of Statistics*, 1991, Vol. 19, No. 2, p. 1725-1747.
13. <http://instruct1.cit.cornell.edu/courses/econ620/reviewm5.pdf>, p. 4
14. <http://www.math.umd.edu/~evs/s701/WilksThm.pdf>, p. 4
15. <http://www.math.montana.edu/~rjboik/classes/502/notes.424.03.pdf>, p. 164

16. Diciccio, T., Hall, P. and Romano, J. Empirical likelihood is Bartlett-correctable. Unpublished manuscript, 1988
17. Diciccio, T. and Romano, J. On adjustments based on the signed root of the empirical likelihood ratio statistic. *Biometrika*, 1989, Vol. 74, p. 481-494.
18. Barndorff-Nielsen, O.E. and Hall P. On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika*, 1988, Vol. 75, p. 374-378
19. Thomas, D. R. and Grunkemeier, G. L. Confidence interval estimation of survival probabilities for censored data. *J. Am. Statist. Assoc.*, 1975, Vol. 70, p. 865-871.
20. Buiķis, M. un Siliņa, B. Matemātika. Definīcijas, formulas, aprēķinu algoritmi. Zvaigzne ABC, 1997, 288 lapas.
21. Hall, P. The bootstrap and edgeworth expansion. Springer-Verlag, 1992. 352 p.

## A Pielikums

### A.1 Programma simulāciju veikšanai vidējai vērtībai

```
#Bartleta korekcija Normalajam sadalijumam N(0,1) ar teoretisko un noverteto a  
el_1<-0  
el_bt_1<-0  
el_be_1<-0  
  
el_2<-0  
el_bt_2<-0  
el_be_2<-0  
  
el_3<-0  
el_bt_3<-0  
el_be_3<-0  
  
alfa1<-0.1  
alfa2<-0.05  
alfa3<-0.01  
  
n<-50  
m<-10000  
m1<-0  
a<-3/2  
  
for (i in 1:m)  
{  
X<-rnorm(n,mean=0,sd=1)  
X<-sort(X)  
  
a1<-round(X[1],2)  
b1<-round(X[n],2)  
if (a1<X[1]) {a1<-a1+0.01}
```

```

if (b1>X[n]) {b1<-b1-0.01}

disp_2<-sum((X-mean(X))^2)/n
mi_3<-(1/n)*sum((X-mean(X))^3)/((disp_2)^(3/2))
mi_4<-(1/n)*sum((X-mean(X))^4)/((disp_2)^2)
a_nov<-(1/2)*mi_4-(1/3)*(mi_3)^2

#EMLR
r<-c()
r11<-c()
r12<-c()
r13<-c()
r21<-c()
r22<-c()
r23<-c()
r31<-c()
r32<-c()
r33<-c()
Z<-seq(a1,b1,by=0.01)
k<-length(Z)
kritiska_vert1<-qchisq(1-alfa1, 1)
kritiska_vert2<-qchisq(1-alfa2, 1)
kritiska_vert3<-qchisq(1-alfa3, 1)

for (v in 1:k)
{
der<-0
f<-function (l) sum((X-Z[v])/(1+l*(X-Z[v])))
l1<-(1-1/n)/(Z[v]-X[n])
l2<-(1-1/n)/(Z[v]-X[1])
intervals<-c(l1,l2)
lambda<-uniroot(f, intervals, lower = min(intervals), upper = max(intervals),
tol = .Machine$double.eps^0.25, maxiter = 1000)

```

```

for (j in 1:n)

{if ((1+lambda[[1]]*(X[j]-Z[v]))>0) {der<-der+1}

if (der==n)

{

r[v]<-2*sum(log(1+lambda[[1]]*(X-Z[v])))

if (r[v]<=kritiska_vert1) {r11[v]<-Z[v]}

if (r[v]<=kritiska_vert1*(1+a/n)) {r12[v]<-Z[v]}

if (r[v]<=kritiska_vert1*(1+a_nov/n)) {r13[v]<-Z[v]}

if (r[v]<=kritiska_vert2) {r21[v]<-Z[v]}

if (r[v]<=kritiska_vert2*(1+a/n)) {r22[v]<-Z[v]}

if (r[v]<=kritiska_vert2*(1+a_nov/n)) {r23[v]<-Z[v]}

if (r[v]<=kritiska_vert3) {r31[v]<-Z[v]}

if (r[v]<=kritiska_vert3*(1+a/n)) {r32[v]<-Z[v]}

if (r[v]<=kritiska_vert3*(1+a_nov/n)) {r33[v]<-Z[v]}

}

}

if (length(r11)>0) {if (0>=na.exclude(r11)[1]& 0<=r11[length(r11)]) {el_1<-el_1+1}

if (length(r12)>0) {if (0>=na.exclude(r12)[1]& 0<=r12[length(r12)]) {el_bt_1<-el_bt_1+1}

if (length(r13)>0) {if (0>=na.exclude(r13)[1]& 0<=r13[length(r13)]) {el_be_1<-el_be_1+1}

if (length(r21)>0) {if (0>=na.exclude(r21)[1]& 0<=r21[length(r21)]) {el_2<-el_2+1}

if (length(r22)>0) {if (0>=na.exclude(r22)[1]& 0<=r22[length(r22)]) {el_bt_2<-el_bt_2+1}

if (length(r23)>0) {if (0>=na.exclude(r23)[1]& 0<=r23[length(r23)]) {el_be_2<-el_be_2+1}

if (length(r31)>0) {if (0>=na.exclude(r31)[1]& 0<=r31[length(r31)]) {el_3<-el_3+1}

if (length(r32)>0) {if (0>=na.exclude(r32)[1]& 0<=r32[length(r32)]) {el_bt_3<-el_bt_3+1}

if (length(r33)>0) {if (0>=na.exclude(r33)[1]& 0<=r33[length(r33)]) {el_be_3<-el_be_3+1}

m1<-m1+1

}

n

m

1-alfa1

el_1/m

el_bt_1/m

```

el\_be\_1/m

1-alfa2

el\_2/m

el\_bt\_2/m

el\_be\_2/m

1-alfa3

el\_3/m

el\_bt\_3/m

el\_be\_3/m

m1

## A.2 Programma simulāciju veikšanai mediānai

```
#Ticamibas intervali kvantilem (normalais sadalijums)
m<-10000
met_111<-0; met_112<-0; met_113<-0;
met_121<-0; met_122<-0; met_123<-0;
met_21111<-0; met_21112<-0; met_21121<-0; met_21122<-0; met_21131<-0; met_21132<-0;
met_22211<-0; met_22212<-0; met_22221<-0; met_22222<-0; met_22231<-0; met_22232<-0;
met_23311<-0; met_23312<-0; met_23321<-0; met_23322<-0; met_23331<-0; met_23332<-0;
met_24411<-0; met_24412<-0; met_24421<-0; met_24422<-0; met_24431<-0; met_24432<-0;
met_311<-0; met_312<-0; met_313<-0; met_321<-0; met_322<-0; met_323<-0;
met_331<-0; met_332<-0; met_333<-0; met_341<-0; met_342<-0; met_343<-0;

m1<-0
n<-50
alpha1<-0.10
alpha2<-0.05
alpha3<-0.01
med<-0
kritiska_vert1<-qchisq(1-alpha1,1)
```

```

kritiska_vert2<-qchisq(1-alpha2,1)
kritiska_vert3<-qchisq(1-alpha3,1)
q<-0.5 #mediana
for (y in 1:m)
{
X<-rnorm(n)
X<-sort(X)
a<-round(X[1],2)
b<-round(X[n],2)
if (a<=X[1]) {a<-a+0.01}
if (b>=X[n]) {b<-b-0.01}
Z<-seq(a,b,by=0.01)
k<-length(Z)

#sign-test
d1<-qbinom(alpha1,n,q,lower.tail=TRUE,log.p=FALSE)
d2<-qbinom(alpha2,n,q,lower.tail=TRUE,log.p=FALSE)
d3<-qbinom(alpha3,n,q,lower.tail=TRUE,log.p=FALSE)
I1_ap1<-X[d1]; I1_aug1<-X[n-d1+1];
I1_ap2<-X[d2]; I1_aug2<-X[n-d2+1];
I1_ap3<-X[d3]; I1_aug3<-X[n-d3+1];
if (med>=I1_ap1&med<=I1_aug1) {met_111<-met_111+1}
if (med>=I1_ap2&med<=I1_aug2) {met_112<-met_112+1}
if (med>=I1_ap3&med<=I1_aug3) {met_113<-met_113+1}

#EL bez gludinasanas
s<-c(); Fn<-c(); kriterija_vert1<-c(); g1<-c(); g2<-c(); g3<-c()

for (j in 1:k)
{
s[j]<-0
for (i in 1:n) {if (X[i]<=Z[j]) {s[j]<-s[j]+1}}
if(s[j]!>0)
}

```

```

{

Fn[j]<-s[j]/n

kriterija_vert1[j]<-2*n*(Fn[j]*log(Fn[j]/q)+(1-Fn[j])*log((1-Fn[j])/(1-q)))
if (kriterija_vert1[j]<=kritiska_vert1) {g1[j]<-Z[j]}
if (kriterija_vert1[j]<=kritiska_vert2) {g2[j]<-Z[j]}
if (kriterija_vert1[j]<=kritiska_vert3) {g3[j]<-Z[j]}

}

}

if (length(g1)>0) {if (med>=na.exclude(g1)[1]&med<=g1[length(g1)]) {met_121<-met_121+1}}
if (length(g2)>0) {if (med>=na.exclude(g2)[1]&med<=g2[length(g2)]) {met_122<-met_122+1}}
if (length(g3)>0) {if (med>=na.exclude(g3)[1]&med<=g3[length(g3)]) {met_123<-met_123+1}}



```

```

library(sm)

#EL ar gludinasanu (Chen&Hall)

#h<-bw.nrd0(X)#rule-of-thumb

h1<-n^(-1); h2<-n^(-3/4); h3<-n^(-1/2); h4<-n^(-1/4);

mi_21<-(1/n)*sum((pnorm((median(X)-X)/h1)-q)^2);
mi_22<-(1/n)*sum((pnorm((median(X)-X)/h2)-q)^2);
mi_23<-(1/n)*sum((pnorm((median(X)-X)/h3)-q)^2);
mi_24<-(1/n)*sum((pnorm((median(X)-X)/h4)-q)^2);
mi_31<-(1/n)*sum((pnorm((median(X)-X)/h1)-q)^3);
mi_32<-(1/n)*sum((pnorm((median(X)-X)/h2)-q)^3);
mi_33<-(1/n)*sum((pnorm((median(X)-X)/h3)-q)^3);
mi_34<-(1/n)*sum((pnorm((median(X)-X)/h4)-q)^3);
mi_41<-(1/n)*sum((pnorm((median(X)-X)/h1)-q)^4);
mi_42<-(1/n)*sum((pnorm((median(X)-X)/h2)-q)^4);
mi_43<-(1/n)*sum((pnorm((median(X)-X)/h3)-q)^4);
mi_44<-(1/n)*sum((pnorm((median(X)-X)/h4)-q)^4);

beta_nov1<-(1/6)*(3*((mi_21)^(-2))*mi_41-2*((mi_21)^(-3))*(mi_31)^2);
beta_nov2<-(1/6)*(3*((mi_22)^(-2))*mi_42-2*((mi_22)^(-3))*(mi_32)^2);
beta_nov3<-(1/6)*(3*((mi_23)^(-2))*mi_43-2*((mi_23)^(-3))*(mi_33)^2);
beta_nov4<-(1/6)*(3*((mi_24)^(-2))*mi_44-2*((mi_24)^(-3))*(mi_34)^2);


```

```

r11<-c(); r12<-c(); r13<-c(); rb11<-c(); rb12<-c(); rb13<-c();
r21<-c(); r22<-c(); r23<-c(); rb21<-c(); rb22<-c(); rb23<-c();
r31<-c(); r32<-c(); r33<-c(); rb31<-c(); rb32<-c(); rb33<-c();
r41<-c(); r42<-c(); r43<-c(); rb41<-c(); rb42<-c(); rb43<-c();
w1<-c(); w2<-c(); w3<-c(); w4<-c();
kriterija_vert21<-c(); kriterija_vert22<-c(); kriterija_vert23<-c();
kriterija_vert24<-c();
for (v in 1:k)
{
der1<-0; der2<-0; der3<-0; der4<-0;
w1<-pnorm((Z[v]-X)/h1)-q; w2<-pnorm((Z[v]-X)/h2)-q;
w3<-pnorm((Z[v]-X)/h3)-q; w4<-pnorm((Z[v]-X)/h4)-q;
f1<-function(l)sum(w1/(1+l*w1))
f2<-function(l)sum(w2/(1+l*w2))
f3<-function(l)sum(w3/(1+l*w3))
f4<-function(l)sum(w4/(1+l*w4))
l11<-(1-1/n)/(q-pnorm((Z[v]-X[n])/h1));
l21<-(1-1/n)/(q-pnorm((Z[v]-X[1])/h1));
l12<-(1-1/n)/(q-pnorm((Z[v]-X[n])/h2));
l22<-(1-1/n)/(q-pnorm((Z[v]-X[1])/h2));
l13<-(1-1/n)/(q-pnorm((Z[v]-X[n])/h3));
l23<-(1-1/n)/(q-pnorm((Z[v]-X[1])/h3));
l14<-(1-1/n)/(q-pnorm((Z[v]-X[n])/h4));
l24<-(1-1/n)/(q-pnorm((Z[v]-X[1])/h4));
intervals1<-sort(c(l11,l21)); intervals2<-sort(c(l12,l22));
intervals3<-sort(c(l13,l23)); intervals4<-sort(c(l14,l24));
lambda1<-uniroot(f1, intervals1, lower = min(intervals1), upper = max(intervals1))
lambda2<-uniroot(f2, intervals2, lower = min(intervals2), upper = max(intervals2))
lambda3<-uniroot(f3, intervals3, lower = min(intervals3), upper = max(intervals3))
lambda4<-uniroot(f4, intervals4, lower = min(intervals4), upper = max(intervals4))
for (t in 1:n)
{if ((1+lambda1[[1]]*w1[t])>0) {der1<-der1+1}}
if (der1==n)

```

```

{

kriterija_vert21[v]<-2*sum(log(1+lambda1[[1]]*w1))

if (kriterija_vert21[v]<=kritiska_vert1) {r11[v]<-Z[v]}

if (kriterija_vert21[v]<=kritiska_vert2) {r12[v]<-Z[v]}

if (kriterija_vert21[v]<=kritiska_vert3) {r13[v]<-Z[v]}

if (kriterija_vert21[v]<=kritiska_vert1*(1+beta_nov1/n)) {rb11[v]<-Z[v]}

if (kriterija_vert21[v]<=kritiska_vert2*(1+beta_nov1/n)) {rb12[v]<-Z[v]}

if (kriterija_vert21[v]<=kritiska_vert3*(1+beta_nov1/n)) {rb13[v]<-Z[v]}

}

for (t in 1:n)

{if ((1+lambda2[[1]]*w2[t])>0) {der2<-der2+1}

if (der2==n)

{

kriterija_vert22[v]<-2*sum(log(1+lambda2[[1]]*w2))

if (kriterija_vert22[v]<=kritiska_vert1) {r21[v]<-Z[v]}

if (kriterija_vert22[v]<=kritiska_vert2) {r22[v]<-Z[v]}

if (kriterija_vert22[v]<=kritiska_vert3) {r23[v]<-Z[v]}

if (kriterija_vert22[v]<=kritiska_vert1*(1+beta_nov2/n)) {rb21[v]<-Z[v]}

if (kriterija_vert22[v]<=kritiska_vert2*(1+beta_nov2/n)) {rb22[v]<-Z[v]}

if (kriterija_vert22[v]<=kritiska_vert3*(1+beta_nov2/n)) {rb23[v]<-Z[v]}

}

for (t in 1:n)

{if ((1+lambda3[[1]]*w3[t])>0) {der3<-der3+1}

if (der3==n)

{

kriterija_vert23[v]<-2*sum(log(1+lambda3[[1]]*w3))

if (kriterija_vert23[v]<=kritiska_vert1) {r31[v]<-Z[v]}

if (kriterija_vert23[v]<=kritiska_vert2) {r32[v]<-Z[v]}

if (kriterija_vert23[v]<=kritiska_vert3) {r33[v]<-Z[v]}

if (kriterija_vert23[v]<=kritiska_vert1*(1+beta_nov3/n)) {rb31[v]<-Z[v]}

if (kriterija_vert23[v]<=kritiska_vert2*(1+beta_nov3/n)) {rb32[v]<-Z[v]}

}
}
}

```

```

if (kriterija_vert23[v]<=kritiska_vert3*(1+beta_nov3/n)) {rb33[v]<-Z[v]}

}

for (t in 1:n)
{if ((1+lambda4[[1]]*w4[t])>0) {der4<-der4+1}}
if (der4==n)
{
kriterija_vert24[v]<-2*sum(log(1+lambda4[[1]]*w4))
if (kriterija_vert24[v]<=kritiska_vert1) {r41[v]<-Z[v]}
if (kriterija_vert24[v]<=kritiska_vert2) {r42[v]<-Z[v]}
if (kriterija_vert24[v]<=kritiska_vert3) {r43[v]<-Z[v]}
if (kriterija_vert24[v]<=kritiska_vert1*(1+beta_nov4/n)) {rb41[v]<-Z[v]}
if (kriterija_vert24[v]<=kritiska_vert2*(1+beta_nov4/n)) {rb42[v]<-Z[v]}
if (kriterija_vert24[v]<=kritiska_vert3*(1+beta_nov4/n)) {rb43[v]<-Z[v]}
}

}

if (length(r11)>0) {if (med>=na.exclude(r11)[1]&med<=r11[length(r11)])
{met_21111<-met_21111+1}}
if (length(rb11)>0) {if (med>=na.exclude(rb11)[1]&med<=rb11[length(rb11)])
{met_21112<-met_21112+1}}
if (length(r12)>0) {if (med>=na.exclude(r12)[1]&med<=r12[length(r12)])
{met_21121<-met_21121+1}}
if (length(rb12)>0) {if (med>=na.exclude(rb12)[1]&med<=rb12[length(rb12)])
{met_21122<-met_21122+1}}
if (length(r13)>0) {if (med>=na.exclude(r13)[1]&med<=r13[length(r13)])
{met_21131<-met_21131+1}}
if (length(rb13)>0) {if (med>=na.exclude(rb13)[1]&med<=rb13[length(rb13)])
{met_21132<-met_21132+1}>

}

if (length(r21)>0) {if (med>=na.exclude(r21)[1]&med<=r21[length(r21)])
{met_22211<-met_22211+1}}
if (length(rb21)>0) {if (med>=na.exclude(rb21)[1]&med<=rb21[length(rb21)])

```

```

{met_22212<-met_22212+1)}

if (length(r22)>0) {if (med>=na.exclude(r22)[1]&med<=r22[length(r22)])}
{met_22221<-met_22221+1)}

if (length(rb22)>0) {if (med>=na.exclude(rb22)[1]&med<=rb22[length(rb22)])}
{met_22222<-met_22222+1)}

if (length(r23)>0) {if (med>=na.exclude(r23)[1]&med<=r23[length(r23)])}
{met_22231<-met_22231+1)}

if (length(rb23)>0) {if (med>=na.exclude(rb23)[1]&med<=rb23[length(rb23)])}
{met_22232<-met_22232+1)}

if (length(r31)>0) {if (med>=na.exclude(r31)[1]&med<=r31[length(r31)])}
{met_23311<-met_23311+1)}

if (length(rb31)>0) {if (med>=na.exclude(rb31)[1]&med<=rb31[length(rb31)])}
{met_23312<-met_23312+1)}

if (length(r32)>0) {if (med>=na.exclude(r32)[1]&med<=r32[length(r32)])}
{met_23321<-met_23321+1)}

if (length(rb32)>0) {if (med>=na.exclude(rb32)[1]&med<=rb32[length(rb32)])}
{met_23322<-met_23322+1)}

if (length(r33)>0) {if (med>=na.exclude(r33)[1]&med<=r33[length(r33)])}
{met_23331<-met_23331+1)}

if (length(rb33)>0) {if (med>=na.exclude(rb33)[1]&med<=rb33[length(rb33)])}
{met_23332<-met_23332+1)}

if (length(r41)>0) {if (med>=na.exclude(r41)[1]&med<=r41[length(r41)])}
{met_24411<-met_24411+1)}

if (length(rb41)>0) {if (med>=na.exclude(rb41)[1]&med<=rb41[length(rb41)])}
{met_24412<-met_24412+1)}

if (length(r42)>0) {if (med>=na.exclude(r42)[1]&med<=r42[length(r42)])}
{met_24421<-met_24421+1)}

if (length(rb42)>0) {if (med>=na.exclude(rb42)[1]&med<=rb42[length(rb42)])}
{met_24422<-met_24422+1)}

if (length(r43)>0) {if (med>=na.exclude(r43)[1]&med<=r43[length(r43)])}
{met_24431<-met_24431+1)}

```

```

if (length(rb43)>0) {if (med>=na.exclude(rb43)[1]&med<=rb43[length(rb43)])
{met_24432<-met_24432+1} }

#EL ar gludinasanu (Zhou&Jing)
kriterija_vert31<-c(); kriterija_vert32<-c(); kriterija_vert33<-c();
kriterija_vert34<-c();
Fn_nov1<-c(); Fn_nov2<-c(); Fn_nov3<-c(); Fn_nov4<-c();
u311<-c(); u312<-c(); u313<-c(); u321<-c(); u322<-c(); u323<-c();
u331<-c(); u332<-c(); u333<-c(); u341<-c(); u342<-c(); u343<-c();
for (d in 1:k)
{
Fn_nov1[d]<-(1/n)*sum(pnorm((Z[d]-X)/h1))
Fn_nov2[d]<-(1/n)*sum(pnorm((Z[d]-X)/h2))
Fn_nov3[d]<-(1/n)*sum(pnorm((Z[d]-X)/h3))
Fn_nov4[d]<-(1/n)*sum(pnorm((Z[d]-X)/h4))
kriterija_vert31[d]<-2*n*(Fn_nov1[d]*log(Fn_nov1[d]/q)+(1-Fn_nov1[d])*
log((1-Fn_nov1[d])/(1-q)))
kriterija_vert32[d]<-2*n*(Fn_nov2[d]*log(Fn_nov2[d]/q)+(1-Fn_nov2[d])*
log((1-Fn_nov2[d])/(1-q)))
kriterija_vert33[d]<-2*n*(Fn_nov3[d]*log(Fn_nov3[d]/q)+(1-Fn_nov3[d])*
log((1-Fn_nov3[d])/(1-q)))
kriterija_vert34[d]<-2*n*(Fn_nov4[d]*log(Fn_nov4[d]/q)+(1-Fn_nov4[d])*
log((1-Fn_nov4[d])/(1-q)))
if (kriterija_vert31[d]<=kritiska_vert1) {u311[d]<-Z[d]}
if (kriterija_vert31[d]<=kritiska_vert2) {u312[d]<-Z[d]}
if (kriterija_vert31[d]<=kritiska_vert3) {u313[d]<-Z[d]}

if (kriterija_vert32[d]<=kritiska_vert1) {u321[d]<-Z[d]}
if (kriterija_vert32[d]<=kritiska_vert2) {u322[d]<-Z[d]}
if (kriterija_vert32[d]<=kritiska_vert3) {u323[d]<-Z[d]}

if (kriterija_vert33[d]<=kritiska_vert1) {u331[d]<-Z[d]}
if (kriterija_vert33[d]<=kritiska_vert2) {u332[d]<-Z[d]}

```

```

if (kriterija_vert33[d]<=kritiska_vert3) {u333[d]<-Z[d]}

if (kriterija_vert34[d]<=kritiska_vert1) {u341[d]<-Z[d]}
if (kriterija_vert34[d]<=kritiska_vert2) {u342[d]<-Z[d]}
if (kriterija_vert34[d]<=kritiska_vert3) {u343[d]<-Z[d]}
}

if (length(u311)>0) {if (med>=na.exclude(u311)[1]&med<=u311[length(u311)])
{met_311<-met_311+1}}
if (length(u312)>0) {if (med>=na.exclude(u312)[1]&med<=u312[length(u312)])
{met_312<-met_312+1}}
if (length(u313)>0) {if (med>=na.exclude(u313)[1]&med<=u313[length(u313)])
{met_313<-met_313+1}}


if (length(u321)>0) {if (med>=na.exclude(u321)[1]&med<=u321[length(u321)])
{met_321<-met_321+1}}
if (length(u322)>0) {if (med>=na.exclude(u322)[1]&med<=u322[length(u322)])
{met_322<-met_322+1}}
if (length(u323)>0) {if (med>=na.exclude(u323)[1]&med<=u323[length(u323)])
{met_323<-met_323+1}}


if (length(u331)>0) {if (med>=na.exclude(u331)[1]&med<=u331[length(u331)])
{met_331<-met_331+1}}
if (length(u332)>0) {if (med>=na.exclude(u332)[1]&med<=u332[length(u332)])
{met_332<-met_332+1}}
if (length(u333)>0) {if (med>=na.exclude(u333)[1]&med<=u333[length(u333)])
{met_333<-met_333+1}}


if (length(u341)>0) {if (med>=na.exclude(u341)[1]&med<=u341[length(u341)])
{met_341<-met_341+1}}
if (length(u342)>0) {if (med>=na.exclude(u342)[1]&med<=u342[length(u342)])
{met_342<-met_342+1}}
if (length(u343)>0) {if (med>=na.exclude(u343)[1]&med<=u343[length(u343)])
{met_343<-met_343+1}}

```

```

m1<-m1+1
}
n
m
#sign_test
met_111/m; met_112/m; met_113/m;
#EL_neglud
met_121/m; met_122/m; met_123/m;
#EL_glud_Chen
met_21111/m; met_21112/m; met_21121/m; met_21122/m; met_21131/m; met_21132/m;
met_22211/m; met_22212/m; met_22221/m; met_22222/m; met_22231/m; met_22232/m;
met_23311/m; met_23312/m; met_23321/m; met_23322/m; met_23331/m; met_23332/m;
met_24411/m; met_24412/m; met_24421/m; met_24422/m; met_24431/m; met_24432/m;
#EL_glud_Zhou
met_311/m; met_312/m; met_313/m; met_321/m; met_322/m; met_323/m;
met_331/m; met_332/m; met_333/m; met_341/m; met_342/m; met_343/m;

```

### A.3 Programma $\chi^2$ kritērija pārbaudei

```

#Hi kvadrata kriterijs, kas parbauda, vai datiem nav Hi kvadrata sadalijums
#VSAA dati
X<-read.table(file="viriesi.dat", header = FALSE)
Y<-read.table(file="sievietes.dat", header = FALSE)
X4<-X[[4]]
Y4<-Y[[4]]
n1<-length(X4)
n2<-length(Y4)
X4<-sort(X4)
Y4<-sort(Y4)
df1<-127.93
df2<-114.68

#Viriesi

```

```

r=n1/2000 # r<=n1/2000, tik dalas sadalam teoretisko blivuma funkcijas laukumu
quant<-qchisq((1:(r-1))/r, df1) # izrekinam attiecigas kvantiles
y<-c()
freq1<-c()
for (i in 1:length(quant)){
y[i]<-length(X4[X4<=quant[i]])
if (i==1){
freq1[i]<-y[i]
}
else{
freq1[i]<-y[i]-y[i-1]
}
freq1[i+1]<- n1-sum(freq1) # pedejais intervals
sum(freq1) # jabut vienadai ar n1

expected<-rep(n1/r,r)
observed<-freq1

sum((expected-observed)^2/expected)
qchisq(0.95,r-1)

#Sievietes
r=n2/3000 # r<=n2/3000, tik dalas sadalam teoretisko blivuma funkcijas laukumu
quant<-qchisq((1:(r-1))/r, df2) # izrekinam attiecigas kvantiles
y<-c()
freq1<-c()
for (i in 1:length(quant)){
y[i]<-length(Y4[Y4<=quant[i]])
if (i==1){
freq1[i]<-y[i]
}
else{
freq1[i]<-y[i]-y[i-1]
}
}

```

```
}

freq1[i+1] <- n2-sum(freq1) # pedejais intervals
sum(freq1) # ja but vienadai ar n2

expected<-rep(n2/r,r)
observed<-freq1

sum((expected-observed)^2/expected)
qchisq(0.95,r-1)
```

Diplomdarbs ”Empīriskās ticamības funkcijas metode un tās pielietojumi statistikā” izstrādāts  
LU Fizikas un matemātikas fakultātē.

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

Autors: Sandra Vucāne

Rekomendēju darbu aizstāvēšanai

Vadītājs: docents Dr. math. Jānis Valeinis

Recenzente: docente Dr. math. Nadežda Siļenko

Darbs iesniegts Matemātikas nodaļā 03.06.2008.

Metodiķe: Dzintra Holsta

Darbs aizstāvēts valsts pārbaudījuma komisijas sēdē

09.06.2008. prot. Nr. , vērtējums

Komisijas sekretāre: Ingrīda Uljane