

LATVIJAS UNIVERSITĀTE
FIZIKAS UN MATEMĀTIKAS FAKULTĀTE
MATEMĀTIKAS NODAĻA

**KVANTIĻU FUNKCIJU ANALĪZE AR PIELIETOJUMU
STRUKTURĀLO ATTIECĪBU MODEĻIEM**

DIPLOMDARBS

Autors: **Odeta Lorence**

Stud. apl. ol07001

Darba vadītājs: doc. Dr.math. Jānis Valeinis

RĪGA 2012

Anotācija

Darbs ir veltīts strukturālo attiecību modeļiem divu izlašu gadījumā. Tiek pētīts lokācijas modelis, kas tiek definēts kā $F_1(x) = F_2(x + \theta) \ \forall x \in R$, kur θ ir lokācijas parametrs. Darba mērķi ir noskaidrot vai eksistē lokācijas modelis, konstruējot vienlaicīgās ticamības joslas lokācijas parametram, kvantiļu funkcijas starpībām un varbūtību-varbūtību $(P - P)$, kvantiļu-kvantiļu $(Q - Q)$ grafikiem. Darbā aspkatīta kvantiļu funkcijas analīze vienas un divu izlašu gadījumā. Darbā tiek salīdzinātas dažādas pieejas, kā ar programmas R palīdzību tiek konstruēti ticamības intervāli.

Atslēgas vārdi: kvantiļu funkcija, strukturālie modeļi, lokācijas modelis, vienlaicīgās ticamības joslas

Abstract

This thesis contains analysis of structural relationship models in the two sample case. The location model has been studied, which is defined as $F_1(x) = F_2(x + \theta) \forall x \in R$, where θ is location parameter. The purpose of this work is to investigate, wheter the location model exist, constructing simultaneous confidence bands for the location parameter, quantile function differences and probability-probability ($P - P$) and quantile-quantile ($Q - Q$). Also we consider quantile function analysis in the one and two- sample cases. We compare different approaches, constructing confidence bounds for real data problem in the program R.

Keywords: quantile function, structural relationship models, simultaneous confidence bounds, location model

Saturs

Ievads	2
1. Strukturālo attiecību modeļi	4
2. Hipotēžu pārbaude fiksēta lokācijas parametra gadījumā	5
3. Vienlaicīgās ticamības joslas lokācijas parametram	7
4. Ticamības intervālu konstruēšana, izmantojot empīriskos procesus	10
4.1. Empīriskā procesa definīcija un piemēri	10
4.2. Empīriskie procesi kvantilēm vienas izlases gadījumā	10
4.3. Empīriskie procesi kvantilēm divu izlašu gadījumā	17
5. Ticamības intervālu konstruēšana, izmantojot empīriskās ticamības metodi	21
5.1. Empīriskās ticamības metode (EL) kvantilēm vienas izlases gadījumā	21
5.2. Empīriskās ticamības metode (EL) kvantilēm divu izlašu gadījumā .	28
6. Pielietojums reālu datu problēmām	29
Nobeigums	36
Izmantotā literatūra un avoti	37
Pielikums	39

Ievads

Dotas divas neatkarīgas izlases $X_1, \dots, X_m$ un $Y_1, \dots, Y_n$ ar attiecīgajām sadalījuma funkcijām F_1 un F_2 . Darbā tiek analizēti vispārēji strukturāli modeļi divu izlašu gadījumā, kas pirmo reizi definēti Freitag un Munk publikācijā [1]. Tie iekļauj divas problēmas - lokācijas-skalēšanas modeli un Lēmaņa alternatīvo modeli. Lokācijas skalēšanas modelis pārbauda vai divas izlases atšķiras ar lokācijas un skalēšanas parametriem

$$F_1(x) = F_2\left(\frac{x - \mu}{\sigma}\right), \quad x \in \mathcal{R}, \quad \mu, \sigma \in \mathcal{R}, \quad \sigma > 0.$$

Ja $\sigma = 1$, tad starp izlasēm pastāv šifta jeb lokācijas modelis, bet pastāv tikai skalēšanas modelis, ja $\mu = 0$. Lokācijas modelis tiek definēts kā $F_1(x) = F_2(x + \theta) \quad \forall x$, kur θ ir lokācijas parametrs, F_1 un F_2 ir attiecīgās sadalījuma funkcijas. Šādi modeļi ir svarīgi, piemēram, medicīnā, kur, ieviešot jaunas zāles, ir sagaidāms uzlabojums. Šajā gadījumā tiek pārbaudīta hipotēze par pašu lokācijas modeli, neinteresējoties par konkrēto sadalījuma funkciju veidu.

Eksistē vairākas iespējas, kā pārbaudīt hipotēzi par lokācijas modeli. Pirmkārt, pārbaudot vienkāršu hipotēzi par fiksētu parametru θ , otrai izlasei atņemot (vai pieskaitot) šo parametru var reducēt uz hipotēzi par lokācijas modeli, hipotēzi par divu sadalījuma funkciju vienādības pārbaudi, tas ir $H_0 : F_1(x) = F_2(x)$. Hipotēzi var pārbaudīt ar klasisko Kolmogorova - Statistikas testu vai konstruēt vienlaicīgās ticamības joslas varbūtību-varbūtību ($P-P$) un kvantiļu-kvantiļu ($Q-Q$) grafikiem. Lai pētītu šo pieeju pārbaudot vispārēju lokācijas modeli (parametram θ) θ sākumā ir jānovērtē. Šo pieeju ir pētījis Juris Cielēns savā diplomdarbā [2].

Viena no alternatīvām metodēm kā analizēt hipotēzi par lokācijas modeli ir aplūkot vienlaicīgās ticamības joslas pašam parametram θ , kā to ir skatījis Doksums un Sievers savā publikācijā [3]. Otra metode arī konstruēt divu izlašu kvantiļu funkciju starpību, kā to ir apskatījuši Petter Laake, Knut Laake, Rolf Aaberge savā publikācijā [4]. Abos šajos darbos tiek aplūkota statistika, kas balstīta uz empīrisku procesu tipa statistikām, kuru robežsadalījumi balstās uz Brauna tilta procesiem.

Darbā tiek apskatīta arī empīriskās ticamības metode (turpmāk EL). Tā ir konkurētspējīga neparametriskā metode, kas, balstoties uz dotajiem novērojumiem, konstruē ticamības intervālus, vai veic hipotēžu pārbaudi. Par tās izveidotāju tiek uzskatīts Owen [5], [6], kas ieviesa EL metodi izmantojot neparametrisko ticamības funkciju, bet šo meto-

di 2000. gadā vispārināja divu izlašu gadījumā Qin un Zhao [7] divu izlašu vidējo vērtību starpībai un sadalījuma funkciju starpībai. Jānis Valeinis doktora disertācijā [8] ir aprakstījis EL metodi kvantiļu funkciju starpībām.

Šī darba mērķis ir veikt pārbaudi vai eksistē lokācijas modelis. Lokācijas modeļa analīze tiek veikta konstruējot vienlaicīgos ticamības joslas lokācijas parametram, kvantiļu funkciju starpībai un $P - P$ un $Q - Q$ grafikiem. Darbā tiek apskatīta kvantiļu funkciju analīze vienas un divu izlašu gadījumā, izmantojot gan empīrisko procesu, gan empīrisko ticamības funkcijas metodes. Dažādas metodes tiks salīdzinātas ar simulāciju palīdzību, kā arī analizējot dažādas datu problēmas. Tiks piedāvāta uzlabota metode kvantiļu starpības analīzei divu izlašu gadījumā, izmantojot empīrisko procesu pieeju. Lai gan publikācijā [4] tiek konstruētas vienlaicīgās joslas gadījumā, kad viens sadalījums ir zināms, ir iespējams šo problēmu atrisināt pilnīgi neparametriski, kad abas sadalījuma funkcijas ir nezināmas.

Darbs sastāv no sešām nodaļām un pielikuma. Pirmajā nodaļā tiek sniegta strukturālo modeļu definīcija. Otrajā nodaļā tiek aprakstīta hipotēžu pārbaude par lokācijas modeļa eksistenci, fiksēta parametra θ gadījumā. Trešajā nodaļā analizēta hipotēze par lokācijas modeli, aplūkojot vienlaicīgās ticamības joslas pašam parametram θ . Ceturtajā nodaļā tiek aprakstīta ticamības intervālu konstruēšana, izmantojot empīriskos procesus vienas un divu izlašu gadījumā. Piektajā nodaļā tiek aprakstīta ticamības intervālu konstruēšana, izmantojot empīriskās ticamības metodi vienas un divu izlašu gadījumā. Sestā nodaļa satur reālu datu piemērus. Pielikumā atrodams programmas R kods.

1. Strukturālo attiecību modeļi

Strukturālo attiecību modeļus pirmo reizi savā publikācijā 2005. gadā aprakstīja vācu matemātiķi G. Freitag un A. Munk [1]. Tie iekļauj divas problēmas - lokācijas-skalēšanas modeli un Lēmaņa alternatīvo modeli.

Definīcija 1. Starp divu izlašu $X_1, \dots, X_m$ un $Y_1, \dots, Y_n$ sadalījuma funkcijām F_1 un F_2 pastāv lokācijas-skalēšanas modelis, ja

$$F_1(x) = F_2\left(\frac{x - \mu}{\sigma}\right), \quad x \in R \quad (1.1)$$

kur σ ir skalēšanas parametrs un μ ir lokācija spparametrs.

Šo attiecību var izteikt arī ar kvantiļu funkcijām.

$$F_1^{-1}(y) = \sigma F_2^{-1}(y) + \mu, \quad y \in [0, 1]. \quad (1.2)$$

Ja $\sigma = 1$, tad starp izlasēm pastāv šifta jeb lokācijas modelis, bet pastāv tikai skalēšanas modelis, ja $\mu = 0$.

Definīcija 2. Par empīrisko kvantiļu funkciju sauc funkciju

$$F_n^{-1}(y) = \inf\{x : F(x) \geq y\}, \quad 0 < y < 1. \quad (1.3)$$

Definīcija 3. Lokācijas modelis tiek definēts kā

$$F_1(x) = F_2(x + \theta) \quad \forall x, \quad (1.4)$$

kur θ ir lokācijas parametrs.

Definīcija 4. Starp divu izlašu $X_1, \dots, X_m$ un $Y_1, \dots, Y_n$ sadalījuma funkcijām F_1 un F_2 pastāv Lēmaņa alternatīvais modelis, ja

$$F_1(x) = 1 - (1 - F_2(x))^{1/h} \quad x \in R, \quad h > 0. \quad (1.5)$$

Izsakot ar kvantiļu funkcijām, iegūst

$$F_1^{-1}(y) = F_2^{-1}(1 - (1 - y)^h), \quad y \in [0, 1]. \quad (1.6)$$

Dotās izlases atbilst proporcionālā riska nosacījumam, ja starp tām pastāv Lēmaņa alternatīvais modelis. Tas tiek pielietots medicīniskajā statistikā un izdzīvošanas datu analizē.

Divu izlašu modeļi, kā (1.2) un (1.6), kas izteikti ar kvantiļu funkcijām, var tikt pārveidoti kā strukturālo attiecību modeļi vispārējā formā. Pieņemsim, ka sadalījuma funkcijas F_1 un F_2 ir elementi no funkcijas klases

$$\mathcal{F}^2 := \{F : F \text{ ir sadalījuma funkcija un } \int y^2 dF < \infty\}.$$

Definīcija 5. [1] Pieņemsim, ka $\mathcal{H} \subseteq R^l$ un $\phi_1 : R \times \mathcal{H} \rightarrow R$, $\phi_2 : [0, 1] \times \mathcal{H} \rightarrow [0, 1]$. Funkcijas F_1 un F_2 ir saistītas ar strukturālu attiecību, kuru veido ϕ_1 un ϕ_2 , ja $(F_1, F_2) \in \mathcal{U}_{\phi_1, \phi_2} =: \mathcal{U}$, kur modeļu klase $\mathcal{U}$ tiek definēta kā

$$\mathcal{U} := \{(F_1, F_2) \in \mathcal{F}^2 \mid \exists h \in \mathcal{H} : F_1(\phi_1(F_2^{-1}(\phi_2(u, h))), h), u \in [0, 1]\}.$$

No definīcijas seko, ka, izvēloties $\phi_2(t, h) = t$ un $\phi_1(y, (h_1, h_2)^T) = h_1 + h_2 x$, tiek iegūts lokācijas - skalēšanas modelis un izvēloties $\phi_1(x, h) = x$ un $\phi_2(t, h) = 1 - (1 - h)^h$, iegūst Lēmaņa alternatīvo modeli.

Lai novērtētu parametru h , var izmantot Mallova attālumu, ar kuru var tikt aprēķināts attālums starp kvantiļu funkcijām. Vispārīgā gadījumā strukturālo attiecību modelim teorētiskais parametrs h_0 tiek aprēķināts kā

$$h_0 = \operatorname{argmin}_{h \in H} \left\{ \frac{1}{b-a} \int_a^b F_1^{-1}(u) - \phi_1(F_2^{-1}(\phi_2(u, h)))^2 du \right\},$$

parametra h novērtējums $\hat{h}$ tiek iegūts sadalījuma funkcijas aizstājot ar empīriskajām sadalījuma funkcijām.

2. Hipotēžu pārbaude fiksēta lokācijas parametra gadījumā

Aplūkosim divas neatkarīgas izlases $X_1, \dots, X_n$ un $Y_1, \dots, Y_m$, kuru izlašu apjomi ir n un m . Pieņemsim, ka izlašu teorētiskās sadalījuma funkcijas ir F_1 un F_2 . Divu izlašu sadalījuma funkciju vienādības pārbaudei var lietot Kolmogorova-Smirnova testu divu izlašu gadījumam. Testa statistika tiek uzdota šādā formā

$$D = \sup_x \sqrt{\frac{nm}{n+m}} |F_{1n}(x) - F_{2m}(x)|,$$

kur n un m ir izlašu apjomi, bet F_{1n} un F_{2m} ir attiecīgās izlašu empīriskās sadalījuma funkcijas.

Definīcija 6. Par empīrisko sadalījuma funkciju $F_n(x)$ sauc funkciju, kuru definē

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{x_i \leq x}, \quad (2.1)$$

kur $x_1, x_2, \dots, x_n$ novērotie dati, bet $I_{x_i \leq x}$ - Indikātorfunkcija.

Hipotēžu pārbaude par lokācijas modeli reducējas kā

$$H_0 : F_1(x) = F_2(x) \text{ pret } H_1 : F_1(x) \neq F_2(x) \forall x, \text{ ja otrai izlasei pieskaita vai atņem } \theta \quad (2.2)$$

Šo hipotēzi var pārbaudīt, arī konstruējot vienlaicīgos ticamības intervālus varbūtību-varbūtību ($P - P$) un kvantiļu-kvantiļu ($Q - Q$) grafikiem.

Ja ir zināms procesa asimptotiskais sadalījums, tad var iegūt kritisko vērtību c pie fiksēta nozīmības līmeņa α . Ticamības joslas tiek konstruētas šādi:

1. $P - P$ procesam

$$P(F_{1n}(F_{2m}^{-1}(t)) - \frac{c}{\sqrt{n}} \leq F_1(F_2^{-1}(t)) \leq F_{1n}(F_{2m}^{-1}(t)) + \frac{c}{\sqrt{n}}) = \alpha, \forall 0 < t < 1$$

2. $Q - Q$ procesam

$$P(F_{1n}^{-1}(F_{2m}(x)) - \frac{c}{\sqrt{n}f(F_1^{-1}(F_2(x)))} \leq F_1^{-1}(F_2(x)) \leq F_{1n}^{-1}(F_{2m}(x)) + \frac{c}{\sqrt{n}f(F_1^{-1}(F_2(x)))}) = \alpha, \forall x \in R$$

Kad nulles hipotēze 2.2 ir spēkā, empīriskajam varbūtību - varbūtību procesam izpildās [2]

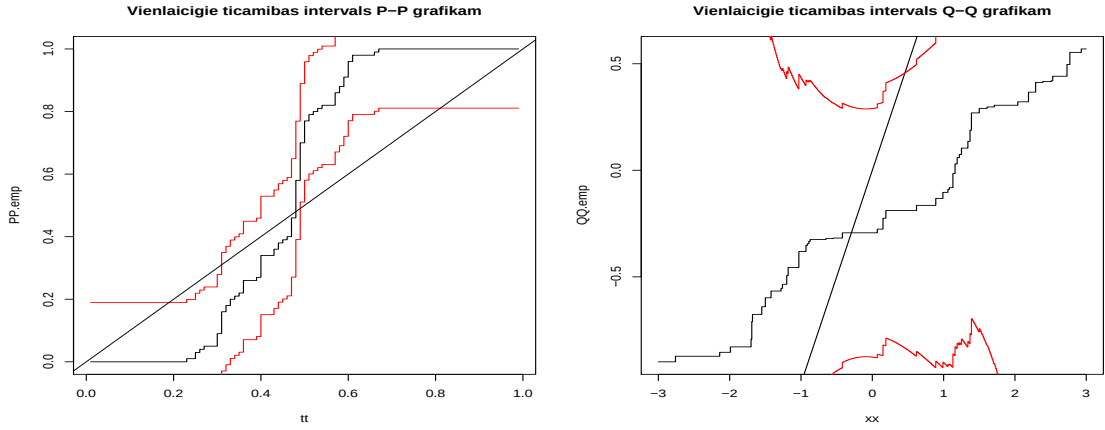
$$\forall \epsilon > 0 \lim_{n,m \rightarrow \infty} P(\sup_{0 < t < 1} |\sqrt{n}(F_{1n}F_{2m}^{-1}(t) - t) - (B^{(n)}(t)) + \frac{\sqrt{n}}{\sqrt{m}}B^{(m)}(t))| > \epsilon) = 0$$

Kad nulles hipotēzes 2.2 ir spēkā, tad kvantiļu - kvantiļu procesam izpildās [2]

$$\forall \epsilon > 0 \lim_{n,m \rightarrow \infty} P(\sup_{-\infty < x < \infty} |\sqrt{n}f_1(x)(F_{1n}F_{2m}^{-1}(x) - x) - (B^{(n)}(F_2(x))) + \frac{\sqrt{n}}{\sqrt{m}}B^{(m)}(F_2(x))| > \epsilon) = 0$$

Definīcija 7. Par Brauna tiltu sauc Gausa procesu $\{B_t, t \in [0, 1]\}$, kura kovariācijas struktūra ir $cov(B_s, B_t) = s(1 - t)$, $s < t$. Ja $W(t)$ ir standarta Vīnera process (t.i. $t \geq 0$, $W(t)$ ir normāli sadalīts ar vidējo vērtību 0 un dispersiju t , un pieaugums ir stacionārs un neatkarīgs), tad $B_t = W_t - tW_1$ ir Brauna tilts.

Redzam, ka procesa asimptotiskais sadalījums ir divu Brauna tiltu lineāra kombinācija. Ticamības intervāli tiek konstruēti, izmantojot J. Cielēna diplomdarbu [2], kur kritiskā vērtība tiek aprēķināta, veicot Brauna tiltu simulācijas ($\alpha = 0.95, c = 1.89$.)



1. att.: $N(0, 1)$ pret $N(1, 4)$, $n = 100$

Attēlā 1. pārbaudīta hipotēze par divu sadalījumu vienādību. Izlases X un Y sadalītas attiecīgi $N(0, 1)$ un $N(1, 4)$, bet $n = 100$. Pirmajā grafikā attēlotas vienlaicīgās ticamības joslas varbūtību-varbūtību ($P - P$) grafikam. Otrajā grafikā attēlotas vienlaicīgās ticamības joslas kvantiļu-kvantiļu ($Q - Q$) grafikam. Var secināt, ka H_0 par divu izlašu vienādu sadalījumu noraidām, jo grafikos redzamā diagonāle iziet ārpus ticamības joslām kaut vienā punktā.

3. Vienlaicīgās ticamības joslas lokācijas parametram

Šajā nodaļā tiks aplūkotas vienlaicīgās ticamības joslas pašam parametram θ , kā to ir skatījušies Doksums un Sievers savā publikācijā [3]. Vispārējā Lokācijas modeļa $F_1(x) = F_2(x + \theta) \forall x$ vietā viņi apskata gadījumu, kad $F_1(x) = F_2(x + \Delta(x))$ kādai funkcijai $\Delta(x) = F_2^{-1}\{F_1(x)\} - x$.

Piezīme 1. Doksums un Sievers savā publikācijā [3] izvirzīja vairākus jautājumus (i) vai apstrāde ir izdevīga visiem izlases datiem, tas ir $\Delta(x) > 0$ visiem x ? (ii) ja nē, tad kurai daļai no izlases izdevīgi, tas ir $\{x : \Delta(x) > 0\}$? (iii) vai lokācijas modeli ietekmē $\Delta(x) = \theta$, ja ņem kādai θ , visiem x ? (iv) ja nē, tad vai $\Delta(x) = \alpha + \beta x$ priekš kādas α un β , visiem x ?

Atbildes uz iepriekšējiem jautājumiem sniedz ticamības joslas $[\Delta_*(x), \Delta^*(x)]$ priekš $\Delta(x)$ vienlaicīgi visiem x . Uz jautājumu (i) atbilde ir apstiprinoša, ja $\Delta_*(x) > 0$ visiem

x . (ii) risināums ir $\Delta_*(x) > 0$, (iii) noraida, ja neviena horizontāla līnija neiekļaujas ticamības joslās un (iv) arī noraida, ja taisna līnija neiekļaujas ticamības joslās.

Ņem $N = n + m$ un $M = mn/N$. Doksums un Sievers [3] aplūkoja vienlaicīgās ticamības joslas balstītas uz Kolmogorova - Smirnova statistiku.

$$T(F_{1n}, F_{2m}) = D_n = \sqrt{M} \sup_x |F_{1n}(x) - F_{2m}(x)|. \quad (3.1)$$

Teorēma 2. [3] $(1 - \alpha)$ vienlaicīgās ticamības joslas priekš $\Delta(x)$ tiek uzdotas

$$[F_{2m}^{-1}\{F_{1n}(x) - K_{S,\alpha}/M^{1/2}\} - x, F_{2m}^{-1}\{F_{1n}(x) + K_{S,\alpha}/M^{1/2}\} - x), \quad (3.2)$$

kad x mainās no $-\infty$ līdz ∞ , un $K_{S,\alpha}$ ir izvēlēts no Kolmogorova-Smirnova tabulām, lai atbilstu $P(T(F_{1n}, F_{2m}) \leq K_{S,\alpha}) = 1 - \alpha$.

Piezīme 3. Apzīmējot ar $[t]$ skaitļa t veselo daļu, kas ir lielākais veselais skaitlis, kas nepārsniedz t , bet ar $\langle t \rangle$ - vismazāko skaitli, kas ir lielāks vai vienāds ar t ; $X_1 < \dots < X_n$ un $Y_1 < \dots < Y_m$, $Y(j) = -\infty (j < 0)$ un $Y(j) = \infty (j \geq m + 1)$. Tad vienlaicīgos ticamības intervālus (3.2) varam pārrakstīt:

$$[S_*(x), S^*(x)] = \left[Y \left\{ \left\langle n \left(\frac{i}{m} - K_{S,\alpha}/M^{\frac{1}{2}} \right) \right\rangle \right\} - x, Y \left\{ \left[n \left(\frac{i}{m} + K_{S,\alpha}/M^{\frac{1}{2}} \right) \right] + 1 \right\} - x \right], \quad (3.3)$$

visiem $x \in [X(i), X(i + 1))$ ($i = 0, 1, \dots, n$) un $X(0) = -\infty$, $X(n + 1) = \infty$.

Doksums un Sievers publikācijā [3] aplūkoja arī ticamības joslas, kas balstītas uz svērto suprēma normu:

$$W_N = W_N(F_{1N}, F_{2M}) = M^{\frac{1}{2}} \sup_{\{x: a \leq F_{1n}(x) \leq b\}} \frac{|F_{1n}(x) - F_{2m}(x)|}{\Psi\{H_N(x)\}}, \quad (3.4)$$

kur $H_N(x) = \lambda F_{1n}(x) + (1 - \lambda)F_{2m}(x)$, $\lambda = m/N$ un $0 \leq a < b \leq 1$. Ja izvēlēsimies $\Psi(t) = \{t(1 - t)\}^{\frac{1}{2}}$, tad iegūtu aptuveni vienādu svaru katram x tadā nozīmē, ka

$$M^{\frac{1}{2}} \{F_{1n}(x) - F_{2m}(x)\} / \Psi\{H_N(x)\}$$

ir asimptotiskā dispersija neatkarīgi no x . Tālāk risina nevienādību $|W_N(F_{1n}, F_{2m})| \leq K$ priekš F_{2m} . Kad ievietojam $\Psi(t) = \{t(1 - t)\}^{\frac{1}{2}}$, šī nevienādība pārveidojas par

$$\{F_{2m}(x) - F_{1n}(x)\}^2 \leq K^2 \{\lambda F_{1n}(x) + (1 - \lambda)F_{2m}(x)\} [1 - \{\lambda F_{1n}(x) + (1 - \lambda)F_{2m}(x)\}] / M,$$

priekš x , kas atbilst $a \leq F_{1n}(x) \leq b$. Apzīmēsim $c = K^2/M$, $v = F_{2m}(x)$, tad nevienādību varam pārrakstīt kā $d(v) \leq 0$, kur

$$d(v) = \{1 + c(1 - \lambda)^2\}v^2 - \{2u - c(1 - \lambda)(2\lambda u - 1)\}v + u^2 - c\lambda u + c\lambda^2 u^2. \quad (3.5)$$

Tā kā koeficients v^2 ir pozitīvs $d(v) \leq 0$ tad un tikai tad, ja v ir starp divām reālām vienādojuma saknēm $d(v) = 0$.

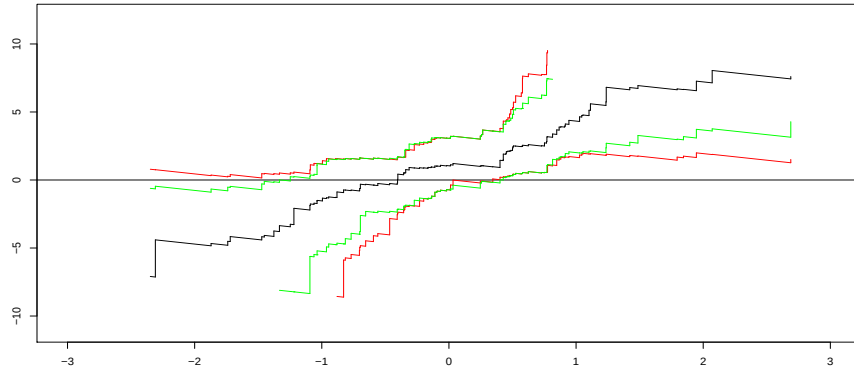
Piezīme 4. Ja $P_{F=G}(W_N \leq K) = 1 - \alpha$, tad vienlaicīgie ticamības intervāli priekš $\Delta(x)$, kas balstīti uz W_N ar $\Psi(t) = \{t(1 - t)\}^{\frac{1}{2}}$ ir

$$[F_{2m}^{-1}[h^-\{F_{1n}(x)\}] - x, F_{2m}^{-I}[h^+\{F_{1n}(x)\}] - x], x \in \{x : a \leq F_{1n}(x) \leq b\}, \quad (3.6)$$

$$h \pm(u) = \frac{u + \frac{1}{2}c(1 - \lambda)(1 - 2\lambda u) \pm \frac{1}{2}\{c^2(1 - \lambda)^2 + 4cu(1 - u)\}^{\frac{1}{2}}}{1 + c(1 - \lambda)^2} \quad (3.7)$$

un $c = K^2/N$.

Turpmāk šos mēs sauksim par W vienlaicīgās ticamības inetrvāliem un rakstīsim $[W_*(x), W^*(x)]$.



2. att.: $N(0, 1)$ pret $N(1, 4)$, $n = m = 100$

Attēlā 2. sarkanā krāsā attēlotas $[S_*(x), S^*(x)]$ vienlaicīgās ticamības joslas un zaļā krāsā $[W_*(x), W^*(x)]$ vienlaicīgās ticamības joslas, kur X un Y sadalītas attiecīgi $N(0, 1)$ un $N(1, 4)$, bet $m = n = 100$ un $\alpha = 0,05$. Abas vienlaicīgās ticamības joslas ir diezgan līdzīgas, bet redzam, ka $[W_*(x), W^*(x)]$ josla ir šaurāka galos. Šajā gadījumā skaidrs, ka horizontāla taisne neiekļaujas vienlaicīgajās ticamības joslās. Tas nozīmē, ka H_0 tiek

noraidīta, un mums nav lokācijas modelis. Bet varētu būt lokācijas - skalēšanas modelis, jo taisne iekļaujas joslās.

4. Ticamības intervālu konstruēšana, izmantojot empīriskos procesus

4.1. Empīriskā procesa definīcija un piemēri

Vienkāršākajā gadījumā empīriskais process tiek veidots no empīriskās un teorētiskās sadalījuma funkcijas starpības. Pieņem, ka $X_1, \dots, X_n$ ir neatkarīgu, vienādi sadalītu (tālāk tekstā tiks izmantots saīsinājums no angļu val. - iid) gadījuma lielumu izlase ar sadalījuma funkciju F .

Apgalvojums 5. [9]

$$\sqrt{n}(F_n(x) - F(x)) \rightarrow_d N(0, F(x)(1 - F(x))).$$

Teorēma 6. *Ir spēkā (Glivenko-Kantelli) [10]*

$$\sup_{-\infty < x < \infty} (|F_n(x) - F(x)|) \xrightarrow[n \rightarrow \infty]{g.d.} 0.$$

Teorēma 7. *Pieņemsim, ka $X_1, \dots, X_n$ ir iid gadījuma leilumi ar sadalījuma funkciju F , tad*

$$\sup_{-\infty < x < \infty} |\sqrt{n}(F_n(x) - F(x))| \rightarrow \sup_{-\infty < x < \infty} |B(F(x))|, \quad (4.1)$$

kur B ir Brauna tilts. Kreisās puses starpība ir pazīstama kā Kolmogorova-smirnova statistika.

Teorēma 8. *Līdzīgi kā sadalījuma funkcijām kvantiļu funkcijām ir spēkā*

$$\sup_{0 < y < 1} (|F_n^{-1}(y) - F^{-1}(y)|) \xrightarrow[n \rightarrow \infty]{as} 0.$$

4.2. Empīriskie procesi kvantilēm vienas izlases gadījumā

Izmantojot blīvuma funkcijas novērtējumu [11]

Dota $X_1, X_2, \dots, X_n$ iid gadījumu izlase ar pilnīgi nepārtrauktu sadalījuma funkciju $F(\cdot)$, ar $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$, tiek apzīmētas sakārtotas statistikas no $X_1 \dots X_n$.

Tālāk tiek definēta empīriskā sadalījuma funkcija $F_n(x)$ un empīriskā kvantiļu funkcija $Q_n(y)$

$$F_n(x) = \begin{cases} 0 & \text{if } X_{1:n} > x, \\ k/n & \text{if } X_{k:n} \leq x < X_{k+1:n}, k = 1, 2, \dots, n-1, \\ 1 & \text{if } X_{n:n} \leq x \end{cases} \quad (4.2)$$

$$Q_n(y) = X_{k:n} \text{ if } (k-1)/n < y \leq k/n, k = 1, 2, \dots, n. \quad (4.3)$$

Ir pašsaprotami izmantot empīrisko kvantiļu funkciju kā kvantiļu funkcijas novērtējumu

$$Q(y) := F^{-1}(y), \text{ kur } F^{-1}(y) = \inf\{x : F(x) \geq y\}, (0 < y < 1). \quad (4.4)$$

Teorēma 9. [11] Dota $X_1, X_2, \dots, X_n$ iid izlase ar nepārtrauktu sadalījuma funkciju F . Nosacījumi

(i) $F(x)$ ir divreiz diferencējam intervāla (a, b) , kur

$$a = \sup\{x : F(x) = 0\}, \quad b = \inf\{x : F(x) = 1\}, \quad -\infty \leq a < b \leq +\infty$$

(ii) $F'(x) = f(x) > 0$ intervālā (a, b)

(iii) kādam $\gamma > 0$ ir

$$\sup_{a < x < b} F(x)(1 - F(x)) \frac{|f'(x)|}{f^2(x)} = \sup_{0 < y < 1} y(1 - y) \frac{|f'(F^{-1}(y))|}{f^2(F^{-1}(y))}.$$

Tad

$$\sup_{\delta_n \leq y \leq 1 - \delta_n} |\rho_n(y) - u_n(y)| = O(n^{-1/2} \log \log n), \quad (4.5)$$

kur $\delta_n = 25n^{-1} \log \log n$.

Ja nosacījumiem (i), (ii) un (iii) pievieno vēl

$$(iv) A := \lim_{x \downarrow a}^- f(x) < \infty, \quad B := \lim_{x \uparrow b}^- f(x) < \infty,$$

$$(v) (v, \alpha) \min(A, B) > 0,$$

(v, β) ja $A = 0$ (attiecīgi $B = 0$) tad f ir nedilstoša (neaugoša) intervālā pa labi no a (pa kreisi no b),

tad, ja ņem (v, α) , iegūst

$$\sup_{0 \leq y \leq 1} |\rho_n(y) - u_n(y)| = O(n^{-1/2} \log \log n), \quad (4.6)$$

un ja ņem (v, β) , iegūst

$$\sup_{0 < y < 1} |\rho_n(y) - u_n(y)| = \begin{cases} O(n^{-1/2} \log \log n) & \text{if } \gamma < 1, \\ O(n^{-1/2} (\log \log n)^2) & \text{if } \gamma = 1, \\ O(n^{-1/2} (\log \log n)^\gamma (\log n)^{(1+\epsilon)(\gamma-1)}) & \text{if } \gamma > 1, \end{cases} \quad (4.7)$$

kur $\epsilon > 0$ ir patvaļīgs un γ ir kā nosacījumā (iii).

Teorēma 10. [11] Dota $X_1, X_2, \dots$ neatkarīga un vienādi sadalīta izlase ar nepārtrauktu sadalījuma funkciju F , nosacījumi no Teorēmas 9 (i), (ii) un (iii). Ņemot $\delta_n = 25n^{-1} \log \log n$, var definēt Brauna tiltu $\{B_n(y); 0 \leq y \leq 1\}$ katram n

$$\sup_{\delta_n \leq y \leq 1 - \delta_n} |f(F^{-1}(y)) n^{1/2} (F_n^{-1}(y) - F^{-1}(y)) - B_n(y)| = O(n^{-1/2} \log n). \quad (4.8)$$

Teorēma 11. [11] Ja papildina Teorēmas 9 nosacījumus (i), (ii) un (iii) ar (iv), (v), tad var definēt Brauna tiltu $\{B_n(y); 0 \leq y \leq 1\}$ katram n

$$\begin{aligned} & \sup_{0 < y < 1} |f(F^{-1}(y)) n^{1/2} (F_n^{-1}(y) - F^{-1}(y)) - B_n(y)| \\ &= \begin{cases} O(n^{-1/2} \log n) & \text{if } \gamma < 2, \\ O(n^{-1/2} (\log \log n)^\gamma (\log n)^{(1+\epsilon)(\gamma-1)}) & \text{if } \gamma \geq 2, \end{cases} \end{aligned} \quad (4.9)$$

kur $\epsilon > 0$ ir patvaļīgs un γ ir kā nosacījumā (iii).

Izmantojot (4.9), var konstruēt vienlaicīgās ticamības joslas kvantiļu funkcijai $F^{-1}(y) = x$:

$$\begin{aligned} F_n^{-1}(y) - n^{-1/2} \frac{c}{f(F^{-1}(y))} &\leq F^{-1}(y) \\ &\leq F_n^{-1}(y) + n^{-1/2} \frac{c}{f(F^{-1}(y))}, \quad 0 < y < 1, \end{aligned} \quad (4.10)$$

kur konstante $c = c(\alpha)$ ir atkarīga no vēlamā ticamības līmeņa $1 - \alpha$, t.i. $c(\alpha)$ katram $\alpha \in (0, 1)$ ir definēts kā $P\{\sup_{0 \leq y \leq 1} |B(y)| \leq c(\alpha)\} = 1 - \alpha$.

Diemžēl (4.10) ir balstīta uz nezināmu kvantiļu blīvuma funkciju $(d/dy)F^{-1}(y) = 1/f(F^{-1}(y))$.

Csorgo un Revesz (1981b) [12] ierosina novērtēt kvantiļu blīvuma funkciju. Viens no mērķiem ir atrast tiešu novērtējumu funkcijai $1/f(F^{-1}(y))$, lai to risinātu ir divas teorēmas, ko izmantot.

Teorēma 12. *Dota $X_1, X_2, \dots$ neatkarīga un vienādi sadalīta izlase ar nepārtrauktu sadalījuma funkciju F , kas atbilst Teorēmas 11 nosacījumiem. Atbilst arī*

(vi)

$$\sup_{a < x < b} \frac{(F(x)(1 - F(x)))^2}{f(x)} = \sup_{0 < y < 1} \frac{(y(1 - y))^2}{f(F^{-1}(y))} \leq C$$

ar kādu $C > 0$.

Tad

$$\lim_{n \rightarrow \infty} \omega(n) \sup_{\epsilon_n \leq y \leq 1 - \epsilon_n} \left| \frac{F_n^{-1}(y + a_n) - F_n^{-1}(y - a_n)}{2a_n} - \frac{1}{f(F^{-1}(y))} \right| = 0, \quad (4.11)$$

kur $a_n = n^{-\alpha}$, $\epsilon_n = n^{-\beta}$, $\omega_n = n^\delta$ un $3\beta + \delta < \alpha < \frac{1}{2}$, $2\delta + 4\beta + \alpha < 1$.

Novērtējums $(2a_n)^{-1}(F_n^{-1}(y + a_n) - F_n^{-1}(y - a_n))$ priekš $1/f(F^{-1}(y))$ ir analogs novērtējumam $f_n^* = (2a_n)^{-1}(F_n^{-1}(x + a_n) - F_n^{-1}(x - a_n))$, kas bieži tiek lietots, lai novērtētu blīvuma funkciju f . Tiek ieviests vispārējs novērtējums

$$\phi_n(y) = a_n^{-1} \int_0^1 \lambda \left(\frac{y - u}{a_n} \right) dF_n^{-1}(y) \quad (4.12)$$

priekš kvantiļu blīvuma funkcijas $1/f(F^{-1}(y))$.

Teorēma 13. *Dota $X_1, X_2, \dots$ neatkarīga un vienādi sadalīta izlase ar nepārtrauktu sadalījuma funkciju $F(\cdot)$, kas atbilst Teorēmas 12 nosacījumiem. Atbilst arī*

(vii)

$$\sup_{a < x < b} |f''(x)| = \sup_{0 < y < 1} |f''(F^{-1}(y))| \leq C$$

ar kādu $C > 0$.

Pieņemam, ka $\lambda(\cdot)$ ir blīvuma funkcija, kas ir nepārtraukta $(-\infty, +\infty)$ un ārpus intervāla $(-\frac{1}{2}, \frac{1}{2})$ tuvojas nullei un

$$\int_{-1/2}^{1/2} x\lambda(x)dx = 0, \quad \int_{-1/2}^{1/2} x^2\lambda(x)dx < -\infty, \quad \sup_{-1/2 \leq x \leq 1/2} |\lambda'(x)| < \infty.$$

Tad iegūst

$$\lim_{n \rightarrow \infty} \omega(n) \sup_{\epsilon_n \leq y \leq 1 - \epsilon_n} \left| \phi_n(y) - \frac{1}{f(F^{-1}(y))} \right| = 0, \quad (4.13)$$

kur $\phi_n(y)$ ir definēts, kā (4.12) un $a_n = n^{-\alpha}$, $\epsilon_n = n^{-\beta}$, $\omega_n = n^\delta$ un $\delta + 10\beta < 2\alpha$ $\alpha + 2\beta + \delta < \frac{1}{2}$.

Sekas 14. *Izmantojot $X_1, X_2, \dots$ iid gadījumu izlasi ar nepārtrauktu sadalījuma funkciju F , kas atbilst Teorēmas 12 nosacījumiem, varam definēt Brauna tiltu $\{B_n(y); 0 \leq y \leq 1\}$ katram n*

$$\sup_{\epsilon_n \leq y \leq 1 - \epsilon_n} \left| \frac{n^{1/2}}{\phi_n^0(y)} (F_n^{-1}(y) - F^{-1}(y)) - B_n(y) \right| = o \left(\frac{(\log \log n)^{1/2}}{\omega(n)} \right), \quad (4.14)$$

un

$$\begin{aligned} \lim_{n \rightarrow \infty} P \{ & F_n^{-1}(y) - \frac{c}{n^{1/2}} \phi_n^0(y) \leq F^{-1}(y) \leq F_n^{-1}(y) \\ & + \frac{c}{n^{1/2}} \phi_n^0(y); \quad \epsilon_n \leq y \leq 1 - \epsilon_n \} \\ & = P \{ \sup_{0 \leq y \leq 1} |B(y)| \leq c \} \\ & = 1 - \sum_{k \neq 0} (-1)^{k+1} e^{-1/2 k^2 c^2}, \quad c \geq 0, \end{aligned} \quad (4.15)$$

kur

$$\phi_n^0(y) = \frac{F_n^{-1}(y + a_n) - F_n^{-1}(y - a_n)}{2a_n}, \quad (4.16)$$

kur $a_n = n^{-\alpha}$, $\epsilon_n = n^{-\beta}$, $\omega_n = n^\delta$ un $3\beta + \delta < \alpha < \frac{1}{2}$, $2\delta + 4\beta + \alpha < 1$ un $\{B_n(y); 0 \leq y \leq 1\}$ ir Brauna tilts.

Taisni ticamības intervāli kvantilēm [11]

$c_1(\alpha)$ tiek definēts izmantojot

$$P \{ \sup_{0 \leq y \leq 1} B(y) \leq c_1(\alpha) \} = 1 - \alpha, \quad (4.17)$$

kur $\{B_n(y); 0 \leq y \leq 1\}$ ir Brauna tilts un $\alpha \in (0, 1)$. Izmantojot $X_1, \dots, X_n$ iid gadījumu izlase ar nepārtrauktu sadalījuma funkciju F ar nosacījumu, ka empīriskā sadalījuma funkcija ir $F_n(\cdot)$.

$$P \{ F_n(x) \leq F(x) + n^{-1/2} c_1(\alpha); -\infty < x < +\infty \} \rightarrow (1 - \alpha) \quad (4.18)$$

un pieņemot, ka $y = F(x)$, vajag iegūt ($n \rightarrow \infty$)

$$\begin{aligned} P \{ & F_n(F^{-1}(y)) \leq y + n^{-1/2} c_1(\alpha); F^{-1}(0) < F^{-1}(y) < F^{-1}(1) \} \\ & P \{ F^{-1}(y) \leq F_n^{-1}(y + n^{-1/2} c_1(\alpha)); 0 < y < 1 \} \rightarrow (1 - \alpha). \end{aligned} \quad (4.19)$$

Visiem $\omega \in \Omega$

$$\begin{aligned}
& \sup_{n^{-1/2}c_1(\alpha) \leq y \leq 1-n^{-1/2}c_1(\alpha)} (F_n^{-1}(y + n^{-1/2}c_1(\alpha)) - F^{-1}(y)) \\
& \geq F_n^{-1}(1) - F^{-1}(1 - n^{-1/2}c_1(\alpha)) \\
& = X_{n:n} - F^{-1}\left(\frac{n}{n+1}\right) + F^{-1}\left(\frac{n}{n+1}\right) - F^{-1}(1 - n^{-1/2}c_1(\alpha)). \tag{4.20}
\end{aligned}$$

Pieņemot, ka $F(x) = 1 - e^{-x}$, $x \geq 0$. Tad $F^{-1}(y) = \log(1/(1-y))$ un (4.20) ir

$$\begin{aligned}
& \sup_{0 \leq y \leq 1-n^{-1/2}c_1(\alpha)} \left(F_n^{-1}(y + n^{-1/2}c_1(\alpha)) - \log \frac{1}{1-y} \right) \\
& \geq (X_{n:n} - \log n) - \log \left(\frac{n+1}{n} \right) + \log \left(\frac{n+1}{n^{1/2}} \right) - \log c_1(\alpha). \tag{4.21}
\end{aligned}$$

Kā sekas no Teorēmas 10 iegūstam.

Sekas 15. *Dota $X_1, X_2, \dots$ neatkarīga un vienādi sadalīta izlase ar nepārtrauktu sadalījuma funkciju F , kas atbilst Teorēmas 10 nosacījumiem, bet $\{B_n(y); 0 \leq y \leq 1\}$ ir Brauna tilts, tad*

$$\begin{aligned}
& \lim_{n \rightarrow \infty} P\{F^{-1}(y) \leq F_n^{-1}(y + n^{-1/2}c); \epsilon_n \leq y \leq 1 - \epsilon_n\} \\
& \lim_{n \rightarrow \infty} P\{F_n^{-1}(y - n^{-1/2}c) \leq F^{-1}(y); \epsilon \leq y \leq 1 - \epsilon_n\} \\
& = P\left\{ \sup_{0 \leq y \leq 1} B(y) \leq c \right\} = 1 - e^{-2c^2}, \quad c \geq 0, \tag{4.22}
\end{aligned}$$

kur $\epsilon_n := n^{-1/2+\delta}$ katram $\delta \in (0, \frac{1}{2})$.

Sekas 16. *Izmantojot Teorēmas 10 nosacījumus iegūst*

$$\begin{aligned}
& \lim_{n \rightarrow \infty} P\{F_n^{-1}(y - n^{-1/2}c) \leq F^{-1}(y) \leq F_n^{-1}(y + n^{-1/2}c); \epsilon_n \leq y \leq 1 - \epsilon_n\} \\
& = P\left\{ \sup_{0 \leq y \leq 1} |B(y)| \leq c \right\} = 1 - \sum_{k \neq 0} e^{-2k^2 c^2}, \quad c \geq 0, \tag{4.23}
\end{aligned}$$

kur $\epsilon_n := n^{-1/2+\delta}$ katram $\delta \in (0, \frac{1}{2})$.

Tātad varam konstruēt ticamības intervālus kvantilēm. Izmantotais ϵ_n tiek izvēlēts tā, lai būtu vienāds ar $\log \log \dots \log n / n^{1/2}$, kur $\log$ var izmantot nierobežotas reizes vai vienkārši kā $n^{1/2}\epsilon_n \rightarrow \infty$, kad $n \rightarrow \infty$. Turpmāk aprakstītās teorēmas izmanto, ja ir nepieciešams konstruēt ticamības intervālus konkrētām kvantilēm.

Sekas 17. *Izmantojot Teorēmas 10 nosacījumus*

$$\begin{aligned}
& \lim_{n \rightarrow \infty} P\{F_n^{-1}(y - n^{-1/2}c) \leq F^{-1}(y); a_1 \leq y \leq a_2\} \\
&= \lim_{n \rightarrow \infty} P\{F^{-1}(y) \leq F_n^{-1}(y + n^{-1/2}c); 1 - a_2 \leq y \leq 1 - a_1\} \\
&= P\left\{\sup_{a_1 \leq y \leq a_2} B(y) \leq c\right\}, \quad c > 0.
\end{aligned} \tag{4.24}$$

Sekas 18. *Izmantojot $a_1 = \epsilon_n$, a_2 ir fiksēts un Teorēmas 10 nosacījumus*

$$\begin{aligned}
& \lim_{n \rightarrow \infty} P\{F_n^{-1}(y - n^{-1/2}c) \leq F^{-1}(y); \epsilon_n \leq Y \leq a_2\} \\
&= \lim_{n \rightarrow \infty} P\{F^{-1}(y) \leq F_n^{-1}(y + n^{-1/2}c); 1 - a_2 \leq y \leq 1 - \epsilon_n\} \\
&= P\left\{\sup_{0 \leq y \leq a_2} B(y) < c\right\}, \quad c > 0.
\end{aligned} \tag{4.25}$$

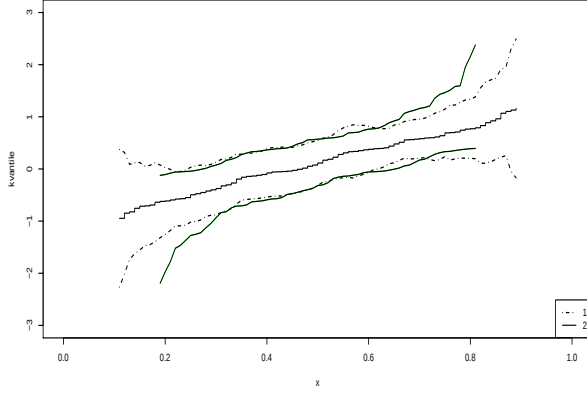
Sekas 19. *Izmantojot $0 < a_1 \leq a_2 < 1$ ir fiksēti un Teorēmas 10 nosacījumus, un ϵ_n kā Sekās 15, iegūst*

$$\begin{aligned}
& \lim_{n \rightarrow \infty} P\{F_n^{-1}(y - n^{-1/2}c) \leq F^{-1}(y); \epsilon_n \leq y \leq a_1, a_2 \leq y \leq 1 - \epsilon_n\} \\
&= \lim_{n \rightarrow \infty} P\{F^{-1}(y) \leq F_n^{-1}(y + n^{-1/2}c); \epsilon_n \leq y \leq 1 - a_2, 1 - a_1 \leq y \leq 1 - \epsilon_n\} \\
&= P\left\{\sup_{0 \leq y \leq a_1, a_2 \leq y \leq 1} B(y) \leq c\right\}, \quad c > 0.
\end{aligned} \tag{4.26}$$

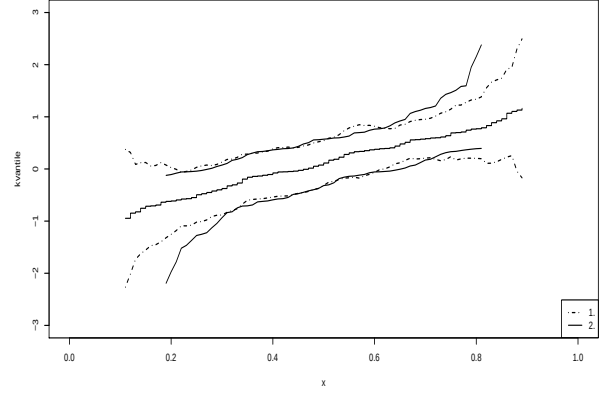
Tātad ja interesē konstruēt augšējo un apakšējo ticamības joslu kavntiļu funkcijai $Q = F^{-1}$ pa "vidu" (tas ir konkrētā apgabalā) tikai, tad formula (4.24) dos daudz ekonomiskāku rezultātu nekā (4.23). Izmantojot $a_1 = \epsilon_n$ un $a_2 = 1 - \epsilon_n$, mēs iegūstam Sekas 15 Ar $c = c_1(\alpha)$ kā (4.17) korekta versija (4.19) ir

$$P\{F^{-1}(y) \leq F_n(y + n^{-1/2}c_1(\alpha)); \epsilon_n < y < 1 - \epsilon\} = 1 - \alpha. \tag{4.27}$$

Attēlā 3.a attēlotas vienlaicīgās ticamības joslas $N(0, 1)$ sadalītiem datiem, 1. konstruētas, izmantojot formulu (4.10) un 2. konstruētas, izmantojot (4.23). Attēlā 3.b attēlotas vienlaicīgās ticamības joslas $N(1, 4)$ sadalītiem datiem, 1. konstruētas, izmantojot, formulu (4.10) un 2. konstruētas, izmantojot (4.23).



(a) $N(0, 1), n = 100$



(b) $N(1, 4), m = 100$

3. att.: Vienlaicīgās ticamības joslas attiecīgo datu kvantilēm izmantojot (4.10) un (4.23)

4.3. Empīriskie procesi kvantilēm divu izlašu gadījumā

Pārbaudīsim hipotēzi vai pastāv lokācijas modelis 1.4, izmantojot vienlaicīgās ticamības joslas divu kvantiļu funkciju starpībai. Izmantosim Petter Laake publikāciju [4]. Vispārējā Lokācijas modeļa gadījumā 1.4, izpildās

$$\Lambda(t) = F_1^{-1}(t) - F_2^{-1}(t),$$

kur $0 \leq t \leq 1$, F_1^{-1} un F_2^{-1} ir attiecīgi F_1 un F_2 kreisās inversās funkcijas. Kreisā inversā funkcija funkcijai H tiek definēta kā $H^{-1}(s) = \inf\{z : H(z) \geq s\}$.

Atzīmēsim, ka

$$\int_0^1 \Lambda(t) dt = E(X) - E(Y).$$

Kvantiļu starpības funkcija ir saistīta ar lokācijas funkciju no Doksuma (1974) [13]. Doksums $\Delta(x) = F_2^{-1}[F_1(x)] - x$ nosauca par lokācijas funkciju. Pieņemot, ka $\Delta(x)$ ir horizontālais attālums starp F_1 un F_2 attiecīgajā punktā x , mēs redzam, ka

$$\Lambda[F_1(x)] = -\Delta(x), \quad (4.28)$$

kur $\Delta(x)$ ir identiska ar lokācijas funkciju $F_2^{-1}[F_1(x)] - x$.

Novērtējums $\Lambda(t)$ ir uzdots kā divu kvantiļu funkciju starpība

$$\hat{\Lambda}(t) = F_{1n}^{-1}(t) - F_{2m}^{-1}(t), \quad (4.29)$$

kur $0 \leq t \leq 1$, kur ar F_{1n}^{-1} un F_{2m}^{-1} apzīmējam F_1 un F_2 kreisās inversās funkcijas.

Tālāk izmantosim, ka $n^{1/2}[F_{1n}^{-1}(t) - F_1^{-1}(t)]$ un $m^{1/2}[F_{2m}^{-1}(t) - F_2^{-1}(t)]$ konverģē uz Gausa procesu.

Definīcija 8. Stohastisku procesu $\{X_t; t \in T\}$ sauc par Gausa procesu, ja jebkura ierobežota lineāra kombinācija no paraugu būs normāli sadalīta.

Pieņemam, ka $f(x)$ un $g(y)$ ir $F_1(x)$ un $F_2(y)$ nepārtraukti atvasinājumi, kas apmierina $0 < f(x) < \infty$ un atiecīgi $0 < g(y) < \infty$. Ar $W_1(t)$ un $W_2(t)$ tiek apzīmēti Brauna tilti, tas nozīmē, ka tiek apskatīts Gausa process ar vidējo nulli un kovariāciju funkciju $s(1-t)$, kur $0 \leq s \leq t \leq 0$.

Zināms, ka

$$n^{1/2}[F_{1n}^{-1}(t) - F_1^{-1}(t)] \xrightarrow{d} \frac{W_1(t)}{f[F_1^{-1}(t)]}$$

un līdzīgi

$$m^{1/2}[F_{2m}^{-1}(t) - F_2^{-1}(t)] \xrightarrow{d} \frac{W_2(t)}{g[F_2^{-1}(t)]}.$$

Pierādījumu skatīt Serflings [14].

Pieņemam, ka $m, n \rightarrow \infty$ tā, ka $(m/N) \rightarrow \theta$, kur $N = m + n$. Tad

$$N^{1/2}[\hat{\Lambda}(t) - \Lambda(t)] = (N/n)^{1/2}n^{1/2}[F_{1n}^{-1}(t) - F_1^{-1}(t)] - (N/m)^{1/2}m^{1/2}[F_{2m}^{-1}(t) - F_2^{-1}(t)],$$

kas dod

$$N^{1/2}[\hat{\Lambda}(t) - \Lambda(t)] \xrightarrow{D} \frac{1}{\theta^{1/2}} \frac{W_1(t)}{f[F_1^{-1}(t)]} - \frac{1}{(1-\theta)^{1/2}} \frac{W_2(t)}{g[F_2^{-1}(t)]}. \quad (4.30)$$

Izmantojot rezultātus, mēs iegūstam, ka $N^{1/2}[\hat{\Lambda}(t) - \Lambda(t)]$ asimptotiskā dispersija ir

$$k^2(t) = \left\{ \frac{1}{\theta f^2[F_1^{-1}(t)]} + \frac{1}{(1-\theta)g^2[F_2^{-1}(t)]} \right\} t(1-t).$$

Tad suprēma statistika būs

$$\sup_{0 < t < 1} \left| \left(\frac{1}{\theta^{1/2}} W_1(t) - \frac{1}{(1-\theta)^{1/2}} W_2(t) \right) \frac{f(F_1^{-1}(t))}{g(F_2^{-1}(t))} \right| \quad (4.31)$$

Turpinot pētīt problēmu publikācijā [4] tiek pieņemts, ka F_2 ir zināma sadalījuma funkcija. Šādā gadījumā ir jāsalīdzina izlase ar zināmu un nezināmu sadalījumu. Tādā gadījumā iegūst, ka

$$n^{1/2}[\hat{\Lambda}(t) - \Lambda(t)] = n^{1/2}[F_{1n}^{-1}(t) - F_1^{-1}(t)]$$

un

$$n^{1/2}[\hat{\Lambda}(t) - \Lambda(t)] \xrightarrow{D} \frac{W_1(t)}{f[F_1^{-1}(t)]}.$$

Tātad $N^{1/2}[\hat{\Lambda}(t) - \Lambda(t)]$ asimptotiskā dispersija ir

$$\gamma^2(t) = \frac{1}{f^2[F_1^{-1}(t)]}t(1-t).$$

Tā kā F_1 nav zināms, tad to novērtē izmantojot empirisko sadalījuma funkciju un blīvuma novērtējums ir uzdots kā

$$f_n(x) = [F_{1n}(x + h_n) + F_{1n}(x - h_n)]/(2h_n).$$

Pieņem, ka $n^{1/2} \sup_t |F_{1n}[F_1^{-1}(t)] - t|$ ir Kolmogorova - Smirnova statistika. Tālāk iegūst, ka

$$n^{1/2} \sup_t |F_{1n}[F_1^{-1}(t)] - t| = n^{1/2} \sup_t |F_{1n}[\Lambda(t) + F_2^{-1}(t)] - t|$$

Tad

$$Pr\{n^{1/2} \sup_t |F_{1n}[\Lambda(t) + F_2^{-1}(t)] - t| \leq k\} = 1 - \epsilon,$$

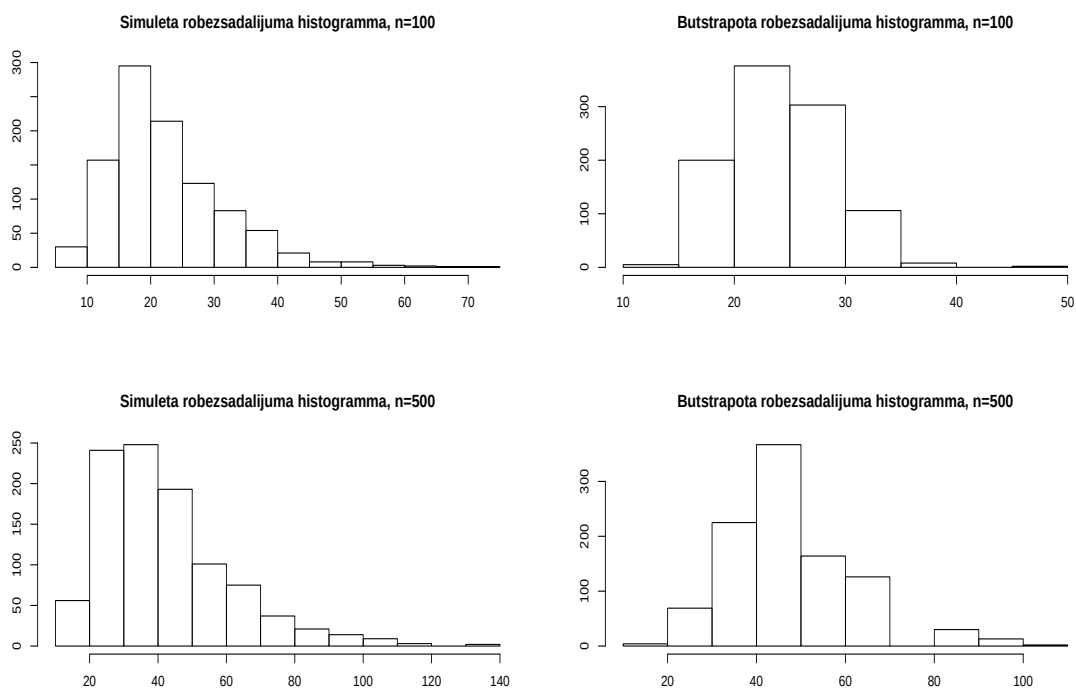
un vienlaicīgās ticamības joslas priekš $\Lambda(t)$ ir iegūtas

$$[F_{1n}^{-1}(t - k/n^{1/2}) - F_2^{-1}(t), F_{1n}^{-1}(t + k/n^{1/2}) - F_2^{-1}(t)].$$

Bet šajā darbā tiek apskatīts, kad F_2 nav zināms. Tā kā nav zināms robežsadalījums, tiek izmantota suprēma statistika

$$\sup_{0 < t < 1} N^{1/2}(\hat{\Lambda}(t) - \Lambda(t)) \tag{4.32}$$

Vienlaicīgās ticamības joslas tiek konstruētas, izmantojot divas iespējas - tiek simulēta un Butstrapota kritiskā vērtība.



4. att.: $N(2, 2)$ un $N(4, 2)$, bet mainīti $n = 100$ un $n = 500$

Attēlā 4. redzam 4 attēlus, kuros ir attēlotas histogrammas vienādām izlasēm $N(2, 2)$ un $N(4, 2)$, bet mainīti izlases apjomi. Pirmajā attēlā ir zīmēta simulēta robežsadaliĵuma histogramma, kur $n = 100$. Otrajā attēlā ir zīmēta Butstrapota robežsadaliĵuma histogramma, $n = 100$. Trešajā attēlā ir zīmēta simulēta robežsadaliĵuma histogramma, kur $n = 500$. Ceturtajā attēlā ir zīmēta Butstrapota robežsadaliĵuma histogramma $n = 500$.

Attēli parāda problemātiku, ka robežsadaliĵumus konverģē aizvien tālāk. Tāpēc jālieto cita statistika, līdzīgi kā reālas izlases gadījumā. Tādēļ tiek mēģināts atrast savādāku pieeju, lai konstruētu vienlaicīgās ticamības joslas kvantiļu funkciju starpībai.

Ideja ir pielietot šādu statistiku

$$\sup_{0 < t < 1} f(F_1^{-1}(t))N^{1/2}[\hat{\Lambda}(t) - \Lambda(t)] \quad (4.33)$$

Robežsadaliĵums konverģēs, bet ir jānovērtē $f(F_1^{-1}(t))$, to varētu darīt līdzīgi, kā vienas izlases gadījumā.

5. Ticamības intervālu konstruēšana, izmantojot empīriskās ticamības metodi

Viena no populārākajām neparametriskajām metodēm ir empīriskās ticamības metode (EL). Par tās pamatlicēju tiek uzskatīts Owen [5], [6], kas iepazīstināja ar EL metodi, izmantojot neparametrisko ticamības funkciju. Galvenā šīs metodes priekšrocība ir tāda, ka EL ticamības intervālu forma netiek balstīta uz pētāmā sadalījuma neesošu simetriju, bet gan uz pētamo datu kopu (tie nav simetriski). Owen parādīja, ka EL metodes statistika tiecas uz χ^2 sadalījumu. Bet Qin un Lawless 1994.gadā savā publikācijā [15] aprakstīja EL metodi vienas izlases gadījumā vispārējā formā, izmantojot novērtējošās funkcijas.

Empīrisko ticamības metodi vispārināt divu izlašu gadījumā ir centušies vairāki. 2000.gadā to mēģināja izdarīt Qin un Zhao [7], kas izveidoja empīriskās ticamības funkcijas kādā punktā divu izlašu vidējo vērtību starpībai un sadalījuma funkciju starpībai. Jānis Valēinis doktora disertācijā [8] ir aprakstījis EL metodi kvantiļu funkciju starpībām, $P - P$ un $Q - Q$ grafikiem, ROC līknēm un strukturālo attiecību modeļiem. Toties 2011. gadā viņš aprakstīja vispārīgā formā empīrisko ticamības metodi divu izlašu gadījumā [16].

5.1. Empīriskās ticamības metode (EL) kvantilēm vienas izlases gadījumā

EL ticamības intervāliem kvantilēm ir liela pārklājuma kļūda ar kārtu $n^{-1/2}$. Bet ar kodolu metožu nogludinātās EL ticamības intervālu pārklājuma kļūda ir ar kārtu n^{-1} . Sākumā tiek apskatīta parastā EL metode ticamības intervālu konstruēšanai no materiāla [17].

EL ticamības intervāli kvantilēm

Pieņemsim, ka $X_1 \dots X_n$ ir iid gadījuma izlase no nezināma sadalījuma $F(x)$ ar attiecīgu blīvuma funkciju $f(x)$, bet $F^{-1}(q)$ ir q -tā kvantile, $0 < q < 1$. Tātad tiks konstruēti ticamības intervāli $\theta_0 = F^{-1}(q)$. $X_1 \dots X_n$ ir sakārtotā izlase un θ_0 novērtējums ir $\hat{\theta}_0 = X_{([nq])}$, kur $[nq]$ ir skaitļa nq veselā daļa. θ_0 var definēt ar sekojošu vienādojumu

$$\int_{-\infty}^{\infty} \phi(x - \theta) dF(x) = 0, \quad (5.1)$$

kur

$$\phi(z) = \begin{cases} -1 & \text{ja } z \leq 0, \\ q/(1-q) & \text{ja } z > 0. \end{cases} \quad (5.2)$$

Empīriskās ticamības attiecība parametram θ_0 ir

$$R(\theta_0) = \sup_{p_1, \dots, p_n} \prod_{i=1}^n (np_i) \quad (5.3)$$

ievērojot, ka

$$p_i \geq 0, \quad i = 1, \dots, n, \quad \sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n p_i \phi(X_i - \theta_0) = 0. \quad (5.4)$$

Maksimizācijas problēmu (5.3) risina ar Lagranža reizinātāju palīdzību. Var pierādīt, ka maksimums tiek sasniegts, kad

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda(\theta_0) \phi(X_i - \theta_0)}, \quad i = 1, \dots, n, \quad (5.5)$$

kur $\lambda(\theta_0)$ apmierina vienādojumu

$$\frac{1}{n} \sum_{i=1}^n \frac{\phi(X_i - \theta_0)}{1 + \lambda(\theta_0) \phi(X_i - \theta_0)} = 0. \quad (5.6)$$

Kad $\theta_0 \in [X_{1:n}, X_{n:n}]$, tad (5.3) atrisinājums $\lambda(\theta_0)$ ir

$$\lambda(\theta_0) = (q - F_n(\theta_0))/q, \quad (5.7)$$

kur F_n ir empīriskā sadalījuma funkcija, kas definēta kā $F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}}$, kur $I\{\cdot\}$ ir indikātorfunkcija. Logaritmiskā empīriskās ticamības attiecība ir

$$\begin{aligned} l(\theta_0) &= -2 \log R(\theta_0) \\ &= 2n \left(F_n(\theta_0) \log \frac{F_n(\theta_0)}{q} + (1 - F_n(\theta_0)) \log \frac{1 - F_n(\theta_0)}{1 - q} \right). \end{aligned} \quad (5.8)$$

Funkcija $\log R(\theta_0)$ no izteiksmes (5.7) ir pakāpienveida funkcija ar lēcieniem datu punktos, tā kā $\log R(\theta_0)$ pieņem galīga skaita vērtības, tad tas padara χ_1^2 aproksimāciju ne īpaši akurātu.

Gludinātie EL ticamības intervāli kvantilēm [9]

Pieņem, ka doti $X_1 \dots X_n$ gadījuma lielumi un $F(x)$ nezināma sadalījuma funkcija. Zinām, ka histogramma aproksimē teorētisko blīvuma funkciju, neatkarīgi no sadalījuma veida. Tādus novērtējumus sauc par neparametriskiem novērtējumiem. Pēc definīcijas

$$f(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h}.$$

Aizstājot nezināmo F ar neparametrisko novērtējumu F_n un atmetot robežu, iegūst blīvuma funkcijas novērtējumu:

$$\hat{f}(x) = \frac{F_n(x+h) - F_n(x)}{h} = \frac{1}{nh} \sum_{i=1}^n 1_{\{x < x_i \leq x+h\}}$$

kur h ir joslas platums, kas ir mazs lielums. Ja intervālu izvēlas simetriski attiecībā pret punktu x , tad iegūst

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{2h} 1_{\{|x - x_i| < h\}} = \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} 1_{\{|\frac{x - x_i}{h}| < 1\}}.$$

Vispārinot šo formulu iegūstam vispārēju formu

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

kur K ir sadalījuma funkcija, kas parasti tiek saukta par kodolu.

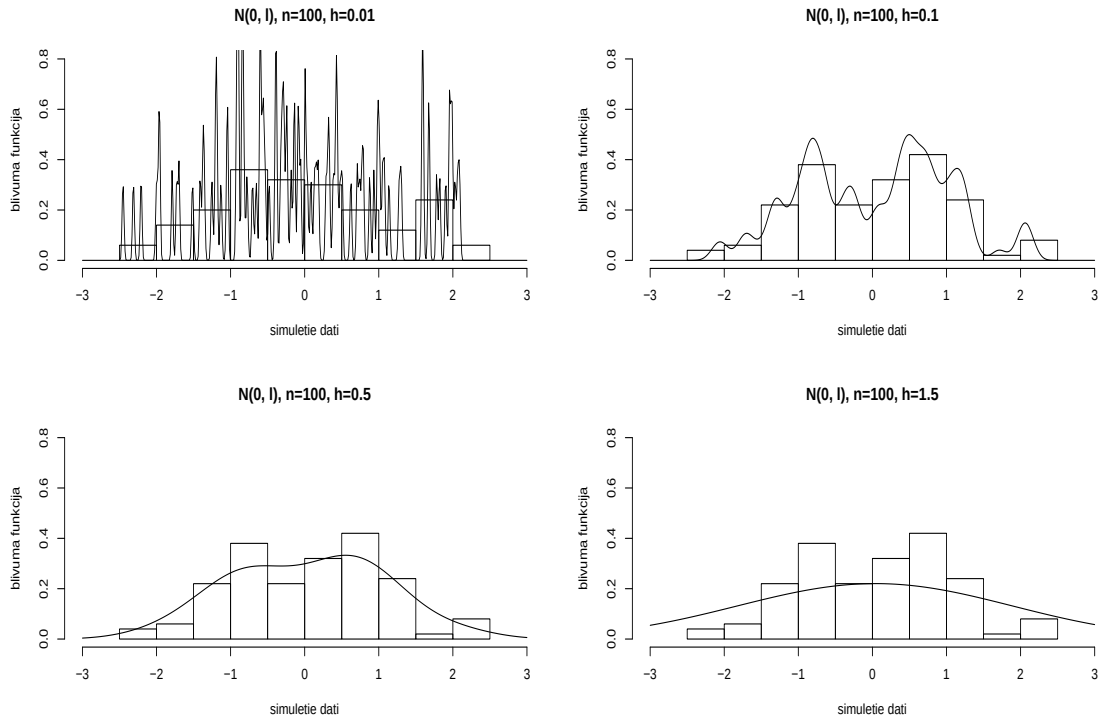
Vispārējā gadījumā būtiskākie pieņēmumi: $h = h_n \rightarrow 0$, $nh \rightarrow \infty$, kad $n \rightarrow \infty$,

$$1) \int K(u) du = 1;$$

$$2) \int u K(u) du = 0; \text{ (simetrija pret nulli)}$$

$$3) \int u^2 K(u) du < \infty.$$

Ir pierādīts, ka kodola izvēle nav būtiskākā, svarīgāka ir joslas platuma h izvēle.



5. att.: Histogramma un tās neparametriskais novērtējums, izmantojot kodolu gludināšanu ar dažādām joslas platuma h izvēlēm.

Attēlā 5. redzams kodolu blīvuma funkcijas novērtējums $N(0, 1)$ sadalījumam ar dažādiem joslas platumiem h . No attēla redzam, ka h izvēle nosaka gludināšanas pakāpi.

Punktveida ticamības intervāli, izmantojot Zhou un Jing ideju [17]

Par joslas platumu tiek ņemts $h = h_n > 0$. Pieņem, ka $h \rightarrow 0$, kad $n \rightarrow \infty$. Tiek izvēlēta kāda Borela mērojama funkcija $K(x)$ kā kodols un definēta $G(t) = \int_{-\infty}^t K(x)dx$. Empīriskā sadalījuma funkcija F_n tiek aizstāta ar nogludinātu empīriskā sadalījuma funkciju

$$\hat{F}_n(x) = \frac{1}{2} \sum_{i=1}^n G\left(\frac{x - X_i}{h}\right). \quad (5.9)$$

Aizstājot $\hat{F}_n(x)$ ar $F_n(\theta_0)$ formulā (5.7), iegūst koriģēto logaritmisko empīriskās ticamības attiecību

$$\hat{l}(\theta_0) = 2n \left(\hat{F}_n(\theta_0) \log \frac{\hat{F}_n(\theta_0)}{q} + (1 - \hat{F}_n(\theta_0)) \log \frac{1 - \hat{F}_n(\theta_0)}{1 - q} \right). \quad (5.10)$$

Nogludinātās empīriskās ticamības intervāls parametram θ_0 ir $I_c = \{\theta_0 : \hat{l}(\theta_0) \leq c\}$, kur c ir konstante, kuru iegūst, izmantojot ticamības līmeni $\alpha = P(\theta \in I_c) = P\{\hat{l}(\theta_0) \leq c\}$

Tālāk tiek uzdoti daži precizējumi.

- (i) $f(x) = F'(x)$. Eksistē f un $f^{(r-1)}$, kur $r \geq 2$, punkta θ_0 apkārtņē, kā arī tie ir nepārtraukti punktā θ_0 . Arī $f(\theta_0) > 0$.
- (ii) kodols $K(\cdot)$ ir ierobežots un tam ir kompakts definīcijas apgabals $[a, b]$. Kādam $[a, b]$ dekompozīcijai $a = u_0 < u_1 < \dots < u_m = b$, kodols $K(\cdot)$ ir vai nu strikti pozitīvs vai strikti negatīvs katrā no intervāliem (u_{j-1}, u_j) , kur $j = 1, \dots, m$.

$$\int u K(u) G(u) du = 0,$$

$$\int u^j K(u) du = \begin{cases} 1, & j = 0, \\ 0, & 1 \leq j \leq r_1 - 1, \\ C, & j = r_1, \end{cases} \quad (5.11)$$

kur C ir kāda ierobežota konstante un r_1 ir vesels skaitlis, kas vienāds vai lielāks par r .

- (iii) $nh/\log n \rightarrow \infty$, nh^2 ir ierobežots $n \rightarrow \infty$.

Teorēma 20. *Nem $\hat{l}(\theta_0)$ tādu kā (5.10) un, pieņemot, ka nosacījumi (i) – (iii) izpildās. Tad, ja $n \rightarrow \infty$*

$$P(\hat{l}(\theta_0) \leq x) - P(\chi_1^2 \leq c) = O(n^{-1}),$$

katram fiksētam x .

No Teorēmas 20 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 $P(\chi_1^2 \leq c) = \alpha$ tad I_{hc} pārklājuma varbūtība aproksimēs α ar kļūdu $O(n^{-1})$, kad $n \rightarrow \infty$.

Ticamības intervāli, izmantojot Chen un Hall ideju [18]

Šajā idejā tiek izmantota kodola gludināšanas metode, lai panāktu rezultātu, ka ticamības intervāla kļūda ir ar kārtu n^{-1} . Tiek parādīts, ka, izmantojot Bartleta korekciju, ticamības intervāla pārklājuma kļūda ir ar kārtu n^{-2} . Šajā idejā $\lambda(\theta_0)$ netiek izteikta kā (5.7) un arī $l(\theta_0)$ nav kā (5.8).

$$l(\theta_0) = 2 \sum_{i=1}^n \log\{1 + \lambda(\theta_0) w_i(\theta_0)\}$$

un $\lambda(\theta_0)$ nosaka no

$$\sum_{i=1}^n w_i(\theta_0) \{1 + \lambda(\theta_0) w_i(\theta_0)\}^{-1} = 0,$$

kur $w_i(\theta_0) = G((\theta_0 - X_i)/h) - q$. Ar K tiek apzīmēts r -tās kārtas kodols, kuram pie $r \geq 2$ un $C \neq 0$, ir spēkā (5.9). Definē $G(t) = \int_{-\infty}^t K(x)dx$. Gadījums, kad $r = 2$ ir visbiežāk sastopams, tad K ir simetriska blīvuma funkcija un $G(x)$ ir sadalījuma funkcija. Ar $f(x) = F'(x)$ tiek apzīmēts funkcijas F pirmais atvasinājums.

Teorēma 21. *Ja K atbilst (5.11), tas ir ierobežots un tam ir kompakts definīcijas apgabals, ja eksistē f un $f^{(r-1)}$ punkta θ_0 apkārtne un tie ir nepārtraukti punktā θ_q , un ja $f(\theta_q) > 0$, un ja dažiem $t > 0$ $nh^t \rightarrow 0$, kad $n \rightarrow \infty$. Tad $\hat{l}(\theta_0)$ ir asimptotisks χ_1^2 sadalījums, ja $nh^{2r} \rightarrow 0$ un $f^{(r-1)}(\theta_q) \neq 0$.*

Teorēmas 21 nosacījumi nodrošina, ka K ir r -tās kārtas kodols, sadalījuma funkcija F ir pietiekami gluda apkārtne θ_q . $f(\theta_q) > 0$ nodrošina, ka asimptotiskās izlases kavntiles dispersija ir ar kārtu n^{-1} . Nosacījums, ka $nh^t \rightarrow 0$, kad $n \rightarrow \infty$, nodrošina, ka joslas platums h nekonverģē uz nulli pārāk lēni.

Ja K ir otrās kārtas kodols ($r = 2$) un $f'(\theta_q) \neq 0$, tad $l(\theta_q)$ ir asimptotiskas χ_1^2 sadalījums tad un tikai tad, ja $h = O(n^{-1/4})$. Ja $f^{(r-1)}(\theta_q) \neq 0$, ir iespējams, ka $l(\theta_q)$ būs asimptotiskais χ_1^2 sadalījums, tomēr nh^{2r} netieksies uz nulli.

Ja izpildās Teorēmas 21 nosacījumi un $nh^{2r} \rightarrow 0$, tad seko, ka

$$P\{l(\theta_q) \leq c\} \rightarrow P(\chi_1^2 \leq c),$$

kur c izvēlēts atbilstoši χ_1^2 tabulām, izpildoties

$$P(\chi_1^2 \leq c) = \alpha.$$

Teorēma 22. *Pieņemot, ka izpildās Teorēmas 21 nosacījumi un joslas platumam h ir spēkā*

$$nh^{2r} \rightarrow 0 \text{ un } nh/\log n \rightarrow \infty \quad (5.12)$$

un $c = c(\alpha)$. Tad priekš

$$P(\theta_q \in I_c) = \alpha + O(n^{-1}) \quad (5.13)$$

pietiekamais nosacījums ir nh^r ir ierobežots, kad $n \rightarrow \infty$. Šis nosacījums ir nepieciešams arī, ja $f^{(r-1)}(\theta_q) \neq 0$. Neatkarīgi no tā, ka $c > 0$, labā puse no (5.13) nevar būt vienāda ar $\alpha + O(n^{-1})$, attiecīgi izvēloties h .

Bartleta korekcija

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

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

kur $\beta = \frac{1}{6}(3\mu_2^{-1}\mu_4 - 2\mu_2^{-2}\mu_3^2)$ un $\mu_j = E[G\{(\theta_q - X_i)/h\} - q]^j$. Tad $l(\theta_q)$ matemātiskā cerība no aproksimētā χ^2 sadalījuma atšķiras par $n^{-1}\beta$. Ja n^3h^{2r} ir ierobežots, tad $l(\theta_q)/(1 + n^{-1}\beta)$ sadalījums atšķiras no χ_1^2 sadalījuma ar kārtu n^{-2} , nevis ar n^{-1} . Praksē β nav zināms un ir jānovērtē. Tādēļ definē

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

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

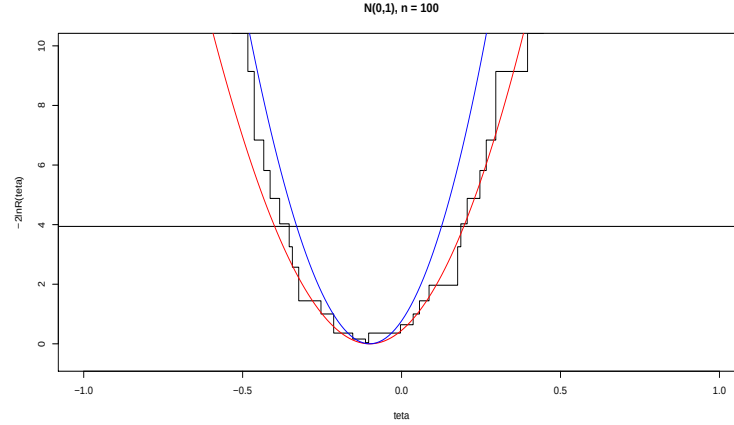
Teorēma 23. *Pieņem, ka izpildās Teorēmas 16. nosacījumi un (5.12), un $c = c(\alpha)$. Tad*

$$P(\theta_q \in I_{d(c,\gamma)}) = \alpha + O(n^{-2}) \quad (5.14)$$

petiekamais nosacījums ir n^3h^{2r} ir ierobežots, kad $\gamma = \beta$ vai $\gamma = \hat{\beta}$, kur $d(c, \gamma) = c(1 + n^{-1}\gamma)$. Šis nosacījums ir nepieciešams arī, lai $f^{(r-1)}(\theta_q) \neq 0$.

Tātad, pielietojot Bartleta korekciju nogludinātās EL ticamības intervāliem, var iegūt pārklājuma kļūdu ar kārtu n^{-2} .

EL metožu salīdzinājums kvantilēm



6. att.: Negludinātā EL un gludinātās EL_{CH}, EL_{ZJ} $-2\ln R(\theta)$ ticamības līknes standartnormālā sadalījuma mediānai

Attēlā 6. redzama logaritmiskās empīriskās attiecības līknes standartnormā $N(0, 1)$ sadalījuma mediānai. Melnā krāsā ir parastā EL metode (EL), sarkanā krāsā ir Zhou un Jing gludinātā EL metode (EL_{ZJ}), bet zilā krāsā ir Chen un Halli gludinātā EL metode (EL_{CH}). Konkrētajā gadījumā redzams, ka Chen un Halli gludinātā metode dod labāku rezultātu. Ticamības intervāli ir šaurāki nekā pārējām divām metodēm: $EL[-0.36; 0.18]$, $EL_{ZJ}[-0.40; 0.19]$, $EL_{CH}[-0.33; 0.12]$

5.2. Empīriskās ticamības metode (EL) kvantilēm divu izlašu gadījumā

Pieņemsim, ka ir dotas divas iid (neatkarīgas un vienādi sadalītas) izlases $X_1 \dots X_n$ un $Y_1 \dots Y_n$ ar attiecīgām sadalījuma funkcijām F_1 un F_2 . $\theta = \theta(t)$ ir funkcija, kas ir saistīta ar vienu no sadalījuma funkcijām F_1 vai F_2 . Tāpat kā vienas izlases gadījumā, pieņemam, ka visa informācija par Δ , θ_0 , F_1 un F_2 ir pieejama nenovirzītu novērtējošo funkciju veidā:

$$E_{F_1} w_1(X, \theta_0, \Delta) = 0, \quad (5.15)$$

$$E_{F_2} w_2(X, \theta_0, \Delta) = 0. \quad (5.16)$$

Piemērs 1. Apzīmējam $\theta_0 = \int x dF_1(x)$ un $\Delta_0 = \int y dF_2(y) - \int x dF_1(x)$. Novērtējošo

funkciju vietā tiek izvēlētas

$$w_1(X, \theta_0, \Delta_0, t) = X - \theta_0, w_2(Y, \theta_0, \Delta_0, t) = Y - \theta_0 - \Delta_0$$

tā, lai būtu spēkā (5.15) un (5.16).

Piemērs 2. (Qin and Zhao, 2000) Divu izlašu vidējo vērtību starpība. Definēsim $\theta_0 = F_1(t)$ un $\Delta_0 = F_2(t) - F_1(t)$. Novērtējošo funkciju vietā izvēlamies

$$w_1(X, \theta_0, \Delta_0, t) = I_{\{X \leq t\}} - \theta_0, w_2(Y, \theta_0, \Delta_0, t) = I_{\{Y \leq t\}} - \theta_0 - \Delta_0$$

tā, lai būtu spēkā (5.15) un (5.16).

Piemērs 3. (Zhou un Jing (2003)) Divu izlašu kvantiļu funkciju starpība. Definē $\theta_0 = F_1^{-1}(t)$ un $\Delta_0 = F_2^{-1}(t) - F_1^{-1}(t)$. Novērtējošo funkciju vietā izvēlamies

$$w_1(X, \theta_0, \Delta_0, t) = I_{\{X \leq \theta_0\}} - t, w_2(Y, \theta_0, \Delta_0, t) = I_{\{Y \leq \theta_0 + \Delta_0\}} - t$$

tā, lai būtu spēkā (5.15) un (5.16).

Teorēma 24. [16] *Izpildoties standarta nosacījumiem*

$$-2\ln EL(\Delta, \hat{\theta}) \xrightarrow{d} \chi_1^2,$$

kur $\hat{\theta}$ ir parametra θ_0 novērtējums.

Izmantojot šo rezultātu, var konstruēt uz EL metodes balstītus punktveida ticamības intervālus parametram Δ formā $\Delta : \{EL(\Delta, \hat{\theta}) > c\}$. Savukārt vienlaicīgās ticamības joslas var iegūt konstruējot punktveida intervālus ar butstrapa metodi. Implementācijai programmā R tiks izmantota pakete EL, ko izveidojuši J. Valienis un E. Cers, kas brīvi pieejama instalēšanai.

6. Pielietojums reālu datu problēmām

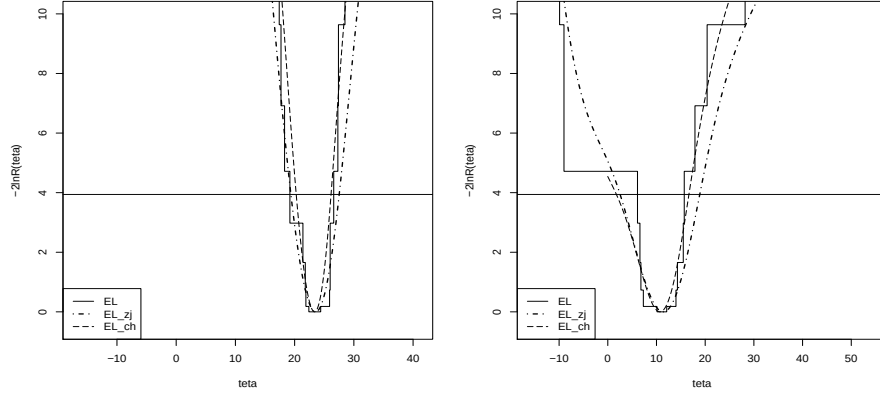
Tiek izmantoti dati no publikācijas [3]. Pirmie dati satur informāciju par 22 žurkām, kas ir turētas 7 dienas ozona vidē un to svara pieaugums gramos tiek apzīmēts ar y_i .

10.1	6.1	20.4	7.3	14.3	15.5	-9.9	6.8	28.2	17.9	-9.0
-12.9	14.0	6.6	12.1	15.7	39.9	-15.9	54.6	-14.7	44.1	-9.0

Otrie dati satur informāciju par 22 žurkām, kas turētas ozona brīvā vidē, šo žurku svara pieaugumu gramos satur x .

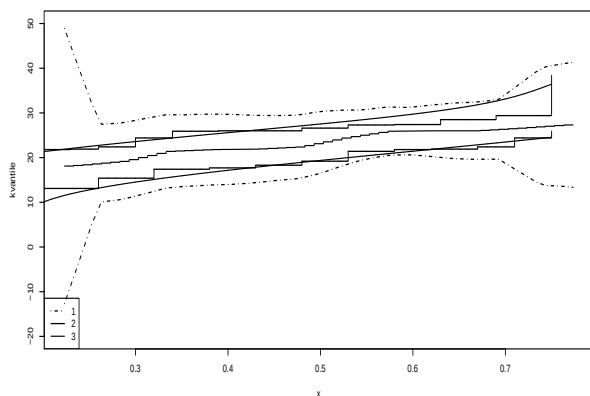
41.1 38.4 24.4 25.9 21.9 18.3 13.1 27.3 28.5 -16.9 26.0
17.4 21.8 15.4 27.4 19.2 22.4 17.7 26.0 29.4 21.4 26.6

Tiek konstruēti ticamības intervāli vidējai vērtībai pie nozīmības līmeņa $\alpha = 0.05$.

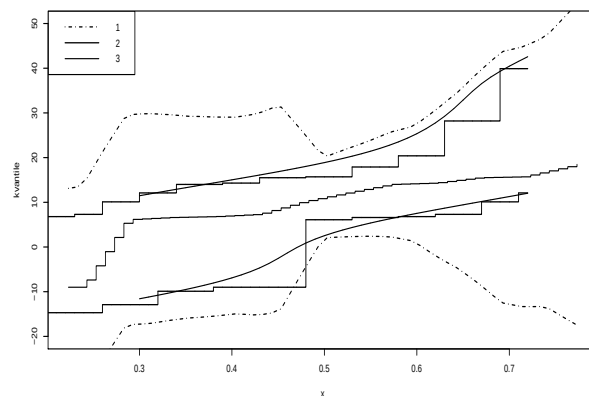


7. att.: Negludinātā EL un gludinātās $EL_{CH}, EL_{ZJ} - 2\ln R(\theta)$ ticamības līknes doto datu mediānai

Attēlā 7. redzama logaritmiskās empīriskās attiecības līknes doto datu mediānai. Priemajā attēlā ir attēloti dati par 22 žurkām, kas ir turētas 7 dienas ozona vidē un to svara pieaugums gramos. Konkrētajā gadījumā redzams, ka Chen un Halli gludinātā metode dod labāku rezultātu. Ticamības intervāli ir šaurāki nekā pārējām divām metodēm: $EL[19.20; 26.60]$, $EL_{ZJ}[19.37; 27.52]$, $EL_{CH}[20.27; 26.15]$. Otrajā attēlā ir attēloti dati par 22 žurkām, kas turētas ozona brīvā vidē, šo žurku svara pieaugumu gramos. Konkrētajā gadījumā redzams, ka negludinātā EL metode dod labāku rezultātu. Ticamības intervāli ir šaurāki nekā pārējām divām metodēm: $EL[6.10; 15.70]$, $EL_{ZJ}[2.53; 18.88]$, $EL_{CH}[1.78; 16.66]$.



(a) Ozona vidē

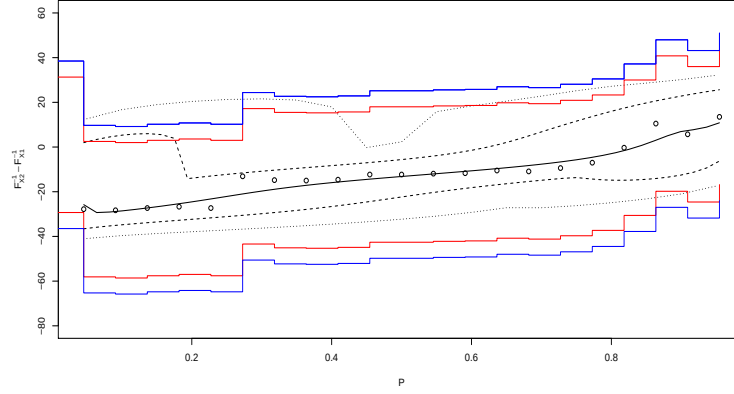


(b) Bez ozona vidē

8. att.: Ticamības intervāli datu kvantilēm

Attēlā 8.a attēloti ticamības intervāli kvantilēm datiem, kas satur informāciju par žurku svara izmaiņu gramos, kad tās uzturējušās ozona vidē, bet attēlā 8.b attēloti ticamības intervāli kvantilēm datiem, kas satur informāciju par žurku svara izmaiņu gramos, kad tās uzturējušās no ozona brīvā vidē. Ar 1. apzīmētas vienlaicīgās ticamības joslas, kas iegūtas, izmantojot (4.10), ar 2. apzīmētas - (4.23), bet ar 3. apzīmētas punktveida ticamības joslas, kas konstruētas izmantojot (5.8)

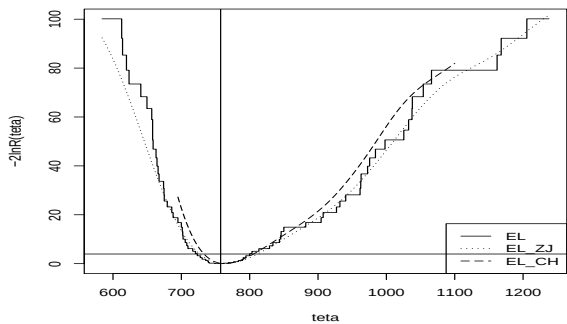
Grafiski zīmēsim vienlaicīgās ticamības joslas kvantiļu funkcijas starpībai un empīriskai ticamības metodei. Lai konstruētu vienlaicīgās ticamības joslas kvantiļu funkcijas starpībai, mums ir jāiegūst kritiskā vērtība. Kritiskā vērtība, kas ir iegūta simulējot, ir 200.9875, bet, izmantojot Butstrapa metodi, ir 248.6564.



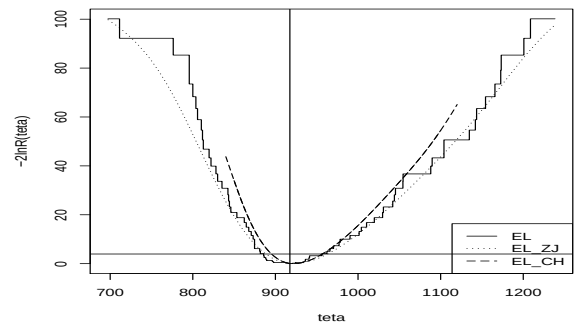
9. att.: Ticamības joslas, kas iegūtas izmantojot kvantiļu funkcijas starpības un empīrisko ticamības modeli

Attēlā 9. redzam ticamības joslas, kas iegūtas izmantojot kvantiļu funkcijas starpības un EL metodi. Tiek izmantoti dati par žurku svara izmaiņām gramos, kas ir turētas ozona vidē un kas turētas no ozona brīvā vidē. Zilā krāsā ticamības joslas tiek zīmētas, izmantojot kritisko vērtību, kas iegūta ar Butstropa metodi. Sarkanā krāsā ticamības joslas tiek zīmētas, izmantojot simulēto kritisko vērtību. Melnā krāsā ir zīmētas ticamības joslas izmantojot EL metodi. Redzam, ka H_0 netiek noraidīta, jo eksistē horizontāla taisne, kas iekļaujas vienlaicīgās ticamības joslās.

Tālāk tiek izmantoti dati no publikācijas [16]. Izlases satur vidējos urbšanas laikus, kuros sasniegts fiksēts dziļums. Tika salīdzināta sausā urbšana un slapjā urbšana.



(a) Sausa urbšana

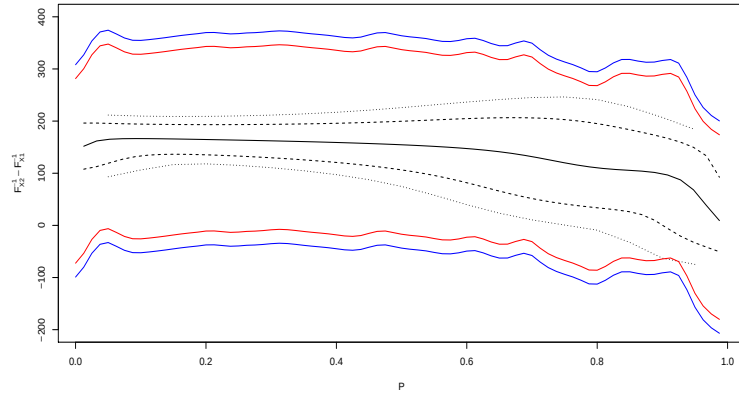


(b) Slapja urbšana

10. att.: Ticamības intervāli datu kvantilēm

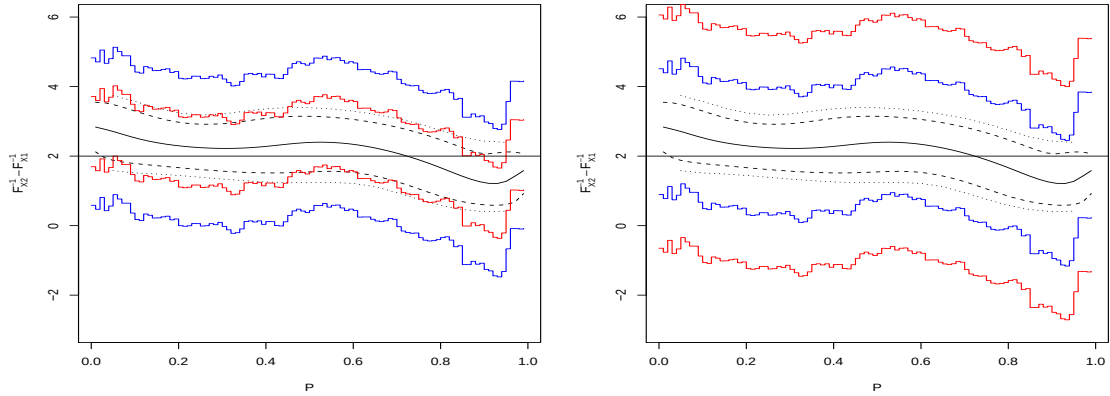
Attēlā 10. redzama logaritmiskās empīriskās attiecības liknes doto datu mediānai. Melnā krāsā ir parastā EL metode (EL). Primajā attēlā ir attēloti dati par sauso urbšanu (80 dati). Konkrētajā gadījumā redzams, ka Chen un Halli gludinātā metode dod labāku rezultātu. Ticamības intervāli ir šaurāki nekā pārējām divām metodēm: $EL[723.33; 800]$, $EL_{ZJ}[726.94; 811.35]$, $EL_{CH}[733.14; 805.57]$. Otrajā attēlā ir attēloti dati par slapju urbšanu. Konkrētajā gadījumā redzams, ka negludinātā EL metode dod labāku rezultātu. Ticamības intervāli ir šaurāki nekā pārējām divām metodēm: $EL[885.67; 956.33]$, $EL_{ZJ}[888.48; 963.02]$, $EL_{CH}[895.30; 956.72]$.

Tālāk tiek konstruētas vienlaicīgās ticamības joslas kvantiļu funkcijas starpībai un empīriskai ticamības metodei datiem par urbšanu. Lai konstruētu vienlaicīgās ticamības joslas kvantiļu funkcijas starpībai, mums ir jāiegūst kritiskā vērtība. Kritiskā vērtība, kas ir iegūta simulējot, ir 2238.893, bet, izmantojot Butstrapa metodi, ir 2575.549.



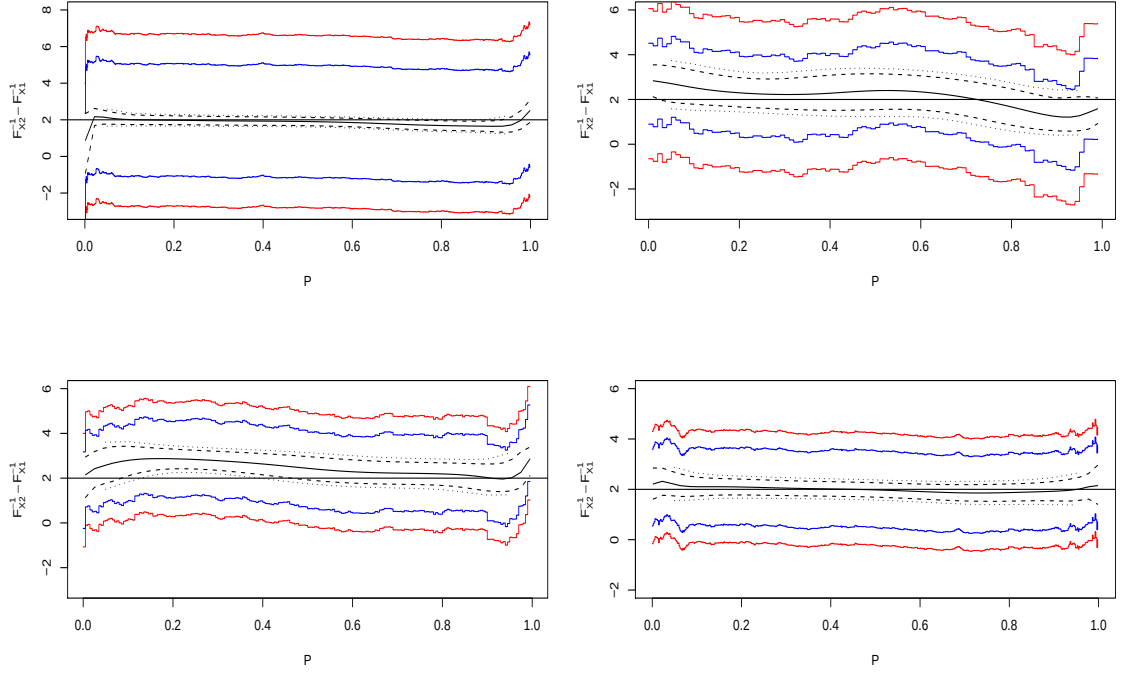
11. att.: Ticamības joslas, kas iegūtas izmantojot kvantiļu funkcijas starpības un empīrisku ticamības modeli

Attēlā 11. redzam ticamības joslas, kas iegūtas izmantojot kvantiļu funkcijas starpības un EL metodi. Tiek izmantoti dati par sausu un slapju urbšanu. Zilā krāsā ticamības joslas tiek zīmētas, izmantojot kritisko vērtību, kas iegūta ar Butstrapa metodi. Sarkanā krāsā ticamības joslas tiek zīmētas, izmantojot simulēto kritisko vērtību. Melnā krāsā ir zīmētas ticamības joslas izmantojot EL metodi. Redzam, ka H_0 netiek noraidīta, jo ir taisne, kas iekļaujas vienlaicīgās ticamības joslās, tas nozīmē, ka mums ir lokācijas modelis.



12. att.: $N(2, 2)$ un $N(4, 2)$, $n = 100$, izmetot 5 izlecējus un $N(2, 2)$ un $N(4, 2)$, $n = 100$

Attēlā 12. redzam ticamības joslas, kas iegūtas izmantojot kvantiļu funkcijas starpības un EL metodi. Zilā krāsā ticamības joslas tiek zīmētas, izmantojot kritisko vērtību, kas iegūta ar Butstropa metodi. Sarkanā krāsā ticamības joslas tiek zīmētas, izmantojot simulēto kritisko vērtību. Pirmajā grafikā zīmētas joslas izmetot arī 5 izlecējus, simulētā kritiskā vērtība ir 14.29185, bet butstrapotā ir 29.97886. Otrā attēlā ir zīmētas ticamības joslas izmantojot visus datus, simulētā kritiskā vērtība ir 19.15858, bet butstrapotā ir 27.78919. No attēliem redzm, ka izmantojot kvantiļu funkcijas starpības, ticamības joslu platumu ietekmē izlecējdati, bet izmantojot EL metodi ticamības joslu paltumi nemainās. Redzam, ka izmantojot EL metodi H_0 tiek noraidīta, bet izmantojot kvantiļu funkcijas starpības H_0 netiek noraidīta, jo ir horizontāla taisne, kas iekļaujas vienlaicīgās ticamības joslās.



13. att.: $N(2, 2)$ un $N(4, 2)$, bet mainīti $n = 100, 200, 500, 1000$

Attēlā 13. redzam 4 attēlus, kuros ir vienādi $N(2, 2)$ un $N(4, 2)$, bet mainīti izlases apjomi. Tiek zīmētas ticamības joslas, kas iegūtas izmantojot kvantiļu funkcijas starpības un EL metodi. Zilā krāsā ticamības joslas tiek zīmētas, izmantojot kritisko vērtību, kas iegūta ar Butstrapa metodi. Sarkanā krāsā ticamības joslas tiek zīmētas, izmantojot simulēto kritisko vērtību. Melnā krāsā ir zīmētas ticamības joslas izmantojot EL metodi. Pirmajā attēlā ir zīmētas ticamības joslas, kur simulētā kritiskā vērtība ir 108.2727, bet butstrapotā ir 122.7485, $n = 800$ Otrajā attēlā zīmētas joslas, kur simulētā kritiskā vērtība ir 19.15858, bet butstrapotā ir 25.56378, $n = 100$. Trešajā attēlā ir zīmētas ticamības joslas, kur simulētā kritiskā vērtība ir 31.20313, bet butstrapotā ir 34.19481, $n = 200$. Ceturtajā attēlā ir zīmētas ticamības joslas, kur simulētā kritiskā vērtība ir 17.39147, bet butstrapotā ir 48.20564, $n = 500$.

Pēc grafikiem redzam, ka ticamības joslas, kur izmantota kritiskā vērtība, kas iegūta simulējot statistiku vai izmantojot Butstrapu metodi, ir platākas par ticamības joslas, kas zīmētas izmantojot EL metodi. To platumu ietekmē izlecējdati, kā arī izlašu apjomi. Redzam, ka H_0 netiek noraidīts, ja izmanto ticamības joslas, kas ir konstruētas, izmantojot kvantiļu funkciju starpību.

Nobeigums

Darbā tika apskatīta grafiskā hipotēžu pārbaude par lokācijas modeli, izmantojot vairākas pieejas. Vienlaicīgās ticamības joslas pašam parametram θ , kā to ir pētījuši Doksums un Sievers savā publikācijā [8]. Tagad tika apskatītas vienlaicīgās ticamības joslas divu kvantiļu funkciju starpībai, izmantojot Petter Laake publikāciju [11], un ar empīrisko ticamības funkcijas metodi, izmantojot Jāņa Valeina publikācijas [8], [16]. H_0 netiek noraidīta, ja eksistē horizontāla taisne, kas iekļaujas vienlaicīgās ticamības joslās. Ja nav taisne, kas iekļautos vienlaicīgās ticamības joslās, tad mēs varam noraidīt H_0 pie izvēlēta nozīmības līmeņa.

Problēmas rada kritiskās vērtības, kas ir nepieciešamas, lai konstruētu vienlaicīgās ticamības joslas empīrisko procesu gadījumā. Konstruējot vienlaicīgās ticamības joslas divu kvantiļu funkciju starpībai tika simulēta un Butstrapota kritiskā vērtība. Butstrapošana šajā gadījumā ir izdevīgāka, jo mums nav zināmi izlašu sadalījumi. Veicot pārbaudi ieguvām rezultātus, ka lokācijas parametra izmaiņas neietekmē kritisko vērtību, bet to ietekmē dispersijas izmaiņas. Veicot simulācijas radās problēmas, jo kritiskās vērtības konverģē uz bezgalību, bet fiksētam n tās konverģē uz kādu skaitli, kas ir labāks rezultāts. Problēmas rada statistikas robežsadalījums (4.32), tāpēc jākonstruē cita statistika (4.33) līdzīgi kā vienas izlases gadījumā.

Salīdzinot visas metodes tika secināts, ka visšaurākos intervālus tomēr dod EL metode. Konstruējot vienlaicīgās ticamības joslas divu kvantiļu funkciju starpībai, rezultātu ietekmē izlases apjoms, izlecējdati un kritiskā vērtība, kas tiek simulēta vai butstrapota.

Darbu turpinot varētu salīdzināt iegūtos rezultātus ar metodi, kad novērtē modeļa parametrus, kā arī implementēt piedāvāto statistiku programmā R.

Izmantotā literatūra un avoti

- [1] Freitag G. and Munk A. On hadamard differentiability in k-sample semiparametric models-with applications to the assessment of structural relationships. *Journal of Multivariate Analysis*, 94(123):158, 2005.
- [2] Cielēns J. Empīrisko procesu pielietojums strukturālo attiecību modeļos. *Diplomdarbs*, pages 14–16, 2010.
- [3] Doksum Kjell A. and Sievers Gerald L. Plotting with confidence: Graphical comparisons of two populations. *Biometrika*, 63:421–433, 1976.
- [4] Laake Petter, Laake Knut, and Aaberge Rolf. On the problem of measuring the distance between distribution functions: Analysis of hospitalization versus mortality. *Biometrika*, 41:515–523, 1985.
- [5] Owen A. Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1):90–120, 1990.
- [6] Owen A.B. *Empirical likelihood*. CRC, 2001.
- [7] Qin Y.S. and Zhao L.C. Empirical likelihood ratio confidence intervals for various differences of two populations. *Systems Sci Math Sci*, 41, 2000.
- [8] Valeinis J. Confidence bands for structural relationship models. *PhD thesis, Niedersächsische Staats-und Universitätsbibliothek Gottingen*, 3:14–16, 2007.
- [9] Valeinis J. Ievads darbam ar statistikas paketi R. *Latvijas Universitāte Fizikas un matemātikas fakultāte*, page 47, 2008.
- [10] Shorak G.R. and Wellner J.A. *Empirical Processes with Applications to Statistics*. John Wiley and Sons, New York, 1986.

- [11] Csorgo M. *Quantile Processes with Statistical Applications*, volume 42. Society for Industrial and Applied Mathematics, 1983.
- [12] Csorgo M. and Revesz P. Strong approximations in probability and statistics. *Academic Press*, 1981.
- [13] Doksum K. A. Empirical probability plots and statistical inference for nonlinear models in the two sample case. *The Annals of Statistic*, 2:267–277, 1974.
- [14] Serfling R.J. Approximation theorems of mathematical statistics. 1980.
- [15] Qin J. and Lawless J. Empirical likelihood and general estimating equations. *The Annals of Statistic*, 22(1):300–325, 1994.
- [16] Cers E. and Valeinis J. Extending the two-sample empirical likelihood. *preprint*, 2011.
- [17] Zhou W. and Jing B. Adjusted empirical likelihood method for quantiles. *Ann. Inst. Statist. Math.*, 21:1166–1181, 2003.
- [18] Chen S. X. and Hall P. Smoothed empirical likelihood condence intervals for quantiles. *The Annals of Statistics*, 21(3):1166–1181, 1993.

Pielikums

```
#Joslas
n<-100; m<-n; N<- n+m; alpha<- 0.05
izlase1<-sort(rnorm(n,0,4)); izlase2<-sort(rnorm(n,1,2))
K<-1.36 #/(sqrt(n)); M<- (n*m)/N
xx<-seq(-3,2,length.out=m)
f<-sort(quantile(izlase2,(ecdf(izlase1)(xx))))

#### S ticam.joslas [S_*(x),S^*(X))
gal1<-sort(floor(n*(((1:m)/m)- (K/(M^(1/2))))))
garums1<-length(gal1[gal1<=0])
gal2<-ceiling((n*(((1:m)/m)+(K/(M^(1/2)))))+1)
garums2<-length(gal2[gal2>m]) # jāzīmē līdz (100 - garums2)
apaksha<-izlase2[gal1[(garums1+1):m]]
augsha<-izlase2[gal2[1:(m-garums2)]]
### G^-1_n{f_m(x)}-x
gals<-floor(n*((1:m)/m))
QQ<-izlase2[gals[1:m]]; X<-izlase1
x1<-floor(min(X)); x2<-ceiling(max(X))
y2<-floor(min(apaksha)); y3<-ceiling(max(augsha))
x1;x2;y2;y3
plot( c(x1,x2), c(y2-1,y3+1), type="n", xlab="(b)", ylab="", )
for (i in 1:m){lines(c(X[i],X[i+1]),
c(QQ[i]-X[i],QQ[i]-X[i+1]))}
for (j in 1:m){lines(c(X[j+1],X[j+1]),
c(QQ[j]-X[j+1],QQ[j+1]-X[j+1]))}; abline(0,0)
#####
Y<-izlase1[(garums1+1):m]
# apaksheja līnija
for (i in 1:m) {lines(c(Y[i],Y[i+1]),
c(apaksha[i]-Y[i],apaksha[i]-Y[i+1]),col = "red",)}
for (i in 1:m){lines(c(Y[i+1],Y[i+1]),
c(apaksha[i]-Y[i+1],apaksha[i+1]-Y[i+1]),col = "red",)}
# augšējā līnija
Y1<-izlase1[1:(m-garums2)]
for (i in 1:(m-garums2)){lines(c(Y1[i],Y1[i+1]),
c(augsha[i]-Y1[i],augsha[i]-Y1[i+1]),col = "red", )}
for (i in 1:(m-garums2)){lines(c(Y1[i+1],Y1[i+1]),
```

```

c(augsha[i]-Y1[i+1],augsha[i+1]-Y1[i+1]),col = "red", )}

#### W ticam.joslas [W_*(x),W^*(x)]
K1<-3.02; lambda<-m/N; c<-(K1^(2))/M
v<-m*(1:n/n); u<-(n*((1:m)/m))
hpl<-sort(ceiling(n*
(((1:m)/m)+(1/2)*c*(1-lambda)*(1-(2*lambda*((1:m)/m)))+(1/2)*(c^(2)*(1-lambda)^(2)
+4*c*((1:m)/m)*(1-((1:m)/m)))^(1/2))/(1+c*(1-lambda)^2))))
garums3<-length(hpl[hpl>m])
augsha2<-izlase2[hpl[1:(m-garums3)]]
hmin<-sort(ceiling(n*
(((1:m)/m)+(1/2)*c*(1-lambda)*(1-(2*lambda*((1:m)/m)))-(1/2)*(c^(2)*(1-lambda)^(2)
+4*c*((1:m)/m)*(1-((1:m)/m)))^(1/2))/(1+c*(1-lambda)^2))))
garums4<-length(hmin[hmin<=0])
apaksha2<-izlase2[hmin[(garums4+1):m]]
# apaksheja linija
Y3<-izlase1[(garums4+1):m]
for (i in 1:m){lines(c(Y3[i],Y3[i+1]),
c(apaksha2[i]-Y3[i],apaksha2[i]-Y3[i+1]),col = "green",)}
for (i in 1:m){lines(c(Y3[i+1],Y3[i+1]),
c(apaksha2[i]-Y3[i+1],apaksha2[i+1]-Y3[i+1]),col = "green",)}
# augšējā līnija
Y4<-izlase1[1:(m-garums3)]
for (i in 1:(m-garums3)){lines(c(Y4[i],Y4[i+1]),
c(augsha2[i]-Y1[i],augsha2[i]-Y4[i+1]),col = "green", )}
for (i in 1:(m-garums3)){lines(c(Y4[i+1],Y4[i+1]),
c(augsha2[i]-Y1[i+1],augsha2[i+1]-Y4[i+1]),col = "green", )}
# stat.lambda<-function(n,loc) {
vec<-sort(dati2)-sort(dati1)-(loc-2)
vec.lab<-sort(abs(vec))[1:(n-5)] # izmetu piecus izlecējus
sqrt(2*n)*(max(abs(vec.lab))) # šeit bija kļūda}

X1 <- dati1; X2 <- dati2
x <- seq(0.05, 0.95, length=19)
y <- EL.smooth("qdiff", X1, X2, x, conf.level=0.95,
simultaneous=TRUE)
## Plot the graph with both pointwise and simultaneous confidence bands
EL.plot("qdiff", X1, X2, conf.level=0.95,ylim=c(-3,6))
lines(x, y$conf.int[1,], lty="dotted")

```

```

lines(x, y$conf.int[2,], lty="dotted")
stat.lambda<-function(n,loc) {
dati1<-rnorm(n,0,2); dati2<-rnorm(n,loc,2)
sqrt(2*n)*(max(abs(sort(dati2)-sort(dati1)-(loc-2)))) # šeit bija kļūda}
N<-1000; rez<-replicate(N,stat.lambda(n, loc)); alpha<-0.05
sort(rez)[(1-alpha)*N]
#hist(rez, xlab="",ylab=" ",main="Simulētā kritiskā vērtība, n=500")
krit<-sort(rez)[(1-alpha)*N]
lines((0:(n-1))/n,sort(dati2)-sort(dati1)-krit/sqrt(2*n),col="red",type="s")
lines((0:(n-1))/n,sort(dati2)-sort(dati1)+krit/sqrt(2*n),col="red",type="s")
abline(2,0) # istā

x1 <- (0:(n-1))/n
y1 <- sort(dati2)-sort(dati1)+krit/sqrt(2*n)
b.spl <- smooth.spline(x1, y1)
lines(b.spl, col = "red")
x <- (0:(n-1))/n
y <- sort(dati2)-sort(dati1)-krit/sqrt(2*n)
b.spl <- smooth.spline(x, y)
lines(b.spl, col = "red")

# Butstraps
stat.lambda.b<-function(n, dati1, dati2)
{lok<-mean(dati2)-mean(dati1) # šeit bija kļūda
izl1.b<-sample(dati1,replace=T)
izl2.b<-sample(dati2,replace=T)
vec1<-sort(izl2.b)-sort(izl1.b)-lok
vec.lab1<-sort(abs(vec1))[1:(n-5)] # izmetu piecus izlecējus
sqrt(2*n)*(max(abs(vec.lab1)))
#sqrt(2*n)*(max(abs(sort(izl2.b)-sort(izl1.b) - lok)))) }
stat.lambda.b<-function(n, dati1, dati2)
{lok<-mean(dati2)-mean(dati1) # šeit bija kļūda
izl1.b<-sample(dati1,replace=T)
izl2.b<-sample(dati2,replace=T)
sqrt(2*n)*(max(abs(sort(izl2.b)-sort(izl1.b) - lok))))}

N<-1000
rez.b<-replicate(N,stat.lambda.b(n, dati1, dati2)); alpha<-0.05
sort(rez.b)[(1-alpha)*N]

```

```

#hist(rez.b, xlab="",ylab=" ", main="Butstropotā kritiskā vērtība, n=500")
krit.b<-sort(rez.b)[(1-alpha)*N]
lines((0:(n-1))/n,sort(dati2)-sort(dati1)-krit.b/sqrt(2*n), type="s",col="blue")
lines((0:(n-1))/n,sort(dati2)-sort(dati1)+krit.b/sqrt(2*n), type="s",col="blue")
points((0:(n-1))/n,sort(dati2)-sort(dati1))

x1 <- (0:(n-1))/n
y1 <- sort(dati2)-sort(dati1)+krit.b/sqrt(2*n)
b.spl <- smooth.spline(x1, y1)
lines(b.spl, col = "green")
x <- (0:(n-1))/n
y <- sort(dati2)-sort(dati1)-krit.b/sqrt(2*n)
b.spl <- smooth.spline(x, y)
lines(b.spl, col = "green")

# grafiks -2lnR(teta) tic.līknes mediānai
n<-100; dati<-rnorm(n,1,4)
fn<-ecdf(dati)
stat<-function(theta,q)
{2*n*(fn(theta)*log(fn(theta)/q)+
(1-fn(theta))*log((1-fn(theta))/(1-q)))}
stat<-Vectorize(stat,vectorize.args="theta")
median(dati); q<-0.5
xx<-seq(min(dati)+0.1,max(dati)-0.1,by=0.01)
plot(xx,stat(xx,q), ylim=c(-0.5,10), xlim=c(-1,1), xlab="teta", ylab="-2lnR(teta)",
type="s", main="N(0,1), n = 100"); abline(h=3.94)
l1<-min(dati); l2<-quantile(dati,q)[[1]]
l3<-max(dati); abline(v=l2)
uniroot(function(x) stat(x,q)-3.94, c(l1+0.1,l2))$root
uniroot(function(x) stat(x,q)-3.94, c(l2,l3-0.1))$root

#Zhou, Jing
library(sm); h<-hcv(dati)
K<-function(x) dnorm(x) #1/sqrt(2*pi)*exp(-x^2/2)
bliv1<-function(x,dati,h)
{n<-length(dati)
(1/(n*h))*sum(K((x-dati)/h))}
bliv<-Vectorize(bliv1,vectorize.args="x")
sad.fun<-function(y,dati,h)

```

```

{integrate(function(x) bliv(x,dati,h),-20,y)$value}
stat2<-function(theta,q)
{2*n*(sad.fun(theta,dati,h)*log(sad.fun(theta,dati,h)/q)+
(1-sad.fun(theta,dati,h))*log((1-sad.fun(theta,dati,h))/(1-q)))}
stat2<-Vectorize(stat2,vectorize.args="theta")
lines(xx,stat2(xx,q),lty=1)
uniroot(function(x) stat2(x,q)-3.94, c(11+0.1,12))$root
uniroot(function(x) stat2(x,q)-3.94, c(12,13-0.1))$root
#abline(h=1:6, lty=1:6)

#Chen, Hall
w<-function(x) pnorm((x-dati)/h)-q
f.lam<-function(lambda, theta)
{sum(w(theta)/(1+lambda*w(theta)))}
f.lam2<-Vectorize(f.lam,vectorize.args="lambda")
f.lam2<-Vectorize(f.lam2,vectorize.args="theta")
fun<-function(theta)
{sdati<-sort(w(theta))
lamda1<-(1-1/n)/sdati[1]; lamda2<-(1-1/n)/sdati[n]
lambda<-uniroot(function(x) f.lam2(x,theta), c(lamda1,lamda2))$root
2*sum(log(1 + lambda*w(theta)))
fun<-Vectorize(fun)
x2<-seq(-0.8,0.5,by=0.01)
lines(x2,fun(x2), col="blue")
uniroot(function(x) fun(x)-3.94, c(-0.39,12))$root
uniroot(function(x) fun(x)-3.94, c(12,0.5))$root

### Pēc M.Corgo (ar blīvuma funkcijas novērtējumu)
n<-100; set.seed(1)
dati<-rnorm(n); alpha1<-0.5 #jāizvēlas <1/2
tt <- seq(n^{-alpha1}+0.01,1-n^{-alpha1}-0.01,by=0.01)
kvantile <- quantile(dati,tt)
plot(tt,kvantile,xlab="x", ylab="kvantile", ylim=c(-3,3), xlim=c(0,1), type="s")
kr.v<-1.36; alpha<-0.95
nov1<-function(x){
(2*(n^{-alpha1}))^{(-1)*(quantile(dati,x+n^{-alpha1}) - quantile(dati,x-n^{-alpha1}))}
nov<-Vectorize(nov1)
lines(tt,kvantile+kr.v/sqrt(n)*nov(tt),lty=4, lwd=2)
lines(tt,kvantile-kr.v/sqrt(n)*nov(tt),lty=4, lwd=2)

```

```

### Pēc M.Corgo (taisnas vienlaicīgās joslas)
xx <- seq(n^{-0.5}+0.09,1-n^{-0.5}-0.09,by=0.01)
kvantile2<- quantile(dati,xx)
nov2<-function(x){quantile(dati,x+kr.v*n^{-0.5})}
nov3<-Vectorize(nov2);nov4<-function(x){
quantile(dati,x-kr.v*n^{-0.5})}
nov5<-Vectorize(nov4)
lines(xx,nov3(xx)); lines(xx,nov5(xx))

kvan<-seq(0.01,0.99,by=0.01)
m<-length(kvan)
for (i in 10:m){ l1<-min(dati); l2<-quantile(dati,kvan[i])[[1]]
l3<-max(dati)
X1<-uniroot(function(x) stat(x,kvan[i])-3.94, c(l1+0.1,l2))$root
X2<-uniroot(function(x) stat(x,kvan[i])-3.94, c(l2,l3-0.1))$root
lines(c(kvan[i],kvan[i+1]),c(X1[i],X1[i+1]),type="s")
lines(c(kvan[i],kvan[i+1]),c(X2[i],X2[i+1]),type="s")}
```

Diplomdarbs “Kvantiļu funkciju analīze ar pielietojumu strukturālo attiecību modeļiem” izstrādāts LU Fizikas un Matemātikas fakultātē.

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

Autors: Odeta Lorence

(paraksts)

(datums)

Rekomendēju darbu aizstāvēšanai.

Vadītājs: doc. Dr.math. Jānis Valeinis

(paraksts)

(datums)

Recenzents: Doc. Nadežda Siņenko

(paraksts)

(datums)

Darbs iesniegts Matemātikas nodaļā _____
(datums)

(darbu pieņēma)

Darbs aizstāvēts gala pārbaudījuma komisijas sēdē

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

Komisijas sekretārs/-e: _____
(Vārds, Uzvārds) (paraksts)