

LATVIJAS UNIVERSITĀTE  
FIZIKAS UN MATEMĀTIKAS FAKULTĀTE  
MATEMĀTIKAS NODAĻA

**GRAFISKĀ HIPOTEŽU PĀRBAUDE LOKĀCIJAS  
MODELIM**

KURSA DARBS

Autors: **Odetta Lorence**

Stud. apl. ol07001

Darba vadītājs: doc. Dr.math. Jānis Valeinis

RĪGA 2010

# Saturs

Ievads	2
1. Hipoztēžu pārbaude fiksēta lokācijas parametra gadījumā	3
2. Vienlaicīgās ticamības joslas lokācijas parametram	6
2.1. Vienlaicīgās ticamības joslas . . . . .	7
3. Vienlaicīgās ticamības joslas divu kvantiļu funkciju starpībai	12
4. Pielietojam datus	15
Nobeigums	17
Izmantotā literatūra un avoti	18
Programmas kodi	19

# Ievads

Kursa darba mērķis ir pārbaudīt grafiski hipotēzi par lokācijas modeli. Tas ir

$$H_0 : F(x) = G(x + \theta) \quad \forall x, \tag{0.1}$$

kur  $\theta$  ir lokācijas parametrs,  $F$  un  $G$  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.

Ir vairākas iespējas, kā pārbaudīt hipotēzi. Pirmkārt izmantojot fiksētu parametru  $\theta$ , otrai izlasei atņemot šo parametru var reducēt hipotēzi par divu sadalījuma funkciju pārbaudi, tas ir

$$H_0 : F(x) = G(X).$$

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 pārbaudi  $\theta$  ir jānovērtē. To ir pētījis Juris Cielēns savā diplomdarbā [1].

Darbā tiks analizētas divas citas pieejas. Pirmkārt tiks konstruētas vienlaicīgās ticamības joslas pašam parametram  $\theta$ . Otrkārt tiks konstruētas divu izlašu kvantiļu funkcijas starpība.

Eksistē arī citi modeļi, kurus sauc par vispārīgiem strukturāliem modeļiem, ko ieviesa Freitag un Munk [2]. Šie modeļi ietver lokācijas - skalēšanas modeli un Lēmaņa alternatīvo modeli. Kursa darbā šādi modeļi netiks apskatīti, bet tos paredzēts analizēt diplomdarbā.

Darbs sastāv no četrām nodaļām un pielikuma. Pirmajā nodaļā tiek aprakstīta hipotēžu pārbaude fiksēta parametra  $\theta$  gadījumā. Otrajā nodaļā tiek apskatītas vienlaicīgās ticamības joslas lokācijas parametram  $\theta$ . Trešajā nodaļā tiek analizētas vienlaicīgās ticamības joslas divu kvantiļu funkcijas starpībai. Ceturtajā nodaļā tiks analizēta hipotēze (0.1) konkretiem datiem, izmantojot abas pieejas. Pielikums satur programmas kodu.

# 1. Hipotēžu pārbaude fiksēta lokācijas parametra gadījumā

Aplūkosim divas izlases  $X_1, \dots, X_m$  un  $Y_1, \dots, Y_n$  ar sadalījuma funkcijām  $F$  un  $G$ . Lokācijas modelis ir  $F(x) = G(x + \theta)$  visiem  $x$ . Fiksētam lokācijas parametram hipotēžu pārbaude reducējas, kā

$$H_0 : F_1(x) = F_2(x) \text{ pret } H_1 : F_1(x) \neq F_2(x) \text{ visiem } x$$

Šo hipotēzi var pārbaudīt ar Kolmogorova - Smirnova testu vai var konstruēt vienlaicīgos tīcamības intervālus varbūtību-varbūtību ( $P - P$ ) un kvantiļu-kvantiļu ( $Q - Q$ ) grafikiem. Kolmogorova-Smirnova testa statistika ir formā

$$\sqrt{n}D_n = \sqrt{n} \sup_{-\infty < x < \infty} |F_n(x) - F(x)|,$$

kur  $n$  izlases apjoms,  $F_n(x)$  - empīriskā sadalījuma funkcija.

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

Ja vienkārša hipotēze  $H_0$  ir spēkā, tad no teorijas zināms, ka

$$\sqrt{n}D_n \xrightarrow[n \rightarrow \infty]{} \sup_{0 < t < 1} |B(t)|,$$

kur  $B(\cdot)$  apzīmē Brauna tiltu. Savukārt, ja spēkā  $H_1$  (tas ir  $H_0$  nepareiza), tad

$$\sqrt{n}D_n \xrightarrow[n \rightarrow \infty]{} \infty.$$

**Definīcija 1.** Par empīrisko varbūtību-varbūtību ( $P - P$ ) grafiku sauc funkciju

$$PP_{nm}(t) = F_{1n}(F_{2m}^{-1}(t)),$$

kur  $0 < t < 1$ ,  $F_{1n}$  - pirmās izlases empīriskā sadalījuma funkcija,  $F_{2m}^{-1}$  - otrās izlases empīriskā kvantiļu funkcija.

**Definīcija 2.** Par empīrisko kvantiļu-kvantiļu ( $Q - Q$ ) grafiku sauc funkciju

$$QQ_{nm}(t) = F_{1n}^{-1}(F_{2m}(t)),$$

kur  $0 < t < 1$ ,  $F_{1n}^{-1}(t) := \inf\{x : F_n(x) \geq t\}$ ,  $0 < t < 1$  - pirmās izlases empīriskā kvantiļu funkcija,  $F_{2m}$  - otrās izlases empīriskā sadalījuma funkcija.

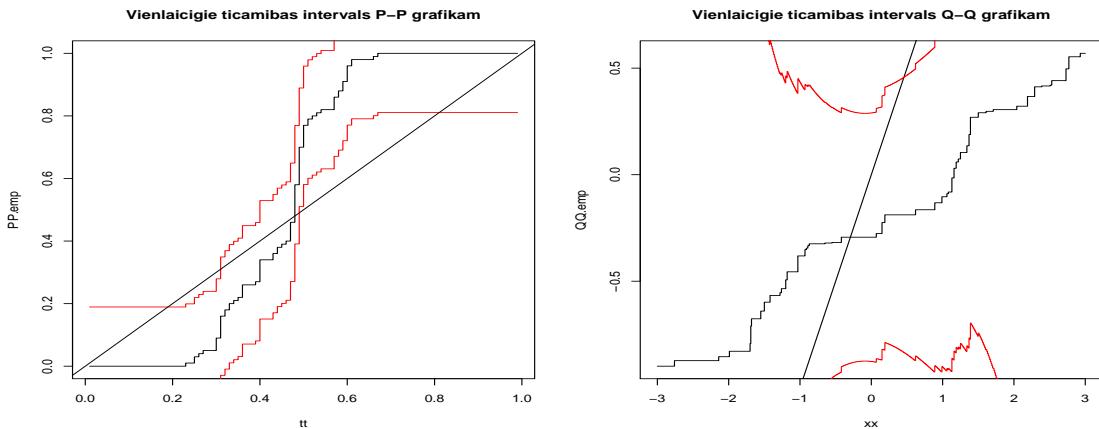
Ticamības intervāli tiek konstruēti, izmantojot J. Cielēna diplomdarbu [1]. Kritiskā vērtība tiek aprēķināta, veicot Brauna tiltu simulācijas ( $\alpha = 0.95, c = 1.89$ )

1.  $P - P$  procesam

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

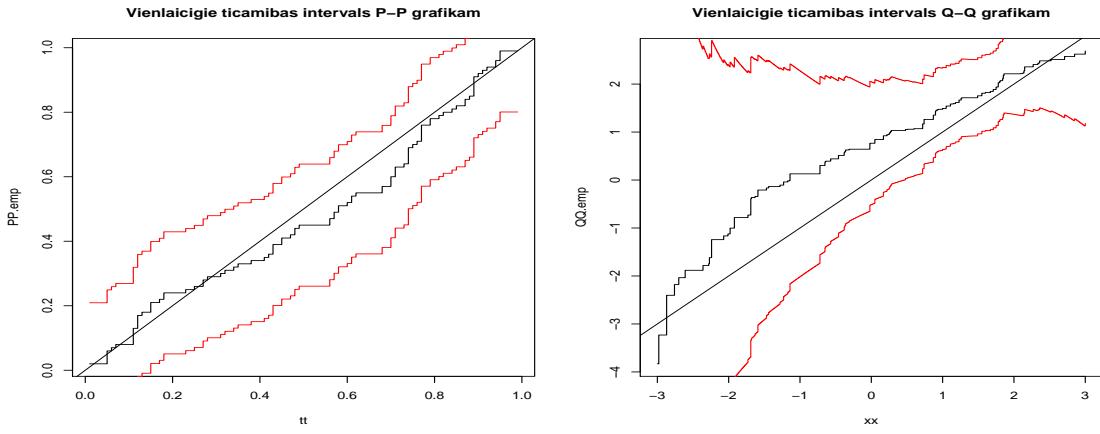
2.  $Q - Q$  procesam

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



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

Attēlā 1.1. pārbaudīta hipotēze par divu sadalījumu vienādību. Izlases  $X$  un  $Y$  sadaļitas attiecīgi  $N(0, 1)$  un  $N(1, 4)$ , bet  $n = 100$ . a) grafikā attēlotas vienlaicīgās ticamības joslas varbūtību-varbūtību ( $P - P$ ) grafikam. b) grafikā attēlotas vienlaicīgās ticamības joslas kvantiļu-kvantiļu ( $Q - Q$ ) grafikam. No grafikiem var secināt, ka  $H_0$  par divu izlašu vienādu sadalījumu noraidam, jo grafikos redzamā diagonāle iziet ārpus ticamības intervāliem kaut vienā punktā.



1.2. att.  $N(1, 2)$  pret  $N(1, 2)$ ,  $n = 100$

Attēlā 1.2. izlases  $X$  un  $Y$  sadalītas attiecīgi  $N(1, 2)$  un  $N(1, 2)$ , bet  $n = 100$ . a) grafikā attēlotas vienlaicīgās ticamības joslas varbūtību-varbūtību ( $P - P$ ) grafikam. b)grafikā attēlotas vienlaicīgās ticamības joslas kvantiļu-kvantiļu ( $Q - Q$ ) grafikam. No grafīkiem var secināt, ka  $H_0$  par divu izlašu vienādu sadalījumu nenoraidam, jo grafikos redzamā diagonāle neiziet ārpus ticamības intervāliem.

## 2. Vienlaicīgās ticamības joslas lokācijas parametram

Lokācijas modeļa  $F(x) = G(x + \theta)$  visiem  $x$  vietā mēs apskatīsim vispārējo gadījumu  $F(x) = G(x + \Delta(x))$ , kur  $\Delta(x)$  ir kāda funkcija. Ja  $X$  ir kontroles rezultāti un  $Y$  ir eksperimentam pakļautie rezultāti, tad  $\Delta(x)$  ir lokācijas parametrs. Saskaņā ar vispārējiem nosacījumiem ( $\Delta_s(x) + x$  jābūt nedilstošam)  $x$  ir vienīgā funkcija, kas apmierina  $X + \Delta(X) \sim Y$ , kur  $\sim$  nozīmē sadalījumu. Tā tad  $\Delta(\cdot)$  ir lokācijas parametrs, kas nepieciešams, lai no  $X$  iegūtu  $Y$ .

Pieņemot, ka  $F$  un  $G$  ir nepārtrauktas un  $F^{-1}(u) = \inf\{x : F(x) \geq u\}$  ir funkcijas  $F$  inversā funkcija. Tad mēs varam rakstīt

$$\Delta(x) = G^{-1}\{F(x)\} - x.$$

Lokācijas modelis  $F(x) = G(x + \theta)$  faktiski ir konstantei  $\theta$ , kur  $\Delta(x) = \theta$ . Funkcijas  $G^{-1}\{F(x)\}$  inversā funkcija ir  $G_n^{-1}\{F_m(x)\}$ , kur  $F_m$  un  $G_n$  ir empīriskās sadalījuma funkcijas.

Doksums un Sievers savā publikācijā [3] izvirzīja vairākus jautājumus (i) vai apstrāde ir izdevīga visiem izlases locekļiem, 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, tas ir  $\Delta(x) = \theta$ , ja ņem kādai  $\theta$ , visiem  $x$ ? (iv) ja nē, tad vai  $\Delta(x) = \alpha + \beta x$  priekš kādas  $\alpha$  un  $\beta$ , 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 neviens horizontāla līnija neiekļaujas ticamības joslās un (iv) arī noraida, ja taisna līnija neiekļaujas ticamības joslās.

**Piezīme 1.** Ja kaut viena taisa līnija iekļaujas  $\Delta_s$  funkcijas ticamības joslās, tad mēs varam nenoraidīt hipotēzi, ka  $\Delta_S(x) = \alpha + \beta x$  visiem  $x$  ar ticamību  $(1 - \alpha)$ , kur  $\alpha, \beta$  ir kādas konstantes. Saskaņā ar hipotēzi  $H_0 : F(x) = G(\alpha + \beta x)$  mēs zinām, ka  $Q - Q$  grafiks ir taisne. Pieņemam, ka  $X \sim N(\mu, \sigma^2)$ . Tād tad

$$P(X \leq q) = P\left(\frac{X - \mu}{\sigma} \leq \frac{q - \mu}{\sigma}\right) = P\left(Z \leq \frac{q - \mu}{\sigma}\right),$$

kur  $Z \sim N(0, 1)$ . Tas ozīmē, ka kvantiles punktam  $q$  no  $N(\mu, \sigma^2)$  atbilst kvantiles punkts  $(q - \mu)/\sigma$  no  $N(0, 1)$ . Nepārprotami hipotēzei  $H_0$  atbilst arī funkcija  $\Delta_S$ , kas ir taisna līnija. Saskaņa ar  $H_0$  vienlaicīgās ticamības joslām priekš funkcijas  $\Delta_S$  jāiekļauj katrs punkts šai līnijai ar ticamību  $(1 - \alpha)$ . Tāpēc, ja nav taisnas līnijas, kas iederas ticamības joslās, tad var noriadīt  $H_0$  hipotēzi.

## 2.1. Vienlaicīgās ticamības joslas

Aplūkojam statistiku  $T(F_m, G_n)$  ar īpašību, ka vienādība  $T(F_m, G_n) \leq K$  ir līdzvērtīga

$$h_*F_m(x) \leq G_n(x) \leq h^*F_m(x) \quad (2.1.1)$$

visiem  $x$ , dažādām funkcijām  $h_*$  un  $h^*$ . Parasti šīs funkcijas ir nedilstošas. Doksums un Sievers [3] aplūkoja vienlaicīgās ticamības joslas balstītas uz Kolmogorova - Smirnova statistiku.

$$T(F_n, G_m) = D_n = \sqrt{M} \sup_x |F_n(x) - G_m(x)|. \quad (2.1.2)$$

Tad  $h_*(x) = x - K/M^{1/2}$  un  $h^*(x) = x + K/M^{1/2}$ .

No (2.1.1) mēs iegūstam vienlaicīgās ticamības joslas. Nemam  $G_n^{-1}(u)$  un  $G_n^{-I}(u) = \sup\{x : G_n(x) \leq u\}$  par kreiso un labo inverso  $G_n$  funkciju. Pieņemsim, ka  $K$  tiks izvēlēts tā, ka

$$P_{F=G}\{T(F_m, G_n) \leq K\} = 1 - \alpha. \quad (2.1.3)$$

Tālāk vienlaikus visiem  $x$

$$\begin{aligned} 1 - \alpha &= P_{F=G}\{T(F_m, G_n) \leq K\} = P_{F,G}\{T(F_m, G_{\Delta,n}) \leq K\} \\ &= P_{F,G}[h_*\{F_m(x)\} \leq G_{\Delta,n}(x) \leq h^*\{F_m(x)\}] \\ &= P_{F,G}[h_*\{F_m(x)\} \leq G_n\{\Delta(x) + x\} \leq h^*\{F_m(x)\}] \end{aligned}$$

$$= P_{F,G}(G_n^{-1}[h_*\{F_m(x)\}] - x \leq \Delta(x) < G_n^{-I}[h^*\{F_m(x)\}] - x).$$

**Teorēma 2.** Ja (2.1.1) un (2.1.3) ir spēkā, tad

$$[G_n^{-1}[h_*\{F_m(x)\}] - x, G_n^{-I}[h^*\{F_m(x)\}] - x) \quad (2.1.4)$$

dod vienlaicīgās ticamības joslas funkcijai  $\Delta(x)$ , kad  $x$  mainās no  $-\infty$  līdz  $\infty$ .

**Piezīme 3.**  $(1 - \alpha)$  vienlaicīgās ticamības joslas priekš  $\Delta(x)$  tiek uzdotas

$$[G_n^{-1}\{F_m(x) - K_{S,\alpha}/M^{1/2}\} - x, G_n^{-1}\{F_m(x) + K_{S,\alpha}/M^{1/2}\} - x).$$

Šī josla, kura nosaukta par  $S$  joslu un apzīmēta ar  $[S_*(x), S^*(x))$ , tika iegūta pēc Switzer [4] un vajadzētu aizstāt ar līdzīgām joslām, kas iegūtas pēc Doksuma [5].

Apzīmēsim 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  $t$ ;  $X_1 < \dots < X_m$  un  $Y_1 < \dots < Y_m$ ,  $Y(j) = -\infty (j < 0)$  un  $Y(j) = \infty (j \geq n+1)$ . Tad vienlaicīgos ticamības intervālus (2.1.4) varam pārrakstīt:

$$[\Delta_*(x), \Delta^*(x)) = \left[ Y\left(\left\langle nh_*\left(\frac{i}{m}\right) \right\rangle\right) - x, Y\left(\left[nh^*\left(\frac{i}{m}\right)\right] + 1\right) - x \right) \quad (2.1.5)$$

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

Tālāk attēlā attēlotas  $S$  vienlaicīgās ticamības joslas

$$[S_*(x), S^*(x)) = \left[ Y\left\{\left\langle n\left(\frac{i}{m} - K_{S,\alpha}/M^{\frac{1}{2}}\right) \right\rangle\right\} - x, Y\left\{\left[n\left(\frac{i}{m} + K_{S,\alpha}/M^{\frac{1}{2}}\right)\right] + 1\right\} - x \right), \quad (2.1.6)$$

kur  $X$  un  $Y$  sadalīti attiecīgi  $N(0, 1)$  un  $N(1, 4)$ , bet  $m = n = 100$  un  $\alpha = 0, 05$ .

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

$$W_N = W_N(F_M, G_N) = M^{\frac{1}{2}} \sup_{\{x:a \leq F_m(x) \leq b\}} \frac{|F_m(x) - G_m(x)|}{\Psi\{H_N(x)\}}, \quad (2.1.7)$$

kur  $H_N(x) = \lambda F_m(x) + (1 - \lambda)G_n(x)$ ,  $\lambda = m/N$  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

$$M^{\frac{1}{2}}\{F_m(x) - G_n(x)\}/\Psi\{H_N(x)\}$$

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

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

priekš  $x$ , kas atbilst  $a \leq F_m(x) \leq b$ . Apzīmēsim  $c = K^2/M$ ,  $v = G_n(x)$ , tad nevienādību varam pārrakstīt kā  $d(v) \leq 0$ , kur

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

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

$$h \pm (u) = \frac{u + \frac{1}{2}c(1 - \lambda)(1 - 2\lambda u) \pm \frac{1}{2}\{c^2(1 - \lambda)^2 + 4cu(1 - u)\}^{\frac{1}{2}}}{1 + c(1 - \lambda)^2} \quad (2.1.8)$$

Izriet, ka ar varbūtību  $(1 - \alpha)$ ,  $G_n(x)$  ir diapozonā

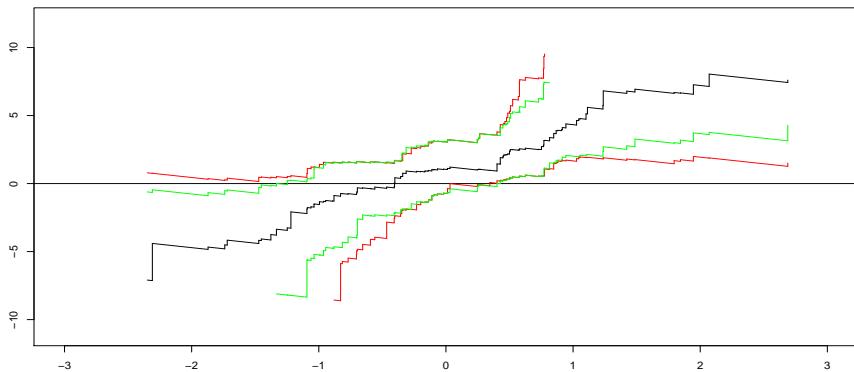
$$h^- \{F_m(x)\} \leq G_n(x) \leq h^+ \{F_m(x)\}$$

priekš visiem  $x \in \{x : a \leq F_m(x) \leq b\}$ .

**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

$$[G_n^{-1}[h^-\{F_m(x)\}] - x, G_n^{-1}[h^+\{F_m(x)\}] - x), x : a \leq F_m(x) \leq b\}. \quad (2.1.9)$$

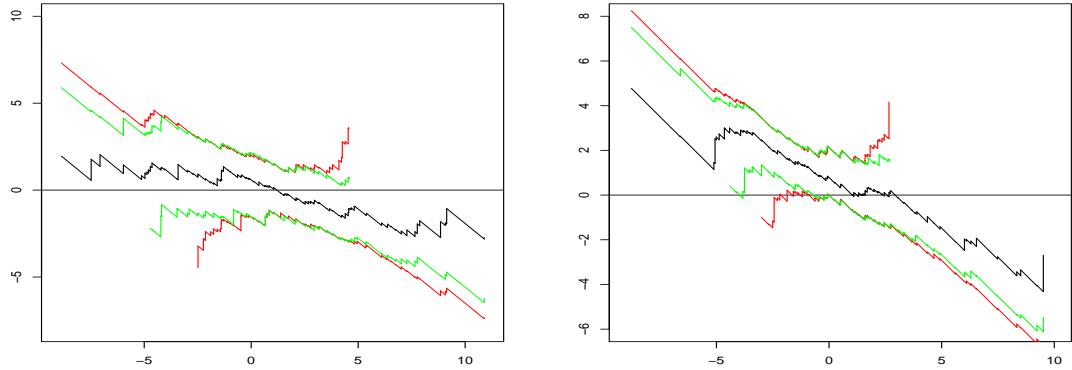
Turpmāk šos mēs sauksim par  $W$  vienlaicīgās ticamības inetrvaliem un rakstīsim  $[W_*(x), W^*(x))$ . Attēlā 2.1 attēlotas  $[S_*(x), S^*(x))$  un  $[W_*(x), W^*(x))$  vienlaicīgās ticamības joslas, kur  $X$



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

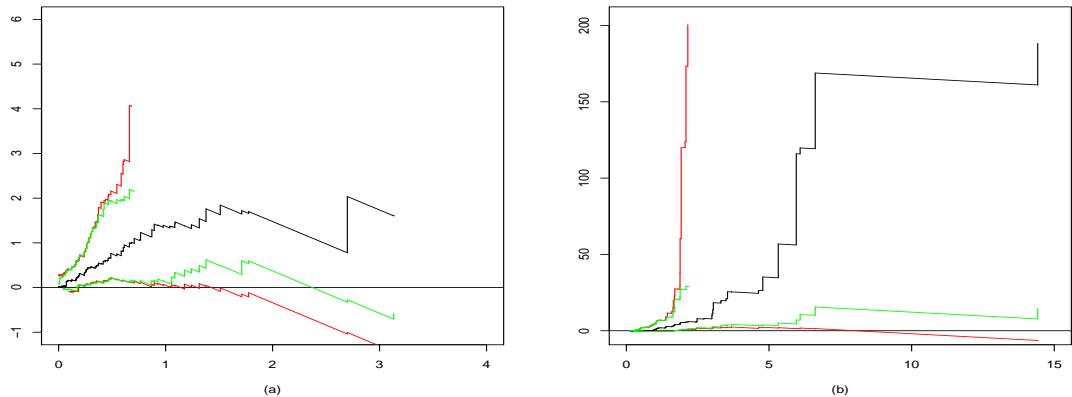
un  $Y$  sadalītas attiecīgi  $N(0, 1)$  un  $N(1, 4)$ , bet  $m = n = 100$  un  $\alpha = 0,05$ . Abas vienlaicīgās ticamības joslas ir diezgan vienādas, bet redzam, ka  $[W_*(x), W^*(x))$  josla ir šaurāka

galos. Šajā gadījumā skaidrs, ka taisne iekļaujas vienlaicīgajās ticamības joslās. Tas nozīmē, ka  $H_0$  tiek noraidīta, un mums nav lokācijas modelis. Bet varam sākt domāt, vai tur neveidojas lokācijas - skalēšanas modelis, kuram vienlaicīgās ticamības joslās jāiekļauj taisne.



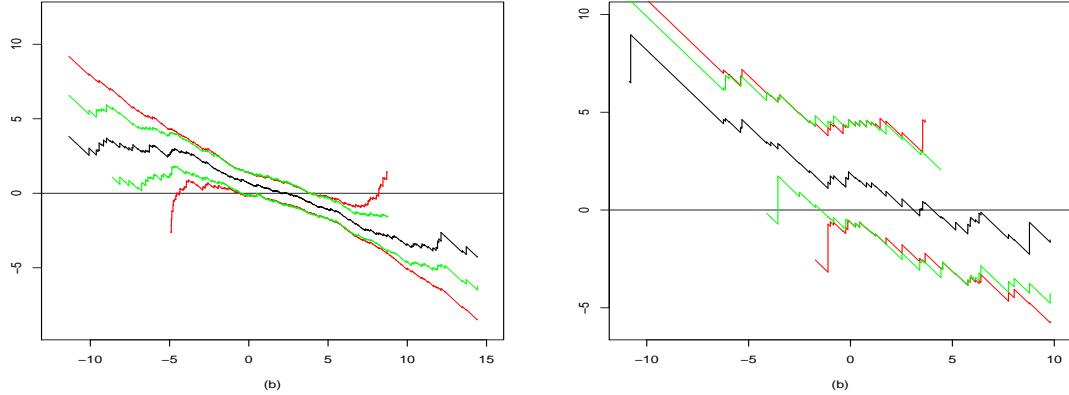
2.2. att.: (a) $N(1, 4)$  pret  $N(1, 3)$ ,  $n = m = 100$       (b) $N(0, 4)$  pret  $N(1, 2)$ ,  $n = m = 100$

Attēlā 2.2 attēlotas  $[S_*(x), S^*(x))$  un  $[W_*(x), W^*(x))$  vienlaicīgās ticamības joslas, kur  $X$  un  $Y$  sadalītas attiecīgi a) $N(1, 4)$  un  $N(1, 3)$ , bet  $m = n = 100$  un  $\alpha = 0, 05$ . b)  $N(0, 4)$  un  $N(1, 2)$ , bet  $m = n = 100$  un  $\alpha = 0, 05$ . Tas nozīmē, ka  $H_0$  netiek noraidīta, un mums ir lokācijas modelis.



2.3. att.: (a) $Exp(2)$  pret  $Exp(1)$ ,  $n = m = 100$    (b) $lnN(0, 1)$  pret  $lnN(1, 2)$ ,  $n = m = 100$

Attēlā 2.3 attēlotas  $[S_*(x), S^*(x)]$  un  $[W_*(x), W^*(x)]$  vienlaicīgās ticamības joslas, kur  $X$  un  $Y$  sadalītas attiecīgi a)  $Exp(2)$  un  $Exp(1)$ , bet  $m = n = 100$  un  $\alpha = 0,05$ . b)  $lnN(0, 1)$  un  $lnN(1, 2)$ , bet  $m = n = 100$  un  $\alpha = 0,05$ . Tas nozīmē, ka  $H_0$  noraidam.



2.4. att. (a)  $N(1, 5)$  pret  $N(2, 3)$ ,  $n = m = 500$       (b)  $N(1, 5)$  pret  $N(2, 3)$ ,  $n = m = 50$

Attēlā 2.4 attēlotas  $[S_*(x), S^*(x)]$  un  $[W_*(x), W^*(x)]$  vienlaicīgās ticamības joslas, kur  $X$  un  $Y$  sadalītas attiecīgi a)  $N(1, 5)$  un  $N(2, 3)$ , bet  $m = n = 50$  un  $\alpha = 0,05$ . b)  $N(1, 5)$  un  $N(2, 3)$ , bet  $m = n = 500$  un  $\alpha = 0,05$ . Tas nozīmē, ka a) gadījumā  $H_0$  noraidam, bet šeit ir lokācijas - skalēšanas modelis, kur jāiekļaujas taisnei. b) gadījumā  $H_0$  nenoraidam, un mums ir lokācijas modelis.

### 3. Vienlaicīgās ticamības joslas divu kvantiļu funkciju starpībai

Pārbaudīsim hipotēzi (0.1) vai ir lokācijas modelis, izmantojot vienlaicīgās ticamības joslas divu kvantiļu funkciju starpībai. Definēsim  $\Lambda(\bullet)$ :

$$\Lambda(t) = F^{-1}(t) - G^{-1}(t),$$

kur  $0 \leq t \leq 1$ ,  $F^{-1}$  un  $G^{-1}$  ir attiecīgi  $X$  un  $Y$  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) [5]. Doksums  $\Delta(x) = G^{-1}[F(x)] - x$  nosauc par lokācijas funkciju. Pieņemot, ka  $\Delta(x)$  ir horizontālais attālums starp  $F$  un  $G$  attiecīgajā punktā  $x$ , mēs redzam, ka

$$\Lambda[F(x)] = -\Delta(x), \quad (3.0.1)$$

kur  $\Delta(x)$  ir identiska ar lokācijas funkciju  $G^{-1}[F(x)] - x$ . Šī formula ir vispārīgāks gadījums, jo ir iespējama situācija, kad  $\Delta(x) + x$  ir dilstoši.

Novērtējums  $\Lambda(t)$  ir uzzīdzots, kā

$$\hat{\Lambda}(t) = F_m^{-1}(t) - G_n^{-1}(t), \quad (3.0.2)$$

kur  $0 \leq t \leq 1$ , kur ar  $F_m^{-1}$  un  $G_n^{-1}$  apzīmējam  $F_m$  un  $G_n$  kreisās inversās funkcijas.

Priekš  $t \in ((i-1)/n, i/n]$ ,  $G_n^{-1}(t) = Y_{(i)}$  un  $F_m^{-1}(t) = X_{(\langle mi/n \rangle)}$ , kur  $\langle v \rangle$  - vismazākais skaitlis, kas ir lielāks vai vienāds ar  $v$  un  $X_1 < \dots < X_m$  un  $Y_1 < \dots < Y_n$ . Tādēļ iegūstam

$$\hat{\Lambda}(t) = X_{(\langle mi/n \rangle)} - Y_{(i)}, \quad t \in ((i-1)/n, i/n]$$

Ja doto izlašu izmēri ir vienādi ( $m = n$ ), tad varam rakstīt

$$\hat{\Lambda}(t) = X_{(i)} - Y_{(i)}, t \in ((i-1)/n, i/n].$$

Tālāk izmantosim, ka  $m^{1/2}[F_m^{-1}(\bullet) - F^{-1}(\bullet)]$  un  $n^{1/2}[G_n^{-1}(\bullet) - G^{-1}(\bullet)]$  konverģē uz Gausa procesu.

Zināms, ka

$$m^{1/2}[F_m^{-1}(\bullet) - F^{-1}(\bullet)] \xrightarrow{D} \frac{W_1(\bullet)}{f[F^{-1}(\bullet)]}$$

un līdzīgi

$$n^{1/2}[G_n^{-1}(\bullet) - G^{-1}(\bullet)] \xrightarrow{D} \frac{W_2(\bullet)}{g[G^{-1}(\bullet)]}.$$

Pierādījuma uzmetumu apskatījis Serflings [6].

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

$$N^{1/2}[\hat{\Lambda}(t) - \Lambda(t)] = (N/m)^{1/2}m^{1/2}[F_m^{-1}(t) - F^{-1}(t)] - (N/n)^{1/2}n^{1/2}[G_m^{-1}(t) - G^{-1}(t)],$$

Kas dod

$$N^{1/2}[\hat{\Lambda}(\bullet) - \Lambda(\bullet)] \xrightarrow{D} \frac{1}{\theta^{1/2}} \frac{W_1(\bullet)}{f[F^{-1}(\bullet)]} - \frac{1}{(1-\theta)^{1/2}} \frac{W_2(\bullet)}{g[G^{-1}(\bullet)]}.$$

Izmantojot rezultātus, mēs iegūstam

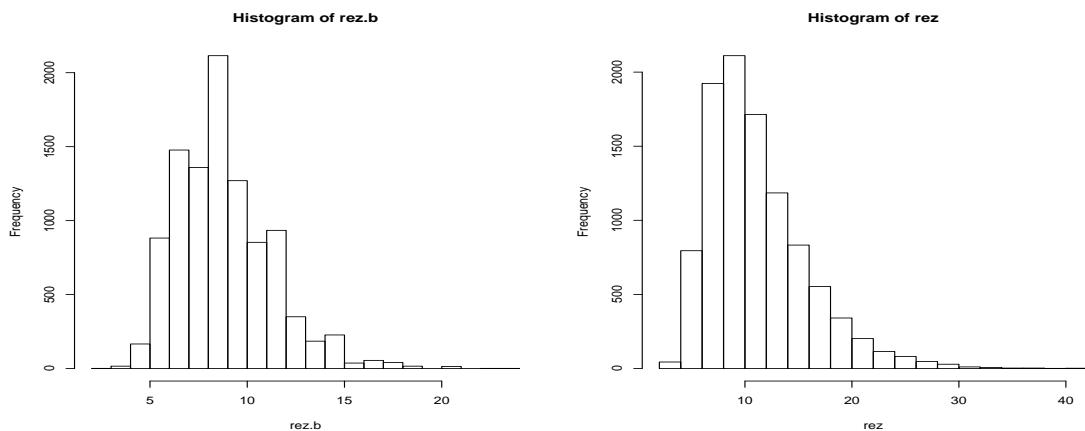
$$k^2(t) = \left\{ \frac{1}{\theta f^2[F^{-1}(t)]} + \frac{1}{(1-\theta)g^2[G^{-1}(t)]} \right\} t(1-t).$$

Vienlaicīgās ticamības joslas tiek konstruētas, izmantojot divas iespējas - tiek simulēta un Bootstrapa kritiskā vērtība. Tabulā ir apkopotas kritiskās vērtības, kas ir iegūtas simulējot kritisko vērtību (apzīmētas ar  $S$ ) un izmantojot Butstrapa metodi (apzīmētas ar  $B$ ).

3.1. tabula Dispersijas un lokācijas parametra izmaiņas

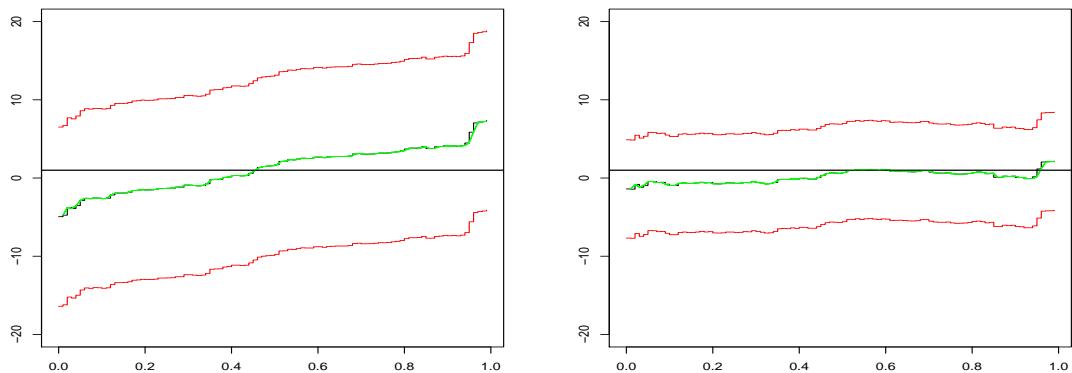
Dispersijas	1	2	4
Lokācijas parametrs 1	$B = 13.60681$	$B = 27.21361$	$B = 54.42722$
	$S = 20.04965$	$S = 40.09931$	$S = 80.19862$
Lokācijas parametrs 2	$B = 13.60681$	$B = 27.21361$	$B = 54.42722$
	$S = 20.04965$	$S = 40.09931$	$S = 80.19862$
Lokācijas parametrs 4	$B = 13.60681$	$B = 27.21361$	$B = 54.42722$
	$S = 20.04965$	$S = 40.09931$	$S = 80.19862$

Tabulā 3.1 redzams, ka lokācijas parametra izmaiņas neietekmē kritisko vērtību, bet to ietekmē dispersijas izmaiņas.



3.1. att. Histogrammas kritiskās vērtības butstrapošanai un simulešanai

Attēlā 3.1 attēlotas histogrammas, kas iegūtas Butstrapojot kritisko vērtību un simulejot kritisko vērtību. Histogrammas ir diezgan līdzīgas. Tikai Butstrapojot tās intervāls ir šaurāks, jo datus ņemam no datiem.



3.2. att.

Attēlā 2.3  $X$  un  $Y$  sadalītas attiecīgi a)  $N(0, 1)$  un  $N(1, 4)$ , bet  $m = n = 100$  un  $\alpha = 0.05$ . b) attēlā  $X$  un  $Y$  sadalītas attiecīgi  $N(0, 1)$  un  $N(1, 2)$ , bet  $m = n = 100$  un  $\alpha = 0.05$ . Tas nozīmē, ka  $H_0$  netiek noraidīta.

## 4. Pielietojam datus

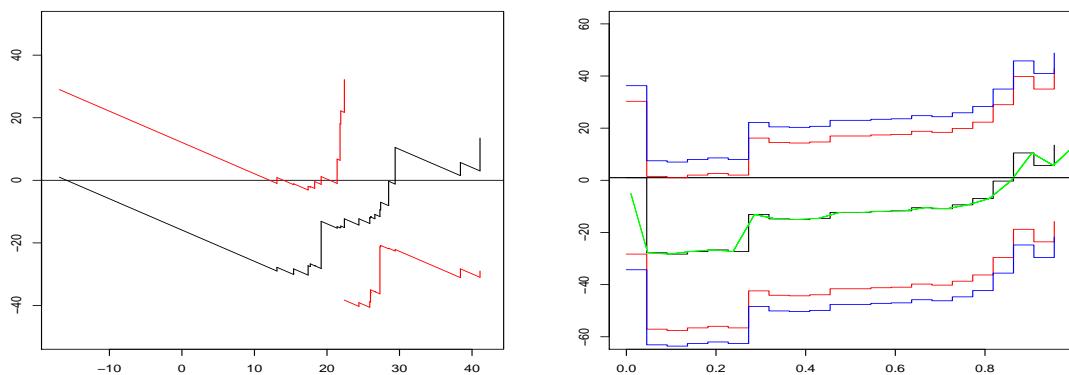
Pārbaudīsim hipotēzi (0.1) par lokācijas modeli datiem no publikācijas [3]. Pirmie dati satur informāciju par 22 žurkām, kas ir turētas 7 dienas ozona vidē un to svara pieaugums tiek apzīmēts ar  $y_i$ .

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 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

Grafiski zīmēsim vienlaicīgās ticamības joslas parametram  $\theta$  un kvantiļu funkcijas starpībai. Lai konstruētu vienlaicīgās ticamības joslas kvantiļu funkcijas starpībai, mums ir jāiegūst kritiskā vērtība. Kritiskā vērtība, kas ir iegūta simulējot, ir 194.3542, bet, izmantojot Butstrapa metodi, ir 233.4904.



4.1. att. (a)  $[S_*(x), S^*(x)]$       (b) kavtniļu starpības

Atēlā 4.1 tiek veikta grafiskā hipotēžu pārbaude vai starp datiem pastāv lokācijas modelis. Gadījumā a) tika apskatītas vienlaicīgās ticamības joslas lokācijas parametram  $\theta$ . Tika apskatītas  $[S_*(x), S^*(x))$  vienlaicīgās ticamības joslas, jo, lai rēķinātu  $[W_*(x), W^*(x))$  vienlaicīgās ticamības joslām, mēs nezinām kritisko vērtību šiem datiem.  $H_0$  netiek noraidīta. Gadījumā b) tika apskatītas vienlaicīgās ticamības joslas divu kvantiļu funkciju starpībai. Redzam, ka  $H_0$  netiek noraidīta.

## Nobeigums

Darbā tika apskatīta grafiskā hipotēžu pārbaude par lokācijas modeli (0.1), izmantojot divas pieejas. Tika apskatītas vienlaicīgās ticamības joslas lokācijas parametram  $\theta$  un vienlaicīgās ticamības joslas divu kvantiļu funkciju starpībai.  $H_0$  netiek noraidīta, ja ir horizontāla taisne, kas iekļaujas vienlaicīgās ticamībasjoslās, tas nozīmē, ka mums ir lokācijas modelis. Ja nav horizontāla taisne, kas iekļaujas vienlaicīgās ticamībasjoslās, tad mēs varam noraidīt  $H_0$ .

Problēmas rada kritiskās vērtības, kas ir nepieciešamas, lai konstruētu vienlaicīgās ticamības joslas. Darbā rēķinot  $W$  vienlaicīgās ticamības joslas parametram  $\theta$  tika ņemta konkrēta kritiskā vērtība, bet mums to vajadzetu pašiem simulēt. 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.

Darbu turpinot vajadzētu salīdzināt iegūtos rezultātus ar metodi, kad novērtē modeļa parametrus. Apskatīt grafisko hipotēžu pārbaudi lokācijas - skalēšanas modelim.

# Izmantotā literatūra un avoti

- [1] J. Cielēns. *Empīrisko procesu pielietojums strukturālo attiecību modeļos*. 2010.
- [2] G. Freitag and A. Munk. *Approximation Theorems of Mathematical Statictics*. 2005.
- [3] Kjell A. Doksum and Gerald L. Sievers. *Plotting with confidence: Graphical comparisons of two populations*, volume 63. 1976.
- [4] P. Switzer. *Confidence procedures for two sample problems*, volume 53. 1976.
- [5] K. A. Doksum. *Empirical probability plots and statistical inference for nonlinear models in the two sample case*, volume 2. 1974.
- [6] R.J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley, New York, 1980.
- [7] Rolf Aaberge Petter Laake, Knut Laake. *On the Problem of Measuring the Distance Between Distribution Functions: Analysis of hospitalization versus Mortality*, volume 41. 1985.

# Programmas kodi

```
\begin{footnotesize}
par(mfrow=c(1,2))
n<-100
m<-n
N<- n+m
alpha<- 0.05
izlase1<-sort(rlnorm(n,0,1))
izlase2<-sort(rlnorm(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))+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]

apaksha
#jāzīmē (apakša)
#(c(Y[20],apaksha[1]-Y[20])
#c(Y[21]),apaksha[1]-Y[21])
#c(Y[21],apaksha[2]-Y[21])
#c(Y[22],apaksha[2]-Y[22])

# apaksheja linija
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)]
augsha
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]
apaksha2
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)]
augsha2

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", )
}

#####
n<-100
set.seed(2)
dati1<-rnorm(n,0,1)
dati2<-rnorm(n,1,2)
par(mfrow=c(1,2))

# Teorētiskais grafiks
lambda<-function(x) {help
  qnorm(x,1,1)-qnorm(x,0,1)
}

# pēc definīcijas inversai funkcijai
plot((0:(n-1))/n,sort(dati2)-sort(dati1),
type="s", ylim=c(-20,20), ylab=" ", xlab=" ")
abline(1,0)

```

```

xx<-seq(0.01,0.99,by=0.0001)

# empiriskais grafiks
lines(xx,quantile(dati2,xx)-quantile(dati1,xx),
type="s", col="green")
abline(1,0)

# kritiskās vērtības simulācijas
stat.lambda<-function(n) {
  loc<-1
  dati1<-rnorm(n,0,1)
  dati2<-rnorm(n,1,2)
  sqrt(2*n)*(max(abs(sort(dati2)-sort(dati1)-1)))
}
## Simulētā kritiskā vērtība un tic.joslas
N<-10000
rez<-replicate(N,stat.lambda(n))
alpha<-0.05
sort(rez)[(1-alpha)*N]
#hist(rez)
krit<-sort(rez)[(1-alpha)*N]
lines((0:(n-1))/n,sort(dati2)-sort(dati1)-krit/sqrt(2*n),
type="s",col="red")
lines((0:(n-1))/n,sort(dati2)-sort(dati1)+krit/sqrt(2*n),
type="s",col="red")
## Butstraps
n<-100
set.seed(1)
dati1<-rnorm(n,0,1)
dati2<-rnorm(n,1,2)
stat.lambda.b<-function(n) {
  lok<-sort(dati2)-sort(dati1)
  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<-10000

rez.b<-replicate(N,stat.lambda.b(n))

alpha<-0.05

sort(rez.b)[(1-alpha)*N]

#hist(rez.b)

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")

#### Hip parbaude

n<-100

x1<-rnorm(n,1,2)

x2<-rnorm(n,1,2)

par(mfrow=c(1,2))

#pp grafiks

tt<-seq(0.01,0.99,by=0.01)

#empīriskais P-P grafiks

PP.emp<-ecdf(x1)(quantile(x2,tt))

plot(tt,PP.emp,type="s",

main="Vienlaicīgie ticamības intervāls P-P grafikam")

kr.v<-1.892441

x<-seq(0.01,0.99,by=0.01)

points(x,PP.emp-kr.v/sqrt(n),col="red",type="s")

points(x,PP.emp+kr.v/sqrt(n),col="red",type="s")

abline(0,1)

# Hipotēžuu pārbaude

x1<-rnorm(n,1,2)

x2<-rnorm(n,1,2)

```

```

# QQ grafiks(empīriskais)
xx<-seq(-3,3,by=0.01)
plot(xx,quantile(x1,(ecdf(x2)(xx))),type="s",ylab="QQ.emp",
main="Vienlaicīgie ticamības intervāls Q-Q grafikam")
abline(0,1)
kr.v<-1.89
library(sm)
# f(x) novērtējums ar kodolu metodi
x<-seq(-3,3,by=0.01)
f<-quantile(x1,(ecdf(x2)(x)))
h1<-hcv(x1)
K<-function(x)1/sqrt(2*pi)*exp(-x^2/2)
bliv<-function(x)
{
  1/n/h1*sum(K((x-x1)/h1))
}
bliv2<-Vectorize(bliv)
points(x,f-kr.v/(sqrt(n)*bliv2(x)),col="red",type="s")
points(x,f+kr.v/(sqrt(n)*bliv2(x)),col="red",type="s")
#####
dati.br<-scan(file="dati1.txt")
dati.oz<-scan(file="dati2.txt")
\end{footnotesize}

```

darbs “Grafiskā hipotēžu pārbaude lokācijas modelim” 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: Oleta Lorence

---

(paraksts)

---

(datums)

Rekomendēju darbu aizstāvēšanai.

Vadītājs: doc. Dr.math. Jānis Valeinis

---

(paraksts)

---

(datums)

Recenzents:

---

(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)