

LATVIJAS UNIVERSITĀTE
FIZIKAS UN MATEMĀTIKAS FAKULTĀTE
MATEMĀTIKAS NODAĻA

**NEPARAMETRISKĀS REGRESIJAS KĻŪDAS
DISPERSIJAS ROBUSTI NOVĒRTĒJUMI**

DIPLOMDABRS

Autors: **Maksims Korotejevs**

Stud. apl. mk08265

Darba vadītājs: doc. Dr.math. Jānis Valeinis

RĪGA 2013. g. 9. janvārī

Anotācija

Darbs ir veltīts neparametriskās regresijas kļūdu dispersijas novērtējumam. Darbā tiek salīdzinātas dažādas piejas kļūdu dispersijas novērtējumam. Tieki analizēti nesen ievesti lokālie $M-$ novērtējumi, kur kļūdu dispersija tiek aplūkota kā funkcija. Tieki pētīta $\widehat{\sigma}_{RICE,n}(x)$ novērtējuma uzvedība atkarībā no gludinoša parametra h . Darbā mērķi ir salīdzināt lokālos $M-$ novērtējumus savā starpā, kā arī pielietot tos reāliem datiem.

Atslēgas vārdi: neparametriska regresija, kļūdu dispersijas funkcija, lokālie $M-$ novērtējumi, gludinošs parametrs, ticamības intervāli

Abstract

This work is about error scale estimation in nonparametric regression models. Different approaches to estimating error scale are compared. Newly introduced local $M-$ estimators, where error scale is observed as a function, are analyzed. $\widehat{\sigma}_{RICE,n}(x)$ estimator behaviour depending on smoothing parameter h is examined. The goals of this work are to compare local $M-$ estimators and use them on real data.

Keywords: nonparametric regression, error scale function, local $M-$ estimators, bandwidth, confidence bands

Saturs

Apzīmejumi	2
Ievads	3
1. Parametriska un neparametriska regresija	5
1.1. Lineāra regresija	5
1.2. Neparametriska regresija	7
1.3. Neparametriskas regresijas dispersijas Yu un Jones novērtējums . .	10
1.4. M – novērtējumi	11
1.5. Dispersijas lokālie M – novērtējumi	13
1.6. Datu simulēšana	15
2. Gludinošā parametra h izvēle	16
3. Punktveida ticamības intervālu konstruēšana	18
3.1. Punktveida ticamības intervālu konstruēšana, kad $\chi(x) = x^2$. . .	18
3.2. $\int K^2(u)du$ aprēķini	20
3.3. Punktveida ticamības intervālu konstruēšana, kad $\chi(x) = x^2$ ar butstrapa metodi	21
4. Praktiskā daļa	22
4.1. Robežsadalījuma konverģence	22
4.2. Gludinošā parametra h izvēle	24
4.3. Kļūdu dispersijas novērtējumu pielietojums simulētiem datiem . .	31
4.4. Kļūdu dispersijas novērtējumu pielietojums reāliem datiem . . .	36
4.5. Novērtējuma $\widehat{\sigma}_{RICE,n}(x)$ punktveida ticamības intervālu konstruē- šana LIDAR datiem izmantojot butstrapa metodi	38
Secinājumi	39
Izmantotā literatūra un avoti	40

Apzīmejumi

$\widehat{\sigma}_{RICE,n}(x)$ lokālais klasiskais Rice novērtējums

$\widehat{\sigma}_{MSD,n}(x)$ lokālais $M -$ novērtējums ar MSD funkciju

$\widehat{\sigma}_{MBT,n}(x)$ lokālais $M -$ novērtējums ar BT funkciju

$\widehat{\sigma}_n(x)$ novērtējums ar Yu un Jones funkciju

$\widehat{\sigma}_{const,n}(x)$ novērtējums kā konstanti

\xrightarrow{D} konverģence pēc sadalījuma

\xrightarrow{P} konverģence pēc varbūtības

$\xrightarrow{g.d.}$ gandrīz droša konverģence

...

Ievads

Nodarbojoties ar neparametrisko regresiju svarīgi ir novērtēt kļūdu dispersiju. Dispersijas kļūdu var novērtēt kā konstanti vai kā funkciju. Vispārīgi regresijai ir svarīgs nosacījums par homoskedastitāti. Sarežģītākais gadījums ir heteroskedastiska modeļa gadījums, kad dispersija nav konstanta. Larry Wasserman savā grāmatā [1] aprakstīja dispersijas novērtējumu, kuru izgudroja Yu un Jones [2], dispersijas novērtējumam pielietojot divas reizes neparametrisko regresiju. Savukārt, Brown un Levin [3] vispārināja Rice [4] dispersijas novērtējumu no homoskedastiska modeļa līdz heteroskedastiskam modelim izmantojot kodola svarus. Samērā nesen tika piedāvāti lokālie M robusti, tā saucamie M – novērtējumi. Kad datiem ir izlecēji, nepieciešams izmantot robusta statistiku kļudas dispersijas funkcijas novērtējumam, lai varētu novērtēt regresijas funkciju šiem datiem. Darbā tika apskatīti lokālie M – novērtējumi dispersijai. Lai novērstu izlēcēju efektu uz neparametrisko modeli, izmantosim Rice [4] piedāvāto izteiksmi un pārveidoto izteiksmi, izmantojot robustu statistiku, kuru piedāvā Boente, Fraiman un Meloche [5].

Robusta statistikas pielietojumam ir liela nozīmē šajā darbā, tāpēc apskatīsim to derīgumu sīkāk. Dispersijas novērtēšanai lieto robusts statistikas, lai ievērotu izlēcējus (Hanning un Lee) [6], lai uzlabotu precizitāti izvēloties atbilstošu gludinošo parametru h , kad novērtējam regresijas funkciju r (Boente [5], Cantoni un Ronchetti [7], Leung [8]).

Robusta statistikas novērtējums, tiek plaši izmantots dispersijas funkcijas novērtēšanai neparametriskā regresijas modelī ar izlēcējiem. Šāds novērtējums ir ļoti noderīgs rēķinot robusta M – novērtējumus regresijas funkcijai (to apskatīja Hardle un Gasser [9], Hardle un Tsybakov [10], Boente un Fraiman [11]).

Šī darba mērķi ir sekojoši:

- 2010. gada publikācijas [12] detalizēta analīze, kur ieviesti kļudas dispersijas M – novērtējumi;
- Dispersijas lokālo M – novērtējumu konstruēšana;
- Gludinošā parametra h izvēle, analizējot publikācijas [12], [13];
- Dispersijas novērtējuma pielietojums simulētiem un reāliem datiem.
- Dažādu dispersijas novērtējumu salīdzināšana;
- Novērtējumam $\widehat{\sigma}_{RICE,n}(x)$ punktveida ticamības intervālu konstruēšana;

Darbs sastāv no ievada, četrām nodaļām, secinājumiem un literatūras avotu saraksta.

1. nodaļā apskatīta dispersijas lokālo M – novērtējumu teorija, 2. nodaļā aprakstīti gludi-

nošā parametra h izvēle, 3. nodaļā apskatīti punktveida ticamības intervāli novērtējumam $\widehat{\sigma}_{RICE,n}(x)$, 4. nodaļā atrodama simulāciju un datu analīze. Visi darbā veiktie aprēķini īstenoti programmā R.

1. Parametriska un neparametriska regresija

Pirms iepazīšanas ar jauniem neparametriskas regresijas kļūdu dispersijas novērtējumiem (lokāliem M – novērtējumiem) sāksim ar vienkāršam un visiem labi pazīstamam lietam. Aprakstīsim dispersijas novērtējumu parametriskai un neparametriskai regresijām. Tas dos priekšstatu par pētāmo objektu. Šīs nodaļas teorētiskais materiāls tiek ņemts no Larry Wasserman grāmatas [1].

1.1. Lineāra regresija

Apskatīsim parametisko lineāro regresiju. Pieņemsim, ka mums ir n dati $(x_1, Y_1), \dots, (x_n, Y_n)$, kur x_i nav gadījuma lielums, $x_i = (x_{i1}, \dots, x_{ip})^T \in \mathbb{R}^p$ ir fiksēti punkti un $Y_i \in \mathbb{R}$. Lineāras regresijas modelis ir:

$$Y_i = r(x_i) + \varepsilon_i \equiv \sum_{j=1}^p \beta_j x_{ij} + \varepsilon_i, i = 1, \dots, n; \beta_j \in \mathbb{R}; \varepsilon_i \in \mathbb{R},$$

kur $E(\varepsilon_i) = 0$ un $D(\varepsilon_i) = \sigma^2$. Būtiski ir novērtēt parametrus β_j . Lai to izdarītu ieviesīsim dizaina matricu X ar dimensiju $n \times p$:

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix},$$

kur $x_{i1} = 1$.

Pārrakstīsim regresijas modeli matricu formā. Pieņem, ka $Y = (Y_1, \dots, Y_n)^T$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ un $\beta = (\beta_1, \dots, \beta_p)^T$. Tad iegūst

$$Y = X\beta + \varepsilon.$$

Lai novērtētu β izmantosim mazāko kvadrātu metodi.

Definīcija 1. Par Mazāko kvadrātu novērtējumu $\hat{\beta}$ sauc vektoru $(\hat{\beta}_1, \dots, \hat{\beta}_p)^T$, kurš minimizē kvadrātu atlikumu summu (RSS):

$$RSS = (Y - X\beta)^T(Y - X\beta) = \sum_{i=1}^n \left(Y_i - \sum_{j=1}^p x_{ij} \beta_j \right)^2.$$

Apgalvojums 1. Ja $X^T X$ ir pilna ranga matrica, t.i., matrica $X^T X$ nav deģenerēta, tad $\exists (X^T X)^{-1}$ un

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

Apgalvojums 2. Regresijas $r(x)$ novērtējums $\hat{r}_n(x)$ pie fiksēta $x = (x_1, \dots, x_p)^T$ pierakstāma sekojošā formā

$$\hat{r}_n(x) = \sum_{j=1}^p \hat{\beta}_j x_j = x^T \hat{\beta}.$$

No tā seko, ka novērtētas vērtības $r = (\hat{r}_n(x_1), \dots, \hat{r}_n(x_n))^T$ varētu būt pierakstītas, šādi

$$r = X \hat{\beta} = LY,$$

kur

$$L = X (X^T X)^{-1} X^T.$$

Pie tam matrica L (*hat matrix*) ir simetriska $L = L^T$ un idempotenta $L^2 = L$. Parametru skaits p tiek noteikts no matricas L pēdas, kas ir diagonālelementu summa:

$$p = \text{tr}(L).$$

Apgalvojums 3. Jebkuram fiksētam x dispersijas σ^2 nenovirzīts novērtējums ir

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{r}_n(x_i))^2}{n - p},$$

kur

$$\hat{r}_n(x) = \ell(x)^T Y = \sum_{i=1}^n \ell_i(x) Y_i,$$

un

$$\ell_i(x)^T = x^T (X^T X)^{-1} X^T.$$

1.2. Neparametriska regresija

Atgādināsim neparametriskas regresijas jēdzienu. Doti n novērojumu pāri $(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n)$, kur x_i ir fiksēti punkti. Neparametriskas regresijas vienādojums ir sekojošs

$$Y_i = r(x_i) + \varepsilon_i, E\varepsilon_i = 0, i = 1, \dots, n, \quad (1.1)$$

kur $r(x)$ ir regresijas funkcija. Regresijas funkcijas $r(x)$ novērtējumu apzīmēsim ar $\hat{r}_n(x)$. Pieņemsim, ka dispersija $D(\varepsilon_i) = \sigma^2$ ir atkarīga no x . Pēc modeļa (1.1) uzskatam, ka x_i ir fiksēti un $(x_{i1}, \dots, x_{ip})^T \in \mathbb{R}^p$. Ja skaidrojošais mainīgais x ir gadījuma lielums, tad parakstīsim datus, šādi $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ un $r(x)$ tiek interpretēts, kā nosacīta matemātiska cerība:

$$r(x) = E(Y|X = x).$$

Definīcija 2. Funkcijas r novērtējums \hat{r}_n ir lineārais gludinātājs, ja $\forall x \exists l(x) : l(x) = (l_1(x), \dots, l_n(x))^T$ tāds, ka

$$\hat{r}_n(x) = \sum_{i=1}^n l_i(x)Y_i.$$

Piezīme: $\forall x : \sum_{i=1}^n l_i(x) = 1$.

Definīcija 3. Par kodolu sauksim jebkuru gludu funkciju K , tādu kurai izpildās sekojoši nosacījumi:

- 1) $K(x) \geq 0$,
- 2) $\int K(x)dx = 1$,
- 3) $\int xK(x)dx = 0$,
- 4) $\int x^2K(x)dx > 0$.

Piemēri:

- 1) Boxcar kodols

$$K(x) = \frac{1}{2}I(x).$$

- 2) Gausa kodols

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

3) Epanečnikova kodols

$$K(x) = \frac{3}{4}(1 - x^2)I(x).$$

4) Tricube kodols

$$K(x) = \frac{70}{81}(1 - |x|^3)^3I(x).$$

kur

$$I(x) = \begin{cases} 1, & \text{ja } |x| \leq 1 \\ 0, & \text{ja } |x| > 1. \end{cases}$$

Definīcija 4. Pieņemsim, ka $h > 0$ ir pozitīvs skaitlis, kuru sauc par gludinošo parametru vai joslas platumu. Nadaraya – Watsona kodola novērtējums tiek definēts sekojoši

$$\hat{r}_n(x) = \sum_{i=1}^n w_i(x)Y_i,$$

kur K ir kodols, $w_i(x)$ ir svari, kas tiek definēti kā

$$w_i(x) = \frac{K\left(\frac{x - x_i}{h}\right)}{\sum_{j=1}^n K\left(\frac{x - x_j}{h}\right)}.$$

Nadaraya – Watsona svaru funkcija w_i nav vienīgais veids, kurā var izvēlēties svarus.

Tā piemēram par w_i varam ņemt Rosenblatt vai Pristley – Chao svarus

$$w_i(x) = \frac{1}{nh}K\left(\frac{x - x_i}{h}\right),$$

$$w_i(x) = \frac{1}{h}(x_{i+1} - x_i)K\left(\frac{x - x_j}{h}\right).$$

Dispersijas novērtējumam mēs izmantisim Gausa kodolu $K(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, tā kā tas bija ņemts pētāmā publikācijā [12]. Gludinošā parametra h izvēle ir svarīga problēma. Eksistē daudz metodes kā to var izdarīt, piemēram, krosvalidācijas metodes, īkšķa likums, ievietošanas metodes.

Parametru h jāizvēlas tāda veidā, lai minimizētu vidējo kvadrātisko funkciju (MSE) jeb riska funkciju

$$R(h) = E\left(\frac{1}{n} \sum_{i=1}^n \hat{r}_n(x_i) - r(x_i)\right)^2.$$

Bet $R(h)$ ir atkarīgs no nezināmas funkcijas $r(x)$. Tāpēc minimizēsim riska funkcijas $R(h)$ novērtējumu $\widehat{R}(h)$. Vidējo kvadrātu atlikumu summu minimizē sekojoši

$$\frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{r}_n(x_i))^2.$$

Bet šīs novērtējums ir novirzīts un nepietiekoši nogludināts, jo datus izmantojam divas reizes, lai novērtētu regresijas funkciju un risku. To iemeslu dēļ novērtēsim risku funkciju, pielietojot vienu — atstāt — ārā krosvalidācijas funkciju.

Definīcija 5. Par vienu — atstāt — ārā krosvalidācijas funkciju sauc

$$CV = \widehat{R}(h) = \frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{r}_{(-i)}(x_i))^2,$$

kur $\widehat{r}_{(-i)}$ ir novērtējums, kuru iegūst izlaižot novērojumu (X_i, Y_i) .

Definīcija 6. Par $\widehat{r}_{(-i)}(x_i)$ novērtējumu ņem

$$\widehat{r}_{(-i)}(x_i) = \sum_{j=1}^n Y_j l_{j,(-i)(x)},$$

kur

$$l_{j,(-i)(x)} = \begin{cases} 0, & \text{ja } j = i \\ \frac{l_j(x)}{\sum_{k \neq i} l_k(x)}, & \text{ja } j \neq i. \end{cases}$$

Ievērosim, ka

$$E(Y_i - \widehat{r}_{(-i)}(x_i))^2 = E(Y_i - r(x_i) + r(x_i) - \widehat{r}_{(-i)}(x_i))^2,$$

$$E(Y_i - r(x_i) + r(x_i) - \widehat{r}_{(-i)}(x_i))^2 = \sigma^2 + E(r(x_i) - \widehat{r}_{(-i)}(x_i))^2,$$

$$\sigma^2 + E(r(x_i) - \widehat{r}_{(-i)}(x_i))^2 \approx \sigma^2 + E(r(x_i) - \widehat{r}_n(x_i))^2,$$

Tādējādi prognozes risks izskatās šādi

$$\Rightarrow E(\widehat{R}) \approx R + \sigma^2.$$

Šī krosvalidācijas funkcija ir gandrīz nenovirzīts riska novērtējums. Sīkāk gludinoša parametra h izvēles problemātiku konkrētai klūdu dispersijas funkcijai apskatīsim citā nodalā.

1.3. Neparametriskas regresijas dispersijas Yu un Jones novērtējums

Dispersijas funkcijas novērtējumam, Yu un Jones [2] piedāvāja divas reizes pielietot neparametrisko regresiju, šī metode ir aprakstīta Larry Wasserman grāmatā [1] Pieņem, ka

$$Y_i = r(x_i) + \sigma(x_i)\varepsilon_i.$$

Apzīmēsim, ka $Z_i = \log(Y_i - r(x_i))^2$ un $\delta_i = \log(\varepsilon_i^2)$.

Tādējādi dispersijas novērtēšanas procedūra ir sekojoša:

- a) Novērtē regresijas funkciju $r(x)$ ar Nadaraya – Watsona neparametrisko metodi, lai iegūtu novērtējumu $\hat{r}_n(x)$.
- b) Aprēķinam $Z_i = \log(Y_i - \hat{r}_n(x_i))^2$.
- c) Veic regresiju Z_i pret x_i , iegūst $\log(\sigma^2(x))$ novērtējumu $\hat{\sigma}^2$

$$Y_i - r(x_i) = \sigma(x_i)\varepsilon_i,$$

$$(Y_i - r(x_i))^2 = \sigma^2(x_i)\varepsilon_i^2,$$

$$\log(Y_i - r(x_i))^2 = \log(\sigma^2(x_i)\varepsilon_i^2),$$

$$\log(Y_i - r(x_i))^2 = \log(\sigma^2(x_i)) + \log(\varepsilon_i^2),$$

tā, kā $Z_i = \log(Y_i - r(x_i))^2$ un $\delta_i = \log(\varepsilon_i^2)$

$$\Rightarrow Z_i - \delta_i = \log(\sigma^2(x_i)),$$

$$\hat{\sigma}^2(x_i) = e^{Z_i - \delta_i}.$$

1.4. M – novērtējumi

Lai labāk saprastu, kas ir lokālo M – novērtējumu funkciju klase, no kurienes viņi rodas, uz ko viņi balstās un to būtību aplūkosim populārus Hūbera novērtējumus jeb M – novērtējumus. M – novērtējumi ir plāša statistisko novērtējumu klase. Piemēram, mazāko kvadrātu metodes novērtējumi ir M – novērtējumi.

Visa zemāk izklāsta informācija paņemta no Victor, Ricardo, Douglas grāmatas [14].

Pieņemsim, ka mums ir x_i novērojumi, kuri ir atkarīgi no nezināma parametra μ (“īsta vērtība”) un gadījuma kļūdas u_i .

Definīcija 7. Par lokācijas modeli sauc izteiksmi

$$x_i = \mu + u_i, i = 1, \dots, n,$$

kur u_1, \dots, u_n ir neatkarīgi gadījuma lielumi ar sadalījumu funkciju F_0 .

Līdz ar to x_1, \dots, x_n ir neatkarīgi identiski sadalīti gadījuma lielumi ar sadalījuma funkciju

$$F(x) = F_0(x - \mu).$$

Pirmais zinātnieks, kurš ieviesa M – novērtējumus bija Peter Huber [15], kurš 1964. gadā piedāvāja minimizēt izteiksmi

$$\sum_{i=1}^n \rho(x_i, \mu).$$

Funkciju ρ izvēlas tāda veidā, lai dati būtu izturīgi pret izlēcējiem.

Parametra μ novērtējums $\hat{\mu}$ ir funkcija no novērojumiem $\hat{\mu} = \hat{\mu}(x_1, x_2, \dots, x_n) = \hat{\mu}(x)$.

Meklēsim tādu novērtējumu $\hat{\mu}$, kurām ar lielu varbūtību izpildās $\hat{\mu} \approx \mu$.

Lai iegūtu novērtējumu $\hat{\mu}$ izmantosim maksimālas ticamības metodi. Pieņemsim, ka F_0 ir sadalījuma funkcija atlikumiem u_i ar blīvuma funkciju $f_0 = F_0'$. Līdz ar to ticamības funkcija ir

$$L(x_1, \dots, x_n; \mu) = \prod_{i=1}^n f_0(x_i - \mu).$$

Parametra μ maksimālas ticamības novērtējums ir

$$\hat{\mu} = \hat{\mu}(x_1, \dots, x_n) = \arg \max_{\mu} L(x_1, \dots, x_n; \mu). \quad (1.2)$$

Tādējādi varam definēt M novērtējumu pozitīvam f_0 .

Definīcija 8. Par M – novērtējumu sauksim

$$\hat{\mu} = \arg \min_{\mu} \left(\sum_{i=1}^n \rho(x_i - \mu) \right),$$

kur $\rho = -\log f_0$.

Apskatīsim piemērus, kuri ilustrēs M – novērtējumus.

Piemērs 1. Ja $F_0 = N(0, 1)$, tad blīvuma funkcija f_0 ir

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Līdz ar to M – novērtējums ir vienāds ar

$$\hat{\mu} = \arg \min_{\mu} \sum_{i=1}^n (x_i - \mu)^2.$$

Piemērs 2. Ja F_0 ir dubultais eksponenciālais sadalījums, tad

$$f_0(x) = \frac{1}{2} e^{-|x|}.$$

Līdz ar to M – novērtējums ir ekvivalenti ar

$$\hat{\mu} = \arg \min_{\mu} \sum_{i=1}^n |x_i - \mu|.$$

Pamatotim, to ka 1. piemēra atrisinājums μ ir vidēja vērtība.

Pienem, ka ρ ir diferencējama, tad diferencējot izteiksmi $\hat{\mu} = \arg \min_{\mu} (\sum_{i=1}^n \rho(x_i - \mu))$ pēc μ iegūstam

$$\sum_{i=1}^n \psi(x_i - \hat{\mu}) = 0,$$

kur $\psi = \rho'$.

Pienemot, ka $\rho(x) = \frac{x^2}{2}$, tad $\psi(x) = x$ un izteiksme $\sum_{i=1}^n \psi(x_i - \hat{\mu}) = 0$ klūst par

$$\sum_{i=1}^n (x_i - \hat{\mu}) = 0,$$

kur $\hat{\mu} = \bar{x}$ ir atrisinājums.

1.5. Dispersijas lokālie M – novērtējumi

Apskatīsim sekojošo neparametriskas regresijas modeli:

$$Y_i = g(x_i) + U_i \sigma(x_i), 1 \leq i \leq n, \quad (1.3)$$

kur $x_0 \leq x_1 \leq \dots \leq x_n \leq 1$, σ nezināma dispersijas funkcija, g nezināma regresijas funkcija, kļūdas ir $U_i \sim F_0$ un U_1, U_2, \dots, U_n ir i.i.d.

Lokālo M – novērtējumu metodes būtība balstās uz rezultātu no homoskedastiskās neparametriskās regresijas modeļa, kuru ieguva Hall [16]. Viņš dispersiju novērtēja šādā veidā:

$$\widehat{\sigma}_{r,n}^2 = \frac{1}{n-r} \sum_{i=m_1+1}^{n-m_2} \left(\sum_{k=-m_1}^{m_2} d_k Y_{i+k} \right)^2,$$

kur Y_{i+k} jābūt sakārtotam. $\{d_i\}_{i=-m_1}^{m_2}$ ir virkne veidota no starpībām uz reāliem skaitļiem, kura apmierina $\sum_{i=-m_1}^{m_2} d_i = 0$ un $\sum_{i=-m_1}^{m_2} d_i^2 = 1$ ar $d_{-m_1} \neq 0$, $d_{-m_2} \neq 0$ priekš $m_1, m_2 \in Z^+$. Pie tam $r = m_1 + m_2$. Kad $r = 1$, $\widehat{\sigma}_{r,n}^2 = \widehat{\sigma}_{Rice,n}^2$, kas ir labi pazīstams, kā novērtējums pēc Rice [4]:

$$\widehat{\sigma}_{Rice,n}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{i+1} - Y_i)^2.$$

Šīs dispersijas funkciju klases paplašināja līdz heteroskedastiskas neparametriskas regresijas modeļiem 2007. gadā Brown un Levin [3], kuri apskatīja lokālo novērtējumu bāzētu uz kodola svariem.

Vispār, homoskedastiskiem neparametriskiem modeļiem, Ghement [17] novērtēja $\widehat{\sigma}_{Rice,n}^2$ izmantojot M – novērtējumu, definētu, kā atrisinājumu $\widehat{\sigma}_0$ izteiksmei

$$\frac{1}{n-1} \sum_{i=1}^{n-1} \chi \left(\frac{Y_{i+1} - Y_i}{a \widehat{\sigma}_0} \right) = b,$$

kur χ ir score funkcija, a pozitīva konstante, b pozitīva konstante, kura dod robust līmeni novērtējumam.

Pamatojoties uz Boente, Ruiz, Zamar [12] publikācijas rezultātiem reprezentēsim funkciju no dispersijas $\sigma(x)$ caur robustu statistiku. Šo $\sigma(x)$ novērtējumu sauksim par *lokālos M – novērtējumus no starpībām*. Aplūkosim datus, kuri apmierina (1.3) modeļa nosacījumus. Sākumā ģenerēsim datus ar daudziem izlecējpunktiem. Lai to izdarītu apskatīsim

sekojošu apakškopu:

$$\mathcal{P}_\varepsilon(F_0) = \{G|G(y) = (1 - \varepsilon)F_0(y) + \varepsilon H(y); H \in D, y \in \mathbb{R}\},$$

kur D ir visu sadalījuma funkcijas kopa, F_0 ir normālais sadalījums, H ir jebkura patvaļīga sadalījuma funkcija, kurā modelē piesārņojumus un $\varepsilon \in [0, 1/2]$.

Vispārīgā veidā definēsim $x \in (0, 1)$ dispersijas funkcijai $\sigma(x)$ lokālos M – novērtējumus, no starpībām:

$$\widehat{\sigma}_{M,n}(x) = \inf\{s > 0 | \sum_{i=1}^{n-1} w_{n,i}(x) \chi\left(\frac{Y_{i+1} - Y_i}{as}\right) \leq b\}, \quad (1.4)$$

kur $w_{n,i}(x)_{i=1}^{n-1}$ ir svaru funkcijas virkne (piemēram, kodols), χ ir skores (*score*) funkcija, $a \in (0, \infty)$ un $b \in (0, 1)$ pie tam

$$E(\chi(Z_1)) = b \text{ un } E\left(\chi\frac{Z_2 - Z_1}{a}\right) = b,$$

kur $\{Z_i\}_{i=1,2}$ ir i.i.d. ar sadalījumu $Z_1 \sim F_0$. Parasti $\chi : \mathbb{R} \rightarrow \mathbb{R}$.

Izteiksmē (1.4) nepieciešams aprēķināt infinimumu tikai, ja funkcija χ ir pārtraukta, bet ja χ ir nepārtraukta, tad viegli redzēt, ka $\widehat{\sigma}_{M,n}(x)$ apmierina izteiksmi

$$\sum_{i=1}^{n-1} w_{n,i}(x) \chi\left(\frac{Y_{i+1} - Y_i}{a\widehat{\sigma}_{M,n}(x)}\right) = b.$$

Aplūkosim trīs lokālos M – novērtējumu piemērus, pēc kuriem tālāk konstruēsim $\sigma(x)$ novērtējumu:

1) Pieņemsim, ka $\chi(x) = x^2$, $a = \sqrt{2}$, $b = 1$, tad mēs iegūsim klasisko Rice novērtējumu

$$\widehat{\sigma}_{RICE,n}(x) = \sqrt{\sum_{i=1}^{n-1} w_{n,i}(x) \left(\frac{Y_{i+1} - Y_i}{\sqrt{2}}\right)^2}.$$

2) Pieņemsim, ka $\chi(y) = I_{y:|y|>\Phi^{-1}(3/4)}(y)$, $a = \sqrt{2}$, $b = 1/2$, tad $\widehat{\sigma}_{MSD,n}(x)$ apmierina vienādojumu

$$\sum_{i=1}^{n-1} w_{n,i}(x) \chi\left(\frac{Y_{i+1} - Y_i}{a\widehat{\sigma}_{MSD,n}(x)}\right) - b = 0,$$

uras saknes, būs $\widehat{\sigma}_{MSD,n}(x)$ vērtības.

3) Pieņemsim, ka $\forall c > 0$ izpildās

$$\chi_c(y) = \begin{cases} 3\left(\frac{y}{c}\right)^2 - 3\left(\frac{y}{c}\right)^4 + \left(\frac{y}{c}\right)^6, & \text{ja } |y| \leq c \\ 1, & \text{ja } |y| > c, \end{cases}$$

tad paņemot $c = 0.70417$, $a = \sqrt{2}$, $b = 3/4$ iegūsim vienādojumu pret $\widehat{\sigma}_{MBT,n}(x)$ vērtības

$$\sum_{i=1}^{n-1} w_{n,i}(x) \chi_c \left(\frac{Y_{i+1} - Y_i}{a \widehat{\sigma}_{MBT,n}(x)} \right) - b = 0.$$

1.6. Datu simulēšana

Atgādināsim, neparametriskās regresijas formu

$$Y_i = g(x_i) + U_i \sigma(x_i), E(U_i) = 0, i = 1, \dots, n,$$

kur $g(x)$ ir regresijas funkcija. x_i dati tiek ģenerēti sekojoši $x_i = i/(n+1)$ un $1 \leq i \leq n$.

Savukārt kļūdas U_i iegūsti simulējot datus $G(y) = (1-\varepsilon)\Phi(y) + \varepsilon H(y)$, $\Phi(y) \sim N(0, 1)$, $\varepsilon = 0, 0.1, 0.2, 0.3, 0.35, 0.4$. Aplūkosim dažādas kļūdas simulēšanas modeļus:

- 1) $H(y) = N(\mu, \tau^2)$, kur $\mu = 10, 100, 1000$ un $\tau^2 = 0.1$
 - 2) $H(y) = C(0, \tau^2)$, kur $\tau = 4$. Paliek noskaidrot kādu regresijas funkciju $g(x)$ un kļūdu dispersijas funkciju $\sigma(x)$ ņemt. Dette un Hetzler [4] piedeva divus modeļus:
- M1) regresijas funkcija ir $g(x) = 2\sin(4\pi x)$ un kļūdu dispersijas funkcija ir $\sigma(x) = e^x$.
- M2) regresijas funkcija ir $g(x) = 1 + x$ un kļūdu dispersijas funkcija ir $\sigma(x) = 1 + (1 + \sin(10x))^2$.

Lai novērtētu cik ir labs izvēlētais gludinošs parametrs h dispersijas novērtējumam $\widehat{\sigma}_{RICE,n}(x)$, apskatīsim *ISEL* kļūdu.

Definīcija 9. Par \widehat{ISEL} kļūdu sauc

$$\widehat{ISEL}_j(\widehat{\sigma}_n^{(j)}(\cdot, \widehat{h}_n)) = \frac{1}{n} \sum_{i=1}^{n-1} \left[\log \left(\frac{\widehat{\sigma}_n^{(j)}(x_i, \widehat{h}_n)}{\sigma(x_i)} \right) \right]^2,$$

kur $\widehat{\sigma}_n^{(j)}(x_i, \widehat{h}_n)$ norāda dispersijas novērtējumu, kurā iegūta j – tā simulācija ar gludinošo parametru h .

Lai ģenerētu $G(y) = (1 - \varepsilon)\Phi(y) + \varepsilon H(y)$, $\Phi(y) \sim N(0, 1)$, $H(y) \sim N(10, 0.1)$ vai $H(y) \sim C(0, 4^2)$ datus izmantosim sekojošus apgalvojumus, alternatīvu gadījuma lielumu ģenerēšanai:

Apgalvojums 4. Ja X_1, X_2, \dots, X_n ir vienādi neatkarīgi sadalīti (i.i.d.) gadījuma lielumi un $X_i \sim F$, tad $Y_1 = F(X_1), Y_2 = F(X_2), \dots, Y_n = F(X_n)$ ir i.i.d. un pie tam Y_i ir vienmērīgi sadalīts gadījuma lielums ar vidējo vērtību 0 un dispersiju 1.

Apgalvojums 5. Ja doti Y_1, Y_2, \dots, Y_n ; Y_i ir vienmērīgi sadalīts gadījuma lielums ar vidējo vērtību 0 un dispersiju 1, tad $X_1 = F^{-1}(Y_1), \dots, X_n = F^{-1}(Y_n)$ ir i.i.d. un X_i ir sadalīts pēc F .

2. Gludinošā parametra h izvēle

Pievērsīsim īpašu uzmanību parametra h izvēlei, jo tas būtiski var uzlabot metodes precizitāti. Papildus informāciju par gludinoša parametra izvēli varētu dabūt izmantojot publikācijas [13], [5] un [4].

Parametru h_n var izvēlēties izmantojot krossvalidācijas un ievietošanas metodes. Detalizēti to apskata Hardle [18], [19]. Šajā darbā tiek apskatīta tikai krossvalidācijas metode. Tā kā sākuma dati speciāli tiek generēti ar daudziem izlēcējiem, gludinoša parametra izvēles parastas procedūras nav pielietojamas.

Levine [20] rekomendēja tā saucamo K – reizes krosvalidācijas metodi, gludinošā parametra h izvēlei. Šī metode izveido dispersijas novērtējumu, kura nav pārāk jūtama pret izlēcējiem.

Metodes būtība ir sekojoša: sadalām datu kopu $\{(x_i, Y_i)\}$ uz K aptuveni vienādām, nešķērsojošām apakškopām. j – tā apakškopa ir izmērā $n_j \geq 2$, $\sum_{j=1}^K n_j = n$. Pieņemsim, ka datu pāris $\{(\tilde{x}_i^{(j)}, \tilde{y}_i^{(j)})\}_{1 \leq i \leq n_j}$ pieder j – tai apakškopai, kur vērtības $\tilde{x}_i^{(j)}$ sakārtoti augošā secībā. Analogiski, $\{(x_i^{(j)}, y_i^{(j)})\}_{1 \leq i \leq n-n_j}$ ir papildinājums j – tai apakškopai, kur $x_i^{(j)}$ sakārtoti augošā secībā. Kopa $\{(x_i^{(j)}, y_i^{(j)})\}_{1 \leq i \leq n-n_j}$ būs training un $\{(\tilde{x}_i^{(j)}, \tilde{y}_i^{(j)})\}_{1 \leq i \leq n_j}$ būs validation kopa. Apzīmēsim $\Delta_i^{(j)} = (y_{i+1}^{(j)} - y_i^{(j)})/\sqrt{2}$ un $\tilde{\Delta}_i^{(j)} = (\tilde{y}_{i+1}^{(j)} - \tilde{y}_i^{(j)})/\sqrt{2}$.

Pieņemsim, ka $\hat{\sigma}_{RICE,n}^{(j)}(x, h)$ ir lokālais klasiskais Rice novērtējums, ar gludinošo parametru h uz j – tās apakškopas $\{(x_i^{(j)}, y_i^{(j)})\}_{1 \leq i \leq n-n_j}$.

Definīcija 10. Par klasisko K reizes krosvalidācijas funkciju $CV_{LS,KCV}(h)$ sauc

$$CV_{LS,KCV}(h) = \frac{1}{n} \sum_{j=1}^K \sum_{i=1}^{n_j-1} \left[\left(\tilde{\Delta}_i^{(j)} \right)^2 - \left(\hat{\sigma}_{RICE,n}^{(j)}(\tilde{x}_i, h) \right)^2 \right]^2.$$

Definīcija 11. Par K reizes krosvalidācijas gludinošo parametru $\hat{h}_{LS,KCV}$ sauc

$$\hat{h}_{LS,KCV} = \operatorname{argmin}_{h \in H} CV_{LS,KCV}(h),$$

kur kopa H satur parametra h visas iespējamas vērtības robežās $[0, 1]$.

Piezīme 6. Varam apskatīt K reizes krosvalidācijas procedūru izmantojot naturālo logaritmu \ln :

$$CV_{LS,KCV}^{\ln}(h) = \frac{1}{n} \sum_{j=1}^K \sum_{i=1}^{n_j-1} \left[\ln \left(\tilde{\Delta}_i^{(j)} \right) - \ln \left(\hat{\sigma}_{RICE,n}^{(j)}(\tilde{x}_i, h) \right) \right]^2,$$

savukārt $\hat{h}_{LS,KCV}^{ln} = \operatorname{argmin}_{h \in H} CV_{LS,KCV}^{ln}(h)$.

Pienemot, ka $K = 1$ iegūstam K reizes krosvalidācijas metodes speciālo gadījumu: klasisko vienu – atstāt – ārā krosvalidācijas procedūru.

Definīcija 12. Par klasisko vienu – atstāt – ārā krosvalidācijas funkciju $CV_{LS,CV}(h)$

sauc

$$CV_{LS,CV}(h) = \frac{1}{n-1} \sum_{i=1}^{n-1} \left[D_i^2 - \left(\widehat{\sigma}_{RICE,n}^{(-i)}(\tilde{x}_i, h) \right)^2 \right]^2,$$

kur $D_i^2 = |Y_{i+1} - Y_i|/\sqrt{2}$ un $\widehat{\sigma}_{RICE,n}^{(-i)}(\tilde{x}_i, h)$ ir klasiskais dispersijas $\sigma(x_i)$ novērtējums, kurš tiek izrēķināts ar parametru h , izmantojot novērojumus

$(x_1, Y_1), \dots, (x_{i-1}, Y_{i-1}), (x_{i+2}, Y_{i+2}), \dots, (x_n, Y_n)$. Tātad mēs izlaižam $(x_i, Y_i), (x_{i+1}, Y_{i+1})$.

Definīcija 13. Par klasisko vienu – atstāt – ārā krosvalidācijas gludinošo parametru $\hat{h}_{LS,CV}$ sauc

$$\hat{h}_{LS,CV} = \operatorname{argmin}_{h \in H} CV_{LS,CV}(h),$$

kur kopa H satur parametra h visas iespējamas vērtības, robežās $[0, 1]$.

Piezīme 7. Krosvalidācijas funkcija izmantojot naturālo logaritmu ir sekojoša

$$CV_{LS,CV}^{ln}(h) = \frac{1}{n-1} \sum_{i=1}^{n-1} \left[\ln(D_i) - \ln\left(\widehat{\sigma}_{RICE,n}^{(-i)}(\tilde{x}_i, h)\right) \right]^2,$$

Savukārt $\hat{h}_{LS,CV}^{ln} = \operatorname{argmin}_{h \in H} CV_{LS,CV}^{ln}(h)$.

3. Punktveida ticamības intervālu konstruēšana

3.1. Punktveida ticamības intervālu konstruēšana, kad $\chi(x) = x^2$

Apskatīsim sekojošas divas teorēmas no [13] publikācijas:

Teorēma 8. $\forall x \in (0, 1)$:

$$\widehat{\sigma}_{M,n}(x, \widehat{h}_n) \xrightarrow{g.d.} S(G_x),$$

kur $S(G_x)$ apmierina izteiksmi $E\left[\chi\left(\frac{\sigma(x)(U_2 - U_1)}{aS(G_x)}\right)\right] = b$, U_1 un U_2 ir neatkarīgi, vienādi sadalīti gadījuma lielumi. Pie tam izpildās sekojošie nosacījumi:

- 1) $K : \mathbb{R} \longrightarrow \mathbb{R}$ ir ierobežota funkcija tāda, ka $\int |K(u)|du < \infty$, $\int K^2(u)du < \infty$ un $\lim_{u \rightarrow \infty} u^2 K(u) = 0$.
- 2) $\int K(u)du = 1$.
- 3) K ir nepārtraukti diferencējama.
- 4) χ ir nepārtraukta pēc Lipšīca.
- 5) $M_n = \max |x_{i+1} - x_i| = O(n^{-1})$, kur $1 \leq i \leq n - 1$
- 6) $\exists \{h_n\}$ virkne tāda, ka
- 6a) $\widehat{h}_n/h_n \xrightarrow{P} 1$,
- 6b) $\lim_{n \rightarrow \infty} (nh_n) = +\infty$ un $\lim_{n \rightarrow \infty} (h_n) = 0$.
Pie tam $\lim_{n \rightarrow \infty} \left(\frac{nh_n}{\log(n)}\right) = \infty$ un $\widehat{h}_n/h_n \xrightarrow{g.d.} 1$.

Teorēma 9.

$$(nh_n)^{\frac{1}{2}} \left[\widehat{\sigma}_{M,n}(x, \widehat{h}_n) - S(G_x) \right] \xrightarrow{D} N\left(0, v \int K^2(u)du\right).$$

Ja $M_n = \max |x_{i+1} - x_i| = O(n^{-1})$ un $\exists \{h_n\}$ tāda virkne, ka $\widehat{h}_n/h_n \xrightarrow{P} 1$, $\lim_{n \rightarrow \infty} (nh_n) = +\infty$ un $\lim_{n \rightarrow \infty} (h_n) = 0$. Pie tam

$$v = \frac{v_1}{v_2^2}$$

$$\begin{aligned} v_1 &= D\left[\chi\left(\frac{\sigma(x)(U_2 - U_1)}{aS(G_x)}\right)\right] + 2\beta cov\left[\chi\left(\frac{\sigma(x)(U_2 - U_1)}{aS(G_x)}\right), \right. \\ &\quad \left. \chi\left(\frac{\sigma(x)(U_4 - U_3)}{aS(G_x)}\right)\right], \\ v_2 &= E\left[\chi'\left(\frac{\sigma(x)(U_2 - U_1)}{aS(G_x)}\right)\left(\frac{\sigma(x)(U_2 - U_1)}{a(S(G_x))^2}\right)\right]. \end{aligned}$$

Izmantosim Teorēmas 9. rezultātu:

$$(nh_n)^{\frac{1}{2}} \left[\widehat{\sigma}_{M,n}(x, \widehat{h}_n) - S(G_x) \right] \xrightarrow{D} N\left(0, v \int K^2(u) du\right).$$

Pēc CRT:

$$\frac{(nh_n)^{\frac{1}{2}} [\widehat{\sigma}_{M,n} - S(G_x)]}{\sqrt{v \int K^2(u) du}} \xrightarrow{D} N(0, 1).$$

Tādējādi punktveida ticamības intervāls izskatās šādi:

$$P\left(-q_{0.975} \leq \frac{(nh_n)^{\frac{1}{2}} [\widehat{\sigma}_{M,n} - S(G_x)]}{\sqrt{v \int K^2(u) du}} \leq q_{0.975}\right) = 0.95,$$

$$P\left(\frac{-q_{0.975} \sqrt{v \int K^2(u) du}}{(nh_n)^{\frac{1}{2}}} \leq [\widehat{\sigma}_{M,n} - S(G_x)] \leq \frac{q_{0.975} \sqrt{v \int K^2(u) du}}{(nh_n)^{\frac{1}{2}}}\right) = 0.95,$$

$$P\left(-\widehat{\sigma}_{M,n} - \frac{q_{0.975} \sqrt{v \int K^2(u) du}}{(nh_n)^{\frac{1}{2}}} \leq -S(G_x) \leq \right.$$

$$\left. \leq -\widehat{\sigma}_{M,n} + \frac{q_{0.975} \sqrt{v \int K^2(u) du}}{(nh_n)^{\frac{1}{2}}}\right) = 0.95,$$

$$P\left(\widehat{\sigma}_{M,n} - \frac{q_{0.975} \sqrt{v \int K^2(u) du}}{(nh_n)^{\frac{1}{2}}} \leq S(G_x) \leq \widehat{\sigma}_{M,n} + \frac{q_{0.975} \sqrt{v \int K^2(u) du}}{(nh_n)^{\frac{1}{2}}}\right) = 0.95.$$

3.2. $\int K^2(u)du$ aprēķini

Rēķināsim integrāli $\int K^2(u)du$.

Tā kā $K(u)$ ir standarta normāla sadalījuma blīvuma funkcija,

$$\int K^2(u)du = \int \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \right)^2 du,$$

$$\int K^2(u)du = \int \frac{1}{2\pi} e^{-u^2} du,$$

$$\text{apzīmesim } u = \frac{t}{\sqrt{2}} \Rightarrow du = \frac{dt}{\sqrt{2}}.$$

$$\int \frac{1}{2\pi} e^{-u^2} du = \frac{1}{2\pi} \int e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2}},$$

$$\frac{1}{2\pi} \int e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2}} = \frac{1}{2\pi\sqrt{2}} \int e^{-\frac{t^2}{2}} dt,$$

Zināms, ka Puasona integrālis ir $\int e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$.

$$\frac{1}{2\pi\sqrt{2}} \int e^{-\frac{t^2}{2}} dt = \frac{\sqrt{2\pi}}{2\pi\sqrt{2}},$$

$$\Rightarrow \int K^2(u)du = \frac{1}{2\sqrt{\pi}}.$$

Līdz ar to punktveida ticamības intervāls vispārīga forma ir šāds:

$$P\left(\widehat{\sigma}_{M,n} - \frac{q_{0.975}\sqrt{v}2^{\frac{-1}{2}}\pi^{\frac{-1}{4}}}{(nh_n)^{\frac{1}{2}}} \leq S(G_x) \leq \widehat{\sigma}_{M,n} + \frac{q_{0.975}\sqrt{v}2^{\frac{-1}{2}}\pi^{\frac{-1}{4}}}{(nh_n)^{\frac{1}{2}}}\right) = 0.95.$$

Problēma ir atrast $S(G_x)$ vērtību un atbilstoši v_1, v_2 vērtības. Tas nav triviāls uzdevums un šajā darbā nebija izpildīts.

3.3. Punktveida ticamības intervālu konstruēšana, kad $\chi(x) = x^2$ ar butstrapa metodi

Atsaucoties uz DasGupta [21] publikāciju pieņemsim, ka X_1, X_2, \dots, X_n reprezentē īsto populācijas sadalījumu. Butstrapa ideja: ģenerēt (simulēt) daudz izlašu no dotās un aproksimēt statistikas T sadalījumu. Tas ir, izvēlēsimies no dotās izlases X_1, X_2, \dots, X_n jaunas i.i.d. izlases no empīriskās sadalījuma funkcijas \widehat{F}_n .

Apzīmēsim iegūtās butstrapa izlases $\{X_{11}^*, \dots, X_{1n}^*\}, \{X_{21}^*, \dots, X_{2n}^*\}, \dots, \{X_{B1}^*, \dots, X_{Bn}^*\}$, kur B apzīmē butstrapoto izlašu skaitu. Statistiskā funkcionāļa T sadalījumu punktā t , tas ir, $P_F(T \leq t)$ aproksimēsim ar $\{j \text{ skaits} : T_j^* \leq t\}/B$, kur $T_1^*, T_2^*, \dots, T_B^*$ apzīmē statistikas T vērtības B dažādajām butstrapa izlasēm.

Piemērs 3. Ja $T(X_1, X_2, \dots, X_n, F) = \sqrt{n}(\bar{X}_n - \mu)/\sigma$, tad attiecīgā butstrapotā statistika ir $T(X_1, X_2, \dots, X_n, F) = \sqrt{n}(\bar{X}_n^* - \bar{X}_n)/S$, kur S apzīmē izlases (empīrisko) dispersiju.

Atkarībā no vēlamās precizitātes un aprēķinu sarežģītības pakāpes ir dažādi veidi, kā konstruēt butstrapa ticamības intervālus.

Definīcija 14. Normālie butstrapa intervāli tiek definēti sekojoši

$$T_n \pm z_{\alpha/2} \widehat{s}e_{boot},$$

kur $\widehat{s}e_{boot}$ ir standartklūdas butstrapa novērtējums statistiskajam funkcionālim T_n . Šie intervāli ir precīzi gadījumā, ja T_n sadalījums ir tuvs Normālajam.

Piemērs 4. Ja $T_n = \bar{X}_n$, tad no Centrālās robežteorēmas ticamības intervāli vidējai vērtībai μ ir formā

$$\bar{X}_n \pm z_{\alpha/2} \sigma / \sqrt{n}.$$

Šajā gadījumā, lai konstruētu ticamības intervālus ir pietiekoši novērtēt parametru σ .

Definīcija 15. Butstrapa procentīļu intervāli tiek definēti, kā

$$C_n = (T_{(B\alpha/2)}^*, T_{(B(1-\alpha/2))}^*),$$

tas ir, lietojam $\alpha/2$ un $1 - \alpha/2$ kvantiles butstrapa izlasei.

Konstruējot butstrapa procentīļu intervālus klūdu dispersijas novērtējumam $\widehat{\sigma}_{RICE,n}(x)$, tika pieņemts, ka $T = \widehat{\sigma}_{RICE,n}(x)$.

4. Praktiskā daļa

4.1. Robežsadalījuma konverģence

Simulējot Y_i un x_i datus pie fiksēta punkta $x = 0.5$ var pārbaudīt robežsadalījuma konverģenci $(nh_n)^{\frac{1}{2}} \left[\widehat{\sigma}_{M,n}(x, \widehat{h}_n) - S(G_x) \right] \xrightarrow{D} N\left(0, v \int K^2(u)du\right)$ izmantojot dažādus testus.

Pārbaudīsim hipotēzes par normālo sadalījumu:

$$H_0 : (nh_n)^{\frac{1}{2}} \left[\widehat{\sigma}_{M,n}(x, \widehat{h}_n) - S(G_x) \right] \longrightarrow N(\mu, \sigma^2),$$

$$H_1 : (nh_n)^{\frac{1}{2}} \left[\widehat{\sigma}_{M,n}(x, \widehat{h}_n) - S(G_x) \right] \not\rightarrow N(\mu, \sigma^2).$$

Šī mērķa sasniegšanai izmantosim Lilliefora un Shapiro – Wilksa testus.

1. tabula p – vērtība, kad $N = 1000$, $\varepsilon = 0$, pie fiksēta $x = 0.5$

n	Lilliefors	Shapiro – Wilks
100	0.0016	0.0023
200	0.3927	0.4
400	0.6335	0.82

Hipotēzi H_0 , sākot ar $n = 200$ par normālo sadalījumu nevaram noraidīt pie nozīmības līmeni $\alpha = 0.05$.

2. tabula p – vērtība, kad $N = 1000$, $\varepsilon = 0.1$, pie fiksēta $x = 0.5$

n	Lilliefors	Shapiro – Wilks
100	0.29	0.0014
200	0.61	0.7328
400	0.77	0.7566

Hipotēzi H_0 , par normālo sadalījumu nevaram noraidīt sākot ar $n = 200$ pie nozīmības līmeni $\alpha = 0.05$.

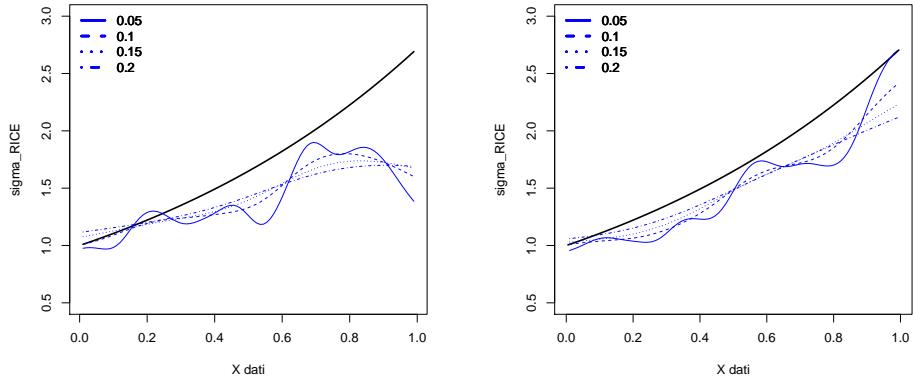
3. tabula p – vērtība, kad $N = 1000$, $\varepsilon = 0.2$, pie fiksēta $x = 0.5$

n	Lilliefors	Shapiro – Wilks
100	0.087	0.082
200	0.203	0.298
400	0.492	0.635

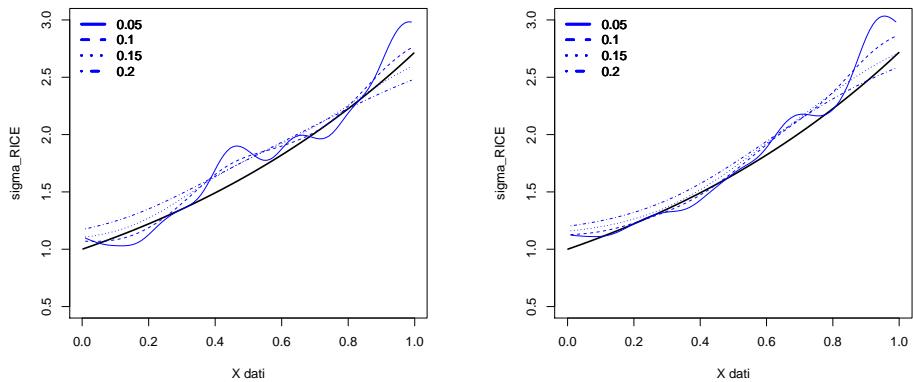
Redzam, ka palielinot izlases apjomu n un mainot ε vērtības nevaram noraidīt hipotēzi par normālo sadalījumu izmantojot Lilliefors un Shapiro – Wilksa testus pie nozīmības līmeni $\alpha = 0.05$ un $N = 1000$.

4.2. Gludinošā parametra h izvēle

Apskatīsim, ka izskatīsies $\widehat{\sigma}_{RICE,n}(x)$ novērtējums dažādiem n un h . Simulēsim datus $G(y) = (1 - \varepsilon)\Phi(y) + \varepsilon H(y)$, $\Phi(y) \sim N(0, 1)$, $H(y) \sim C(0, 16)$ pie zināmas īstas dispersijas funkcijas $\sigma(x) = e^x$. Nemsim $h = 0.05, 0.1, 0.15, 0.2$, $n = 100, 200, 400, 1000$ un $\varepsilon = 0$.



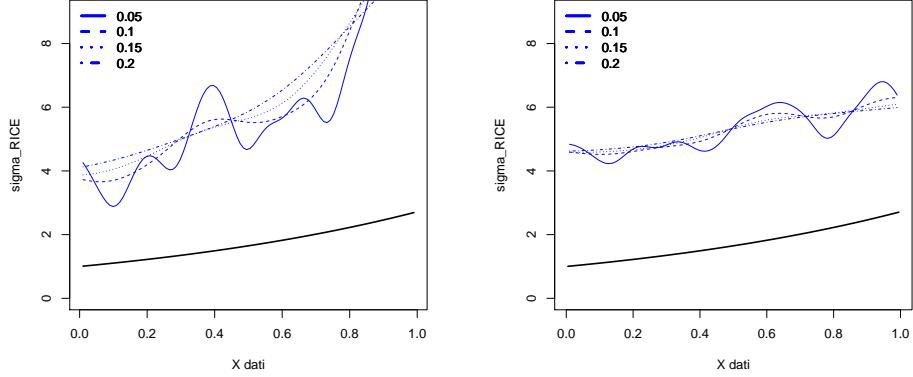
1. att.: $\widehat{\sigma}_{RICE,n}(x)$ grafiks pie $n = 100$ un $n = 200$. Trekna, melna līnija ir īsta dispersijas, punktētas zilas līnijas ir krosvalidācijas funkcijas pie dažāda h .



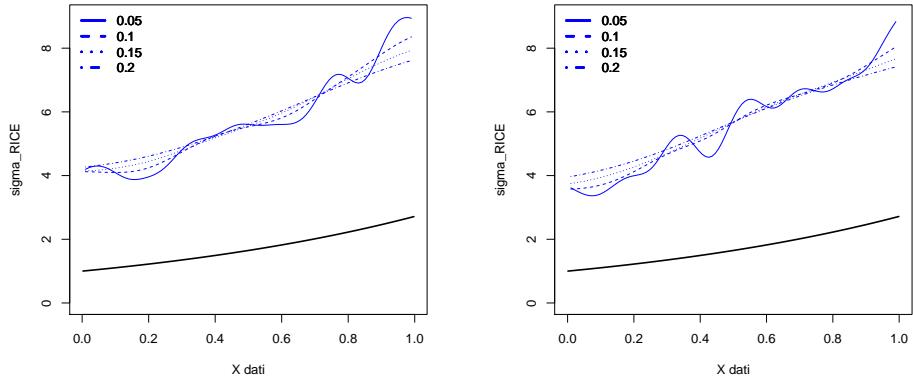
2. att.: $\widehat{\sigma}_{RICE,n}(x)$ grafiks pie $n = 400$ un $n = 1000$. Trekna, melna līnija ir īsta dispersijas, punktētas zilas līnijas ir krosvalidācijas funkcijas pie dažāda h .

Acīmredzams, ka palielinot izlases apjomu n novērtējums $\widehat{\sigma}_{RICE,n}(x)$ konverģē uz īsto dispersiju $\sigma(x)$ palielinot izlases apjomu n .

Simulēsim citus datus $G(y) = (1 - \varepsilon)\Phi(y) + \varepsilon H(y)$, $\Phi(y) \sim N(0, 1)$, $H(y) \sim N(10, 0.1)$ pie zināmas īstas dispersijas funkcijas $\sigma(x) = e^x$ nesmot $h = 0.05, 0.1, 0.15, 0.2$, $n = 100, 200, 400, 1000$ un $\varepsilon = 0.1$. Novērtējums $\widehat{\sigma}_{RICE,n}(x)$ izskatās šādi



3. att.: $\widehat{\sigma}_{RICE,n}(x)$ grafiks pie $n = 100$ un $n = 200$. Trekna, melna līnija ir īsta dispersijas, punktētas zilas līnijas ir krosvalidācijas funkcijas pie dažāda h .



4. att.: $\widehat{\sigma}_{RICE,n}(x)$ grafiks pie $n = 400$ un $n = 1000$. Trekna, melna līnija ir īsta dispersijas, punktētas zilas līnijas ir krosvalidācijas funkcijas pie dažāda h .

Dispersijas novērtējums nobīdās uz augšu. Kas principā atbilst publikācijas dotam rezultātam, rēķinot \widehat{ISEL} kļūdu:

$$\widehat{ISEL} = \frac{1}{n} \sum_{i=1}^n \left(\log \left(\frac{\widehat{\sigma}_n^j(x_i)}{\sigma(x_i)} \right) \right)^2,$$

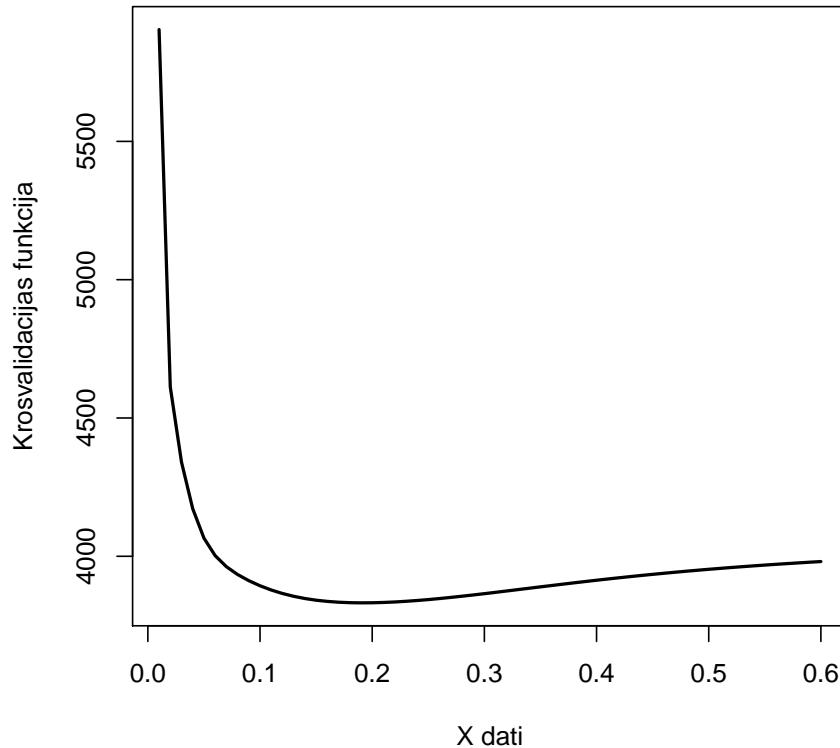
kur $\widehat{\sigma}_n^{(j)}$ norāda dispersijas novērtējumu, kurā iegūta j – tā simulācija ar gludinošo parametru h .

4. tabula: Vidēja \widehat{ISEL} vērtība, simulāciju skaits $N = 500$, $n = 100$, $G(y) = (1 - \varepsilon)\Phi(y) + \varepsilon H(y)$, $\Phi(y) \sim N(0, 1)$, $H(y) \sim N(10, 0.1)$

novērtējums	$\varepsilon = 0.1$	$\varepsilon = 0.2$	$\varepsilon = 0.3$	$\varepsilon = 0.35$	$\varepsilon = 0.4$
$\widehat{\sigma}_{RICE,n}(x)$	1.353	2.060	2.453	2.574	2.656

Tālāk vajadzēs noskaidrot optimālo joslas platumu h katrā izlases gadījumā.

Pielietojot vienu — atstāt — ārā krosvalidācijas funkciju simulētiem datiem $n = 100$, $\varepsilon = 0.1$, $G(y) = (1 - \varepsilon)\Phi(y) + \varepsilon H(y)$, $\Phi(y) \sim N(0, 1)$, $H(y) \sim N(10, 0.1)$ iegūstam sekojošo grafiku



5. att. $\varepsilon = 0.1$, $n = 100$

Minimāla h vērtība ir 0.191. Tādējādi novērtējot gludinošo parametru h mēs varēsim pielietot izpētīto metodi gan simulētiem datiem, gan reāliem. Kā arī salīdzināt publikācijās dotas vidējas joslas pluma h un kļūdas \widehat{ISEL} vērtības ar mūsu iegūtajiem.

Atsaucoties uz Boente, Ruiz, Zamar jauno publikāciju [13], zinām vidējas joslas platura h un klūdas \widehat{ISEL} vērtības simulētiem datiem M1) un M2) modelim, pie $N = 500$. Salīdzināsim tos ar mūsu iegūtiem.

5. tabula: Vidēja vērtība gludinošam parametram h un \widehat{ISEL} klūdai. Apskatāmais modelis ir M1: $g(x) = 2\sin(4\pi x)$, $\sigma(x) = e^x$. Pie tam $H(y) = N(10, 0.1)$. Permutāciju skaits $N = 500$, izlases apjoms $n = 100$.

ε	\widehat{h}_{CV} publikācijā	\widehat{h}_{CV} mūsu rezultāts	\widehat{ISEL} publikācijā	\widehat{ISEL} mūsu rezultāts
$\varepsilon = 0.1$	0.267	0.332	1.313	1.339
$\varepsilon = 0.2$	0.206	0.233	2.029	2.030
$\varepsilon = 0.3$	0.164	0.190	2.407	2.411

6. tabula: Vidēja vērtība gludinošam parametram h un \widehat{ISEL} klūdai. Apskatāmais modelis ir M1: $g(x) = 2\sin(4\pi x)$, $\sigma(x) = e^x$. Pie tam $H(y) = N(10, 0.1)$. Permutāciju skaits $N = 500$, izlases apjoms $n = 100$.

ε	\widehat{h}_{CV}^{\log} publikācijā	\widehat{h}_{CV}^{\log} mūsu rezultāts	\widehat{ISEL} publikācijā	\widehat{ISEL} mūsu rezultāts
$\varepsilon = 0.1$	0.058	0.108	1.152	1.152
$\varepsilon = 0.2$	0.059	0.083	1.892	1.823
$\varepsilon = 0.3$	0.058	0.071	2.309	2.250

No 5. un 6. tabulas redzam, ka modelim M1 pie $H(y) = N(10, 0.1)$ vērtības \widehat{h}_{CV} un \widehat{ISEL} dotās publikācijā ir līdzīgas mūsu iegūtajiem. Palielinot ε gludinoša parametra h vērtības \widehat{h}_{CV} samazinās, bet \widehat{ISEL} klūdas vērtības palielinās. Salīdzinot 5. un 6. tabulas redzam, ka nēmot logaritmu h vērtība klūst ievērojami mazāk.

7. tabula: Vidēja vērtība gludinošam parametram h un \widehat{ISEL} kļūdai. Apskatāmais modelis ir M2: $g(x) = 1 + x$, $\sigma(x) = 1 + (1 + \sin(10x))^2$. Pie tam $H(y) = N(10, 0.1)$. Permutāciju skaits $N = 500$, izlases apjoms $n = 100$.

ε	\widehat{h}_{CV} publikācijā	\widehat{h}_{CV} mūsu rezultāts	\widehat{ISEL} publikācijā	\widehat{ISEL} mūsu rezultāts
$\varepsilon = 0.1$	0.232	0.349	1.711	1.889
$\varepsilon = 0.2$	0.096	0.152	2.303	2.367
$\varepsilon = 0.3$	0.059	0.071	2.636	2.636

8. tabula: Vidēja vērtība gludinošam parametram h un \widehat{ISEL} kļūdai. Apskatāmais modelis ir M2: $g(x) = 1 + x$, $\sigma(x) = 1 + (1 + \sin(10x))^2$. Pie tam $H(y) = N(10, 0.1)$. Permutāciju skaits $N = 500$, izlases apjoms $n = 100$.

ε	\widehat{h}_{CV}^{\log} publikācijā	\widehat{h}_{CV}^{\log} mūsu rezultāts	\widehat{ISEL} publikācijā	\widehat{ISEL} mūsu rezultāts
$\varepsilon = 0.1$	0.034	0.098	1.190	1.241
$\varepsilon = 0.2$	0.032	0.055	1.955	1.866
$\varepsilon = 0.3$	0.031	0.052	2.384	2.310

No 7. un 8. tabulas redzam, ka modelim M2 pie $H(y) = N(10, 0.1)$ vērtības \widehat{h}_{CV} un \widehat{ISEL} dotās publikācijā ir līdzīgas mūsu iegūtajiem. Palielinot ε gludinoša parametra h vērtības \widehat{h}_{CV} samazinās, bet \widehat{ISEL} kļūdas vērtības palielinās. Salīdzinot 7. un 8. tabulas redzam, ka ņemot logaritmu h vērtība kļūst ievērojami mazāk.

9. tabula: Vidēja vērtība gludinošam parametram h un \widehat{ISEL} kļūdai. Apskatāmais modelis ir M1: $g(x) = 2\sin(4\pi x)$, $\sigma(x) = e^x$. Pie tam $H(y) = C(0, 16)$. Permutāciju skaits $N = 500$, izlases apjoms $n = 100$.

ε	\widehat{h}_{CV} publikācijā	\widehat{h}_{CV} mūsu rezultāts	\widehat{ISEL} publikācijā	\widehat{ISEL} mūsu rezultāts
$\varepsilon = 0$	0.190	0.212	0.032	0.030
$\varepsilon = 0.1$	0.430	0.520	4.28	10.735
$\varepsilon = 0.2$	0.435	0.517	7.121	15.984
$\varepsilon = 0.3$	0.433	0.534	9.362	19.976

10. tabula: Vidēja vērtība gludinošam parametram h un \widehat{ISEL} kļūdai. Apskatāmais modelis ir M2: $g(x) = 1 + x$, $\sigma(x) = 1 + (1 + \sin(10x))^2$. Pie tam $H(y) = C(0, 16)$. Permutāciju skaits $N = 500$, izlases apjoms $n = 100$.

ε	\widehat{h}_{CV} publikācijā	\widehat{h}_{CV} mūsu rezultāts	\widehat{ISEL} publikācijā	\widehat{ISEL} mūsu rezultāts
$\varepsilon = 0$	0.069	0.109	0.097	0.124
$\varepsilon = 0.1$	0.421	0.530	4.877	11.6293
$\varepsilon = 0.2$	0.434	0.532	7.781	17.1741
$\varepsilon = 0.3$	0.439	0.530	10.063	21.3725

No 9. un 10. tabulas redzam, ka \widehat{h}_{CV} atšķiras aptuveni uz 0.1. \widehat{ISEL} kļūdas atšķiras vairāk, bet kopējā tendence saglabājas. Palielinot ε vērtību, samazinās \widehat{h}_{CV} un palielinās \widehat{ISEL} vērtība. Jo mazāka ir \widehat{ISEL} vērtība, jo labāk. Vismazākās \widehat{ISEL} vērtības ir pie $\varepsilon = 0$ vai $\varepsilon = 0.1$, bet vislielākās pie pārejas ε vērtības.

Apskatīsim, ka uzvedas krosvalidacijas funkcija pie dažādiem n vērtībām. Pēc idejas palielinot izlases apjomu n , joslas platumam h jāsamazinās. Tas ir viens no panēmieniem, kurš dod iespēju saprast, vai mēs pareizi novērtējam h caur vienu – atstāt – ārā metodi.

11. tabula: Vidēja vērtība gludinošam parametram h un \widehat{ISEL} klūdai. Apskatāmais modelis ir M1: $g(x) = 2\sin(4\pi x)$, $\sigma(x) = e^x$. Pie tam $H(y) = N(10, 0.1)$. Permutāciju skaits $N = 500$.

ε	$\widehat{h}_{CV}, n = 50$	$\widehat{h}_{CV}, n = 100$	$\widehat{h}_{CV}, n = 200$
$\varepsilon = 0.1$	1.042	0.332	0.248
$\varepsilon = 0.2$	0.642	0.233	0.170
$\varepsilon = 0.3$	0.430	0.190	0.141

12. tabula: Vidēja vērtība gludinošam parametram h un \widehat{ISEL} klūdai. Apskatāmais modelis ir M1: $g(x) = 2\sin(4\pi x)$, $\sigma(x) = e^x$. Pie tam $H(y) = N(10, 0.1)$. Permutāciju skaits $N = 500$.

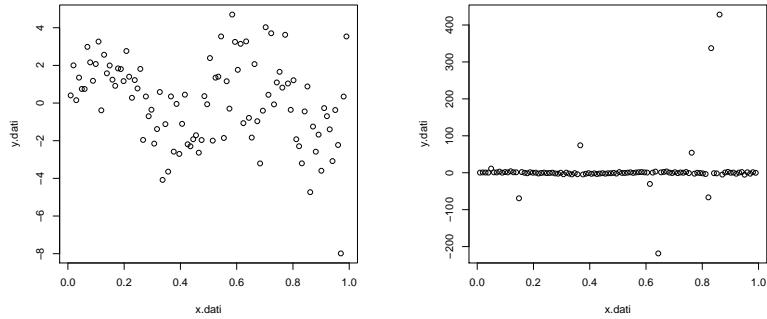
ε	$\widehat{ISEL}, n = 50$	$\widehat{ISEL}, n = 100$	$\widehat{ISEL}, n = 200$
$\varepsilon = 0.1$	1.326	1.339	1.344
$\varepsilon = 0.2$	2.000	2.030	2.024
$\varepsilon = 0.3$	2.390	2.411	2.401

No 11. un 12. tabulas redzam, ka palielinot izlases apjomu n gludinoša parametra vērtības \widehat{h}_{CV} dilst, bet klūdas \widehat{ISEL} aug, jebkuram ε vērtībām.

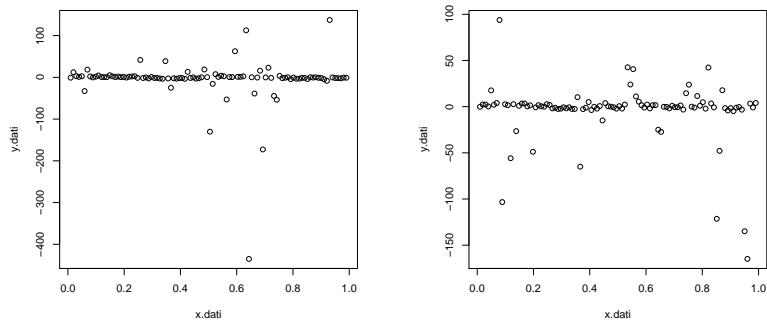
4.3. Klūdu dispersijas novērtējumu pielietojums simulētiem datiem

Iepriekšējas nodaļas tika aprakstīts, dažādu dispersijas funkciju novērtējumi. Šajā apakšnodaļā aplūkosim šīs metodēs pielietojumu simulētiem datiem. Simulēsim datus atbilstoši M1) modelim pie $H(y) = C(0, 16)$.

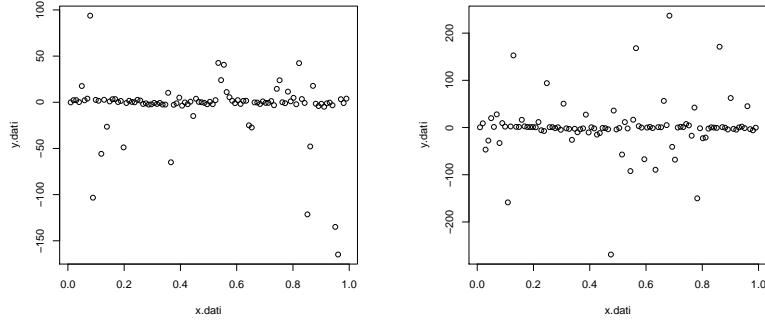
Paņemot $\varepsilon = 0, 0.1, 0.2, 0.3, 0.35, 0.4$



6. att. Datu grafiks pie $\varepsilon = 0$ un $\varepsilon = 0.1$.

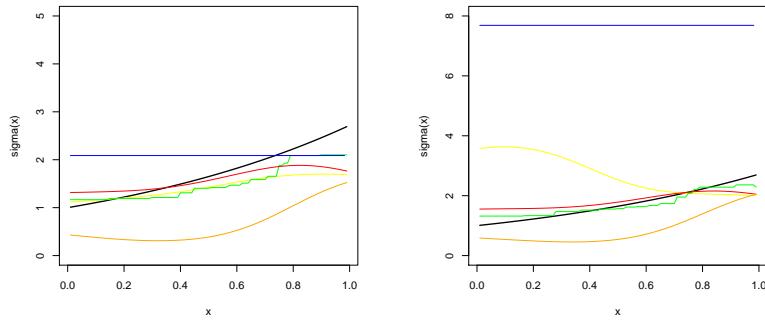


7. att. Datu grafiks pie $\varepsilon = 0.2$ un $\varepsilon = 0.3$.

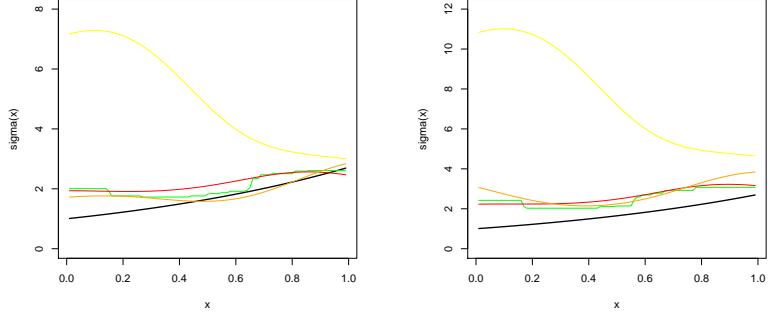


8. att. Datu grafiks pie $\varepsilon = 0.35$ un $\varepsilon = 0.4$.

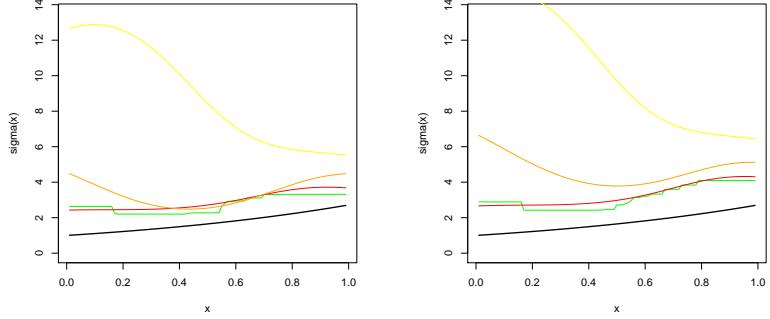
Palielinot ε vērtību datiem simulejusi vairāk izlēcēji. Uzzīmēsim īstas dispersijas funkcijas grafiku e^x simulētiem datiem $G(y) = (1 - \varepsilon)\Phi(y) + \varepsilon H(y)$, $\Phi(y) \sim N(0, 1)$, $H(y) \sim C(0, 4^2)$ ar $\widehat{\sigma}_{RICE,n}(x)$, $\widehat{\sigma}_{MSD,n}(x)$, $\widehat{\sigma}_{MBT,n}(x)$, $\widehat{\sigma}_n(x)$ un $\widehat{\sigma}_{const,n}(x)$.



9. att.: Simulētiem datiem, dispersijas novērtējums kā funkcijas pie $\varepsilon = 0$ un $\varepsilon = 0.1$. Melna līnija ir īsta dispersija e^x , $\widehat{\sigma}_{RICE,n}(x)$ dzeltena, $\widehat{\sigma}_{MSD,n}(x)$ zaļa, $\widehat{\sigma}_{MBT,n}(x)$ sarkana, $\widehat{\sigma}_{const,n}(x)$ zila un $\widehat{\sigma}_n(x)$ oranža.



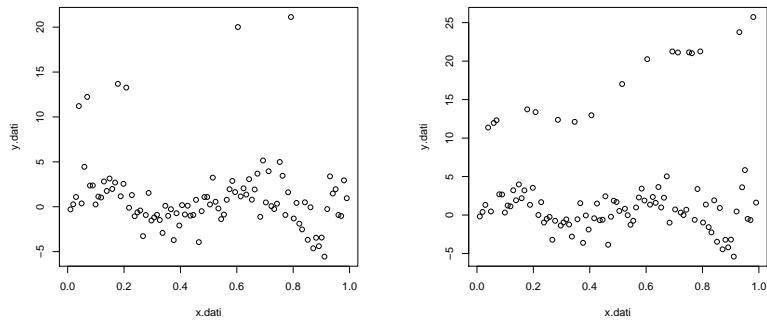
10. att.: Simulētiem datiem, dispersijas novērtējums kā funkcijas pie $\varepsilon = 0.2$, $\widehat{\sigma}_{const,n}(x) = 27.63$ un $\varepsilon = 0.3$, $\widehat{\sigma}_{const,n}(x) = 63.18$. Melna līnija ir īsta dispersija $exp(x)$, $\widehat{\sigma}_{RICE,n}(x)$ dzeltena, $\widehat{\sigma}_{MSD,n}(x)$ zaļa, $\widehat{\sigma}_{MBT,n}(x)$ sarkana un $\widehat{\sigma}_n(x)$ oranža.



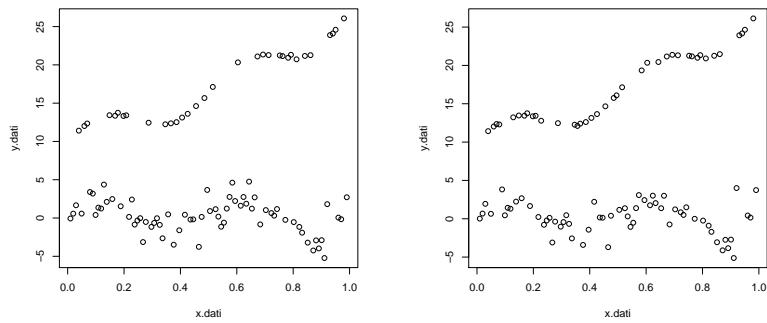
11. att.: Simulētiem datiem, dispersijas novērtējums kā funkcijas pie $\varepsilon = 0.35$, $\widehat{\sigma}_{const,n}(x) = 86.93$ un $\varepsilon = 0.4$, $\widehat{\sigma}_{const,n}(x) = 114.72$. Melna līnija ir īsta dispersija $exp(x)$, $\widehat{\sigma}_{RICE,n}(x)$ dzeltena, $\widehat{\sigma}_{MSD,n}(x)$ zaļa, $\widehat{\sigma}_{MBT,n}(x)$ sarkana un $\widehat{\sigma}_n(x)$ oranža.

Vislabāk aproksimē īsto dispersiju $\sigma(x) = e^x$ novērtējumi $\widehat{\sigma}_{MSD,n}(x)$ un $\widehat{\sigma}_{MBT,n}(x)$, jo tās līnijas iet pietiekoši tuvu melnai un dažās vietās sakrīt ar $\sigma(x)$ līniju. Vissliktāk aproksimē $\widehat{\sigma}_{const,n}(x)$, jo zila līnija iet ļoti tālu no īstas dispersijas līnijas. $\widehat{\sigma}_{RICE,n}(x)$ aproksimē īsto dispersiju, pietiekoši tuvi pie $\varepsilon = 0$, bet palielinot ε vērtību šīs novērtējums nobīdās uz augšu, kas principā atbilst kļūdas \widehat{ISEL} aprēķinātajām 10. tabulas vērtībām. Līdz ar to nevar viennozīmīgi pateikt par labāko dispersijas novērtējumu funkciju. To iemeslu dēļ, apskatīsim metodes pielietojumu M1) modelim pie $\sim N(\mu, \sigma^2)$.

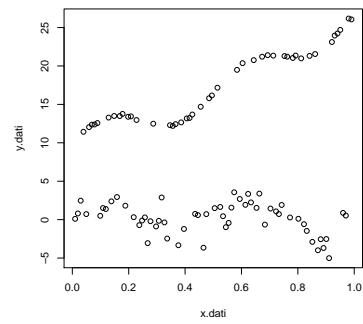
Simulēsim datus $G(y) = (1 - \varepsilon)\Phi(y) + \varepsilon H(y)$, $\Phi(y) \sim N(0, 1)$, $H(y) \sim N(\mu, \sigma^2)$, kur $\sigma = 0.1$



12. att. Datu grafiks pie $\varepsilon = 0.1$ un $\varepsilon = 0.2$, $\mu = 10$.

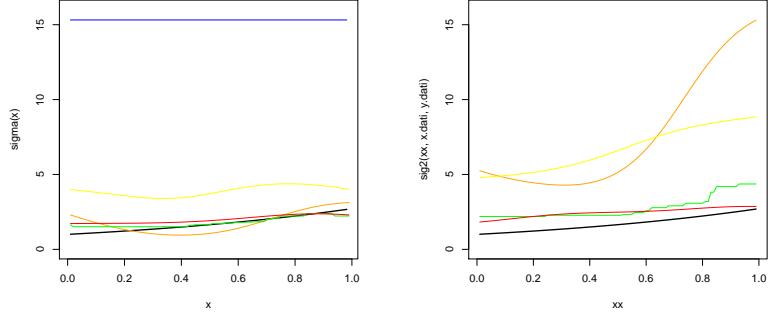


13. att. Datu grafiks pie $\varepsilon = 0.3$ un $\varepsilon = 0.35$, $\mu = 10$.

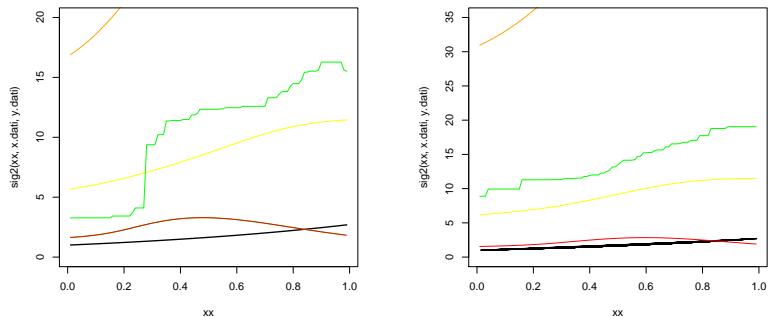


14. att. Datu grafiks pie $\varepsilon = 0.4$, $\mu = 10$.

Uzzīmēsim īstas dispersijas funkcijas grafiku e^x simulētiem datiem $G(y) = (1 - \varepsilon)\Phi(y) + \varepsilon H(y)$, $\Phi(y) \sim N(0, 1)$, $H(y) \sim N(\mu, \sigma^2)$ ar $\widehat{\sigma}_{RICE,n}(x)$, $\widehat{\sigma}_{MSD,n}(x)$, $\widehat{\sigma}_{MBT,n}(x)$, $\widehat{\sigma}_n(x)$ un $\widehat{\sigma}_{const,n}(x)$.



15. att.: Simulētiem datiem, dispersijas novērtējums kā funkcijas pie $\varepsilon = 0.1$ un $\varepsilon = 0.2$, $\mu = 10$. Melna līnija ir īsta dispersija $exp(x)$, $\widehat{\sigma}_{RICE,n}(x)$ dzeltena, $\widehat{\sigma}_{MSD,n}(x)$ zaļa, $\widehat{\sigma}_{MBT,n}(x)$ sarkana, $\widehat{\sigma}_n(x)$ oranža, $\widehat{\sigma}_{const,n}(x)$ zila, pie $\varepsilon = 0.2$ $\widehat{\sigma}_{const,n}(x) = 48.023$.



16. att.: Simulētiem datiem, dispersijas novērtējums kā funkcijas pie $\varepsilon = 0.3$ un $\varepsilon = 0.35$, $\mu = 10$. Melna līnija ir īsta dispersija $exp(x)$, $\widehat{\sigma}_{RICE,n}(x)$ dzeltena, $\widehat{\sigma}_{MSD,n}(x)$ zaļa, $\widehat{\sigma}_{MBT,n}(x)$ sarkana, $\widehat{\sigma}_n(x)$ oranža, pie $\varepsilon = 0.3$ $\widehat{\sigma}_{const,n}(x) = 80.11$, $\varepsilon = 0.35$ $\widehat{\sigma}_{const,n}(x) = 86.26$.

Vislabāk aproksimē īsto dispersiju $\widehat{\sigma}_{MSD,n}(x)$, jo tā līnija iet pietiekoši tuvu visiem ε vērtībām un dažās vietās sakrīt ar e^x līniju. Vissliktāk aproksimē $\widehat{\sigma}_{const,n}(x)$, jo zila līnija iet ļoti tālu no īstas dispersijas līnijas.

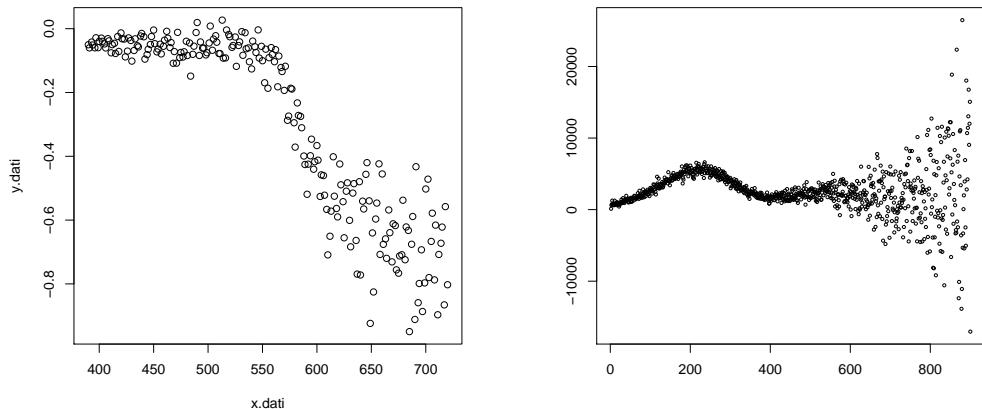
4.4. Klūdu dispersijas novērtējumu pielietojums reāliem datiem

Pielietosim aplūkoto metodi reāliem datiem LIDAR un CMB, kurus ņemam no Larry Wasserman grāmatas [1].

LIDAR dati ir iegūti no gaismas detektēšanas un apgabalu noteikšanas. LIDAR dati tiek lietoti, lai noteiktu piesārņojumu. X_i dati attēlo mērijuma distanci un Y_i dati reprezentē mērijuma ātrumu.

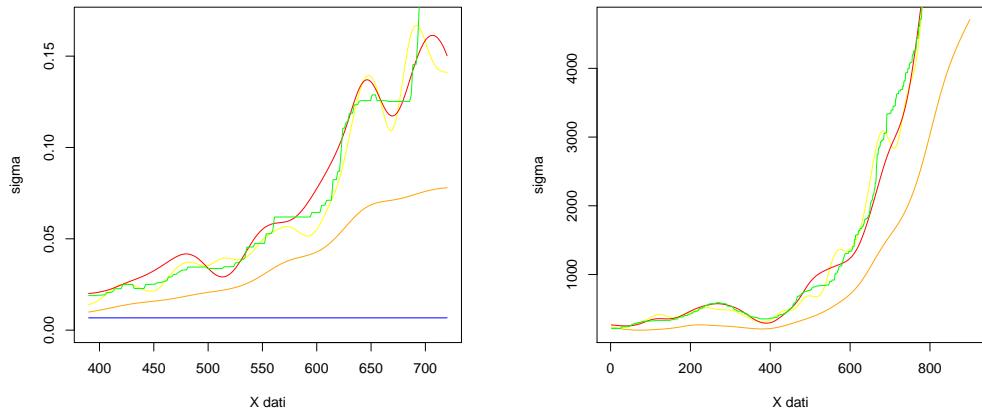
CMB ir kosmiskas radiācijas dati, kuri rodas lielo sprādzienu rezultātā. X_i dati attēlo temperatūras fluktuācijas frekvenci un Y_i dati reprezentē fluktuācijas spēku katrā frekvencē. Fluktuācija ir fizikāla lieluma nejauša novirze no vidējās vērtības. Fluktuācija parādās procesos, kas pakļauti statistikas likumiem (molekulu, atomu, elektronu u. c. daļiņu kustība dažādās sistēmās). Daudzas fizikālas parādības var izskaidrot tikai ar fluktuācijas palīdzību, piemēram, Brauna kustību, gaismas molekulāro izkliedi, kas nosaka debesu zilo krāsu.

Sākuma dati izskatās šādi



17. att. LIDAR un CMB dati

Novērtējot dispersijas funkciju ar mūsu apskatītājam metodēm iegūstam:

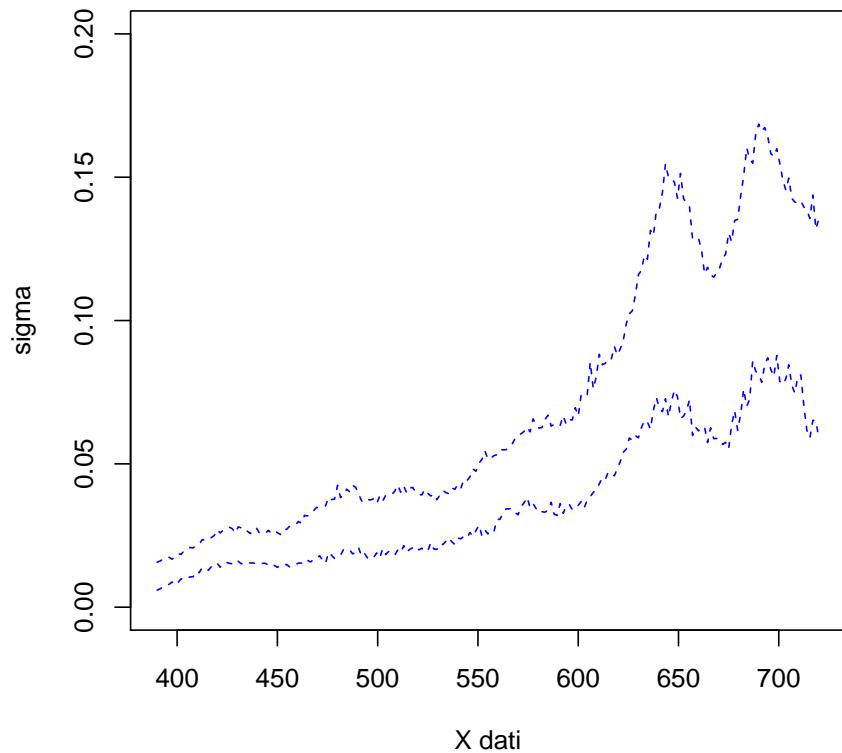


18. att.: LIDAR un CMB datiem, dispersijas novērtējums, kā funkcijas. $\widehat{\sigma}_{RICE,n}(x)$ dzeltena, $\widehat{\sigma}_{MSD,n}(x)$ zaļa, $\widehat{\sigma}_{MBT,n}(x)$ sarkana, $\widehat{\sigma}_n(x)$ oranža, $\widehat{\sigma}_{const,n}(x)$ zila, CMB datiem $\widehat{\sigma}_{const,n}(x)$ vērtība ir 3166.

Redzam, ka lokālie M – novērtējumi uzvedas līdzīgi LIDAR un CMB datiem. Sa-vukārt $\widehat{\sigma}_n(x)$ novērtējums iet zemāk nekā $\widehat{\sigma}_{MBT,n}(x)$, $\widehat{\sigma}_{RICE,n}(x)$, $\widehat{\sigma}_{MSD,n}(x)$ novērtējumi. $\widehat{\sigma}_{const,n}(x)$ novērtējums atrodas tālu no pārējiem novērtējumiem. Šie rezultāti atgādina situāciju, kurā bija ar simulētiem datiem. Tas varētu nozīmēt, kā lokālie M – novērtējumi attaisno sevi datiem, kuriem nav konstanta dispersija un tos ir lietderīgi izmantot kļūdu dispersijas novērtējumam neparametriskas regresijas modelī.

4.5. Novērtējuma $\widehat{\sigma}_{RICE,n}(x)$ punktveida ticamības intervālu konstruēšana LIDAR datiem izmantojot butstrapa metodi

Pieņemsim, ka $B = 200$. Tad butstrapa 95% procentīlu intervāli izskatīsies sekojoši:



19. att.: LIDAR datiem, dispersijas novērtējums. $\widehat{\sigma}_{RICE,n}(x)$ novērtējumam punktveida ticamības intervāli kļūdu dispersijai. Ticamības intervāli ir zilas punktētas līnijas.

Tiek, secināts, ka nevar noraidīt nulles hipotēzi par konstantu kļūdas dispersiju, jo nevar novilkrt horizontālu līkni starp ticamības intervāliem.

Secinājumi

Darbā tika konstruēti dispersijas novērtējumi simulētiem un reāliem datiem, izmantojot dažādas pieejas. Tika apskatīti dispersijas lokālie M – novērtējumi ($\widehat{\sigma}_{MSD,n}(x)$, $\widehat{\sigma}_{MBT,n}(x)$, $\widehat{\sigma}_{RICE,n}(x)$) salīdzinājumā ar Yu un Jones [2] piedāvāto metodi $\widehat{\sigma}_n(x)$. Simulētiem datiem pie dažāda sadalījuma $H(y)$ izvēles, dispersijas funkciju labāk aproksimē lokālie M – novērtējumi. Tas izskaidrojams, ar to, ka lokālajos M novērtējumos tiek izmantota robusta statistika, bet $\widehat{\sigma}_n(x)$ tā netiek izmantota. Apskatot dispersijas novērtējumu ar konstanti $\widehat{\sigma}_{const}(x)$, esam pārliecinājušies, ka tā neatnaisno sevi, jo gan simulētiem datiem, gan reāliem, tā atrodas pārāk tālu no īstās dispersijas līnijas.

Dispersijas funkcijas $\sigma(x)$ novērtējums pēc robusta ir svarīga problēma neparametriskās regresijas analīzē. Viennozīmīgi no salīdzinātajām M – novērtējuma metodēm $\widehat{\sigma}_{MSD,n}(x)$ ir vislabākā, jo visos simulētajos datu grafikos tā atrodas tuvāk īstajai dispersijas funkcijai nekā parējās līnijas.

Runājot par lokāliem M – novērtējumiem, jāpievērš uzmanību gludinošā parametra h un svaru $w_i(x)$ izvēlei. Simulētiem un reāliem datiem mēs nesakām h izmantojot vienu – atstāt – ārā krosvalidācijas metodi. Teorētiski h var izvēlēties arī pēc citām metodēm un paskatīties vai rezultāti uzlabosies.

Rakstot diplomdarbu, tika mēģināts izvest $S(G_x)$, ν_1 , ν_2 formulas priekš $\widehat{\sigma}_{RICE,n}(x)$ dispersijas novērtējumam un uzkonstruēt šīm novērtējumam punktveida ticamības intervālus, pamatojoties uz I. Ghement, M. Ruiz, R. Zamar publikācijas [17] rezultātiem. Šie rezultāti bija pārbaudīti ar Kolmogorova - Smirnova (Lillifora) un Shapirova testiem. Līdz ar to varam secināt, ka pie lieliem n simulētais sadalījums tiecas uz normālo sadalījumu, kas liecina par teorētisko rezultātu patiesumu. Funkcionāla $S(G_x)$ sarežģītības dabas dēļ, punktveida ticamības intervāli nebija konstruēti izmantojot ν_1 , ν_2 .

Izmantojot butstrapa metodi tika konstruēti butstrapa procentīlu intervāli lokālajam $\widehat{\sigma}_{RICE,n}(x)$ novērtējumam ar programmas R palīdzību.

Darba turpinājumā varētu nodarboties ar lokālo M – novērtējumu uzlabojumu, nemot citu χ funkciju un svarus $w_i(x)$ ar citiem paņēmieniem. Funkcijas $S(G_x)$ dziļo izpēte un uzlabotus punktveida ticamības intervālu konstruēšana novērtējumam $\widehat{\sigma}_{RICE,n}(x)$ kā arī punktveida ticamības intervālu konstruēšanu $\widehat{\sigma}_{MSD,n}(x)$ un $\widehat{\sigma}_{MBT,n}(x)$ funkcijām.

Izmantotā literatūra un avoti

- [1] L. Wasserman. All of nonparametric statistics. 99:80–120, 2006.
- [2] M. Jones. K. Yu. Likelihood-based local linear estimation of the conditional variance function. *Journal of the American Statistical Association*, 99:139–144, 2004.
- [3] M. Levine. L. Brown. Varinace estimation in nonparametric regression via the difference sequence method. *Annals of Statistics*, 35:2219–2232, 2007.
- [4] J. Rice. Bandwidth choice for nonparametric regression. *Annals of statistics*, 12:1215–1230, 1984.
- [5] J. Meloche G. Boente, R. Fraiman. Robust plug-in bandwidth estimators in nonparametric regression. *Journal of Statistical Planning and Inference*, 57:109–142, 1997.
- [6] T. Lee J. Hanning. Robust sizer for exploration of regression structures and outliers detection. *Journal of Computational and Graphical Statistics*, 15:101–117, 2006.
- [7] E. Ronchetti. E. Cantoni. Resistant selection of the smoothing parameter for smoothing splines. *Statistics and Computing*, 11:141–146, 2001.
- [8] D. Leung. Cross-validation in nonparametric regression with outliers. *Annals of Statistics*, 33:2291–2310, 2005.
- [9] T. Gasser W. Hardle. Robust nonparametric function fitting. *Journal of the Royal Statistical Society, Series B* 46, 9:42–51, 1984.
- [10] A. Tsybakov W. Hardle. Robust nonparametric regression with simultaneous scale curve estimation. *Annals of Statistics*, 25:443–456, 1988.
- [11] R. Fraiman. G. Boente. Robust nonparametric regression estimation for dependent observations. *Annals of statistics*, 17:1242–1256, 1989.

- [12] R. Zamar G. Boente, M. Ruiz. On a robust local estimator for the scale function in heteroscedastic nonparametric regression. *Annals of Statistics*, 30:50–211, 2010.
- [13] R. Zamar G. Boente, M. Ruiz. Bandwidth choice for robust nonparametric scale function estimation. *Annals of Statistics*, 30:1567–1877, 2011.
- [14] J. Victor A. Ricardo, R. Douglas. Robust statistics: Theory and methods. *Annals of statistics*, 30:17–23, 2006.
- [15] P. Huber. Robust estimation of a location parameter. *Annals of Statistics*, 35:73–101, 1964.
- [16] D. Titterington P. Hall, J. Kay. Asymptotically optimal difference-based estimation of variance in nonparamteric regression. *Biometrika*, 77:521–528, 1990.
- [17] R. Zamar I. Ghement, M. Ruiz. Robust estimation of error scale in nonparametric regression models. *Journal of Statistical Planning and Inference*, 138:3200–3216, 2008.
- [18] W. Hardle. Applied nonparametric regression. *Cambridge*, 30:10–103, 1990.
- [19] S. Sperlich A. Werwatz W. Hardle, M. Muller. Nonparametric and semiparametric models. *Springer*, 30:57–344, 2004.
- [20] M. Levine. Bandwidth selection for a class of difference-based variance estimators in the non- parametric regression: A possible approach. *Comput. Statist. Data Anal.*, 50:3405–3431, 2006.
- [21] A. DasGupta. Asymptotic theory of statistics and probability. *Springer*, 30:121–143, 2008.

Programmas R kods

```
x.dat1<-function(n) #x datu genereshana
{
  1:n/(n+1)
}

g<-function(x) #regresijas f-jas g(x) defineshana
{
  2*sin(4*pi*x)
}

sigma<-function(x) #dispersijas f-jas sigma(x) defineshana
{
  exp(x)
}

#Yi vertibas defineshana
# eps ir parametrs epsilon robezhas no 0 lidz 0.4

y.dat1<-function(n,eps,x.dat){
  x<-runif(n,0,1)
  Gc<-c()
  for (i in 1:n)
  {
    Gc[i]<-uniroot(function(y) (1-eps)*pnorm(y,0,1)+eps*pnorm(y,10,0.1)-x[i],c(-1200,1200))$root
  }

  g(x.dat)+Gc*sigma(x.dat) # y datu genereshana
}

#####
#1. metode
#Rice funkcija

RICE<-function(a,x,h,x.dat,y.dat)
{
  NW.svari<-function(x,x.dat,y.dat,h,i)
  {
    dnorm((x-x.dat[i])/h)/sum(dnorm((x-x.dat)/h))
  }
  NW.svari<-Vectorize(NW.svari,vectorize.args="x")

  NW.svari2<-Vectorize(NW.svari,vectorize.args="i")

  sigLRE<-function(x,x.dat,y.dat,h)
  {
```

```

n<-length(x.dat)
sum(NW.svari2(x,x.dat,y.dat,h,1:(n-1))*((diff(y.dat))/a)^2)
}
sigLRE<-Vectorize(sigLRE,vectorize.args="x")

sigLRE(x,x.dat,y.dat,h)
}

#####
#CV(h) funckija
#CV funkcija RICE metodei

CV<-function(h)
{
n<-length(x.dat)
s<-0
for (i in 1:(n-1))
{
s<-s+(((abs(y.dat[i+1]-y.dat[i]))/a )^2-RICE(a,x.dat[i],h,x.dat[-i][-i],y.dat[-i][-i]))^2
}
s/(n-1)
}
CV<-Vectorize(CV)

#####
#atsevishkam gadijumam h izvele
a<-sqrt(2)
n<-100
x<-seq(min(x.dat),max(x.dat),by=(max(x.dat)-min(x.dat))/(n-1))
length(x)

hh<-seq(0.01,0.6,by=0.01)
length(hh)

eps<-0
#saakuma datu genereshana
x.dat<-x.dat[1:n]
y.dat<-y.dat[1:n,eps,x.dat]

plot(x.dat,y.dat)

CV1<-CV(hh) #f-jas CV vertibas

plot(hh,CV1,type="l",lwd=2,xlab="X dati",ylab="Krosvalidacijas funkcija")

#grafiks atkaribaa no dazadiem h vertibam

```

```

h.opt<-optimize(CV,c(0.01,0.6))$minimum
h.opt #mazaka h vertiba

#####
#Ticamibas joslas konstrueshana

library(KernSmooth)

X<-matrix(data = NA, nrow = n, ncol = 3, byrow = FALSE,
           dimnames = NULL)

sk<-seq(1,n,by=1)

X[,1]<-x.dat
X[,2]<-y.dat
X[,3]<-sk

butstr<-c()
rez<-c()
int1<-c()
int2<-c()
x.dat3<-c()
y.dat3<-c()
index<-c()

for (j in 1:n)
{
  for(B in 1:200)
  {
    butstr<-sample(X[,3],replace=T)

    x.dat2<-c()
    y.dat2<-c()

    for (i in 1:n)
    {
      x.dat2[i]<-X[,1][butstr[i]] #sakarto datus atbilstoshi bootstrapam
      y.dat2[i]<-X[,2][butstr[i]]
    }

    x.dat3<-sort(x.dat2) #sakarto bootstrapotus x.datus augosha seciba

    index<-order(x.dat2)
    y.dat3<-y.dat2[index] #sakarto y.datus tapat ka bootstrapotus x.datus

    x3<-seq(min(x.dat3),max(x.dat3),by=(max(x.dat3)-min(x.dat3))/(n-1))
  }
}

```

```

h.xy3<-dpill(x.dat[i3],y.dat[i3])

rez[B]<-sqrt(RICE(a,x3[j],h.xy3,x.dat[i3],y.dat[i3]))
}

int1[j]<-sort(rez)[5]
int2[j]<-sort(rez)[195]
}

h.xy<-dpill(x.dat[i],y.dat[i])
t<-sqrt(RICE(a,x,h.xy,x.dat[i],y.dat[i]))

plot(x,t,type="l",ylim=c(0,0.2),xlab="X dati",ylab="sigma") #novertējums pēc Rice
lines(x,int1,lty=2,col="blue") #punktveida tīcamības intervals
lines(x,int2,lty=2,col="blue") #punktveida tīcamības intervals

#####
##Robezsadalījuma parbaude
#H0 dati ir normāli sadalīti
#H1 dati nav normāli sadalīti

N<-1000 #permutaciju skaits

x.fix<-0.5

#generesim sakotnejos datus

n<-100
eps<-0
dati2<-c()

for (t in 1:N)
{
  x.dat[i]<-x.dat[i1(n)]
  y.dat[i]<-y.dat[i1(n,eps,x.dat[i])

  h.xy<-optimize(CV,c(0.01,10))$minimum
  dati2[t]<-((n*h.xy)^(0.5))*(sqrt(RICE(a,x.fix,h.xy,x.dat[i],y.dat[i]))-sigma(x.fix))#simuletie dati
}

hist(dati2,prob=T)

library(nortest)
lillie.test(dati2)
shapiro.test(dati2)

#####

```

```

#Sigma_m,n zimejums pie dazadiem h, pie n=100
sigLRE_h<-function(h.xy) #dispersijas f-jas sigma(x) defineshana
{
  lines(xx,sqrt(RICE(a,xx,h.xy,x.dat,i,y.dat,i)),lty=c(i),type='l',col="blue")
  legend("topleft",
  c("0.05","0.1","0.15","0.2"),
  lwd=3, bty="n",lty=c(1,2,3,4),col="blue")
}

plot(x,sigma(x.dat,i),type='l',ylim=c(0,9),lwd=2,xlab="X dati",ylab="sigma_RICE") #ista dispersijas funkcija
#,ylim=c(0,10)
#sigLRE_h(0.2)

t<-c(0.05,0.1,0.15,0.2)

#t<-c(0.35,0.4,0.45,0.5)

for (i in 1:4)
{
  sigLRE_h(t[i])
}

#####
#ISEL kljuda, prieksh dispersijai exp(x)

ISEL<-function(signov,x.dat,i,disp)
{
  w<-0
  n<-length(x.dat)
  for (i in 1:n)
  {
    w<-w+( log ( (sqrt(signov[i]))/(disp[i]) ) )^2 )
  }
  (1/n)*w
}

#Rice<-RICE(a,x,h.opt,x.dat,i,y.dat)
#ISEL(Rice,x.dat,i,sigma(x.dat,i))

#####
#Videjas h un ISEL vertibas rekkini

n<-100
eps<-c()

a<-sqrt(2)

```

```

x<-seq(min(x.dat),max(x.dat),by=(max(x.dat)-min(x.dat))/(n-1))
length(x)

eps[1]<-0.1
eps[2]<-0.2
eps[3]<-0.3

ISEL_kl<-c(3)
h_opt<-c(3)

for (m in 1:3){

N<-500
t_ISEL<-0
h_ISEL<-0

for (v in 1:N){

x.dat<-x.dat1(n)
y.dat<-y.dat1(n,eps[m],x.dat)

h.xy<-optimize(CV,c(0.01,0.6))$minimum

#Rice funkcijas vertibas atrashana
Rice<-RICE(a,x,h.xy,x.dat,y.dat)

h_ISEL<-h_ISEL+h.xy
t_ISEL<-t_ISEL+ISEL(Rice,x.dat,sigma(x.dat))
}

ISEL_kl[m]<-t_ISEL/N
h_opt[m]<-h_ISEL/N
}

print(eps)
print(h_opt)
print(ISEL_kl)

#####
#dispersijas novertejums prieksh lidar datiem
dati<-read.table(file="lider.txt",header=T)

x.dat<-dati$range
y.dat<-dati$logratio
length(x.dat)

n<-length(x.dat)

```

```

xx<-seq(min(x.dat),max(x.dat),by=(max(x.dat)-min(x.dat))/(n-1))
length(x)

plot(x.dat,y.dat)

#####
##### dispersijas novertejums prieksh CMB datiem
dati<-read.table(file="CMB.txt",header=T)
##### dispersijas novertejums prieksh CMB datiem

x.dat<-read.table(file="CMB.txt",header=T)[,1]
y.dat<-read.table(file="CMB.txt",header=T)[,2]

n<-length(x.dat)

plot(x.dat,y.dat,cex=0.5,xlab="",ylab="") #izkliedes grafiks visiem datiem

#sigma ka funkcija no x pec Waserman Nonparametric gramatas
#ta ir taada funkcija ar kuru mees salidzinam parejos metodes, jebkuriem datiem
library(KernSmooth)

#Nadarya-Wats kodolu regresija, raksta ar roku

NW.reg<-function(x,x.dat,y.dat,h)
{
  sum(dnorm((x-x.dat)/h)*y.dat)/
  sum(dnorm((x-x.dat)/h))
}

NW.reg<-Vectorize(NW.reg,vectorize.args="x") #novertējam f-ju r(x)

sig2<-function(x,x.dat,y.dat)
{
  h.YX<-dpill(x.dat,y.dat)
  z.dat<-log((y.dat-NW.reg(x.dat,x.dat,y.dat,h.YX))^2)
  h.ZX<-dpill(x.dat,z.dat)
  exp(NW.reg(x,x.dat,z.dat,h.ZX))
}

xx<-seq(min(x.dat),max(x.dat),by=(max(x.dat)-min(x.dat))/(n-1))
length(xx)

sig2<-Vectorize(sig2,vectorize.args="x")

t1<-sqrt(sig2(xx,x.dat,y.dat))

plot(x,sigma(x),type="l",ylim=c(0,5))

```

```

lines(xx,t1,type="l",col="orange",xlab="X dati",ylab="sigma") #ista dispersijas f-ja lidar datiem

h.opt<-optimize(CV,c(0.01,40))$minimum
h.opt #optimala h vertiba ir 10.032 LIDAR, CMB ir 16.21

t2<-sqrt(RICE(a,xx,h.opt,x.dat,y.dat))
lines(xx,t2,type="l",col="yellow")

#####
#2.metode

a<-sqrt(2)
b<-1/2

Q<-qnorm(0.75,0,1) #75 % kvantile

I<-c() #vieniba vektors
for (i in 1:length(y.dat)-1)
I[i]<-1

O<-c() #nulles vektors
for (i in 1:length(y.dat)-1)
O[i]<-0

Hic<-function(yy,sig,i)
{
yy<-diff(y.dat)/a
{if (abs(yy[i]/sig)>Q)
I[i]
else
O[i]}
}

Hic2<-Vectorize(Hic,vectorize.args="i") #katrai i-tai vertibai bus sava f-jas Hic vertiba

h.xy<-dpill(x.dat,y.dat)

#Nadarya-Watsona kodolu novertejums
NW.svari<-function(x,x.dat,y.dat,h,i) #nem katro i-to
{
dnorm((x-x.dat[i])/h)/sum(dnorm((x-x.dat)/h))
}
NW.svari<-Vectorize(NW.svari,vectorize.args="x")
x<-seq(min(x.dat),max(x.dat),by=((max(x.dat))-(min(x.dat)))/(n-1))

#novertesim pashu f-ju sigma(x), ta ka sigma(x) defineta publikacija ar MT palidzibu

```

```

NW.svari2<-Vectorize(NW.svari,vectorize.args="i")
sigMSD<-function(x,x.dat,y.dat,h,sig)
{
n<-length(x.dat)
sum((NW.svari2(x,x.dat,y.dat,h,1:(n-1))*(Hic2(y.dat,sig,1:(n-1)))))-b
}
sig2<-Vectorize(sigMSD,vectorize.args="sig")

#atrodīsim visas saknes
nov.sig<-function(x)
{
uniroot(function(sig) sig2(x,x.dat,y.dat,h.xy,sig),c(0,1000000000))$root
}
sigMSD<-Vectorize(nov.sig)

#katram xi bus sava sigma

t3<-c()
t3<-sigMSD(x)

lines(x,t3,col="green") #uzzimejam novertetas disperijas grafiku

#####
#3.metode
x<-seq(min(x.dat),max(x.dat),by=((max(x.dat))-min(x.dat))/(n-1))

#novertesim pashu f-ju sigma(x), ta ka sigma(x) defineta publikacija ar BT f-ju
c<-0.70417
a<-sqrt(2)
b<-0.75

I<-c()
for (i in 1:length(y.dat)-1)
I[i]<-1

Hic<-function(y.dat,sig,i)
{
yy<-diff(y.dat)/a
{if (abs(yy[i]/sig)<=c)
3*(yy[i]/(sig*c))^2-3*(yy[i]/(sig*c))^4+(yy[i]/(sig*c))^6
else
I[i]}
}

Hic2<-Vectorize(Hic,vectorize.args="i") #katrai i-tai vertibai bus sava f-jas Hic vertiba

#Nadarya-Watsona kodolu novertejums
NW.svari<-function(x,x.dat,y.dat,h,i) #nem katro i-to

```

```

{
dnorm((x-x.dat[i])/h)/sum(dnorm((x-x.dat)/h))
}

NW.svari<-Vectorize(NW.svari,vectorize.args="x")

#novertesim pashu f-ju sigma(x), ta ka sigma(x) defineta publikacija ar BT palidzibu

NW.svari2<-Vectorize(NW.svari,vectorize.args="i")
sigMBT<-function(x,x.dat,y.dat,h,sig)
{
n<-length(x.dat)
sum((NW.svari2(x,x.dat,y.dat,h,1:(n-1))*(Hic2(y.dat,sig,1:(n-1)))))-b
}

sigMBT<-Vectorize(sigMBT,vectorize.args="sig")

#atrodisim visas saknes
nov.sig<-function(x)
{
uniroot(function(sig) sig2(x,x.dat,y.dat,h.xy,sig),c(0,1000000000))$root
}
sigMBT<-Vectorize(nov.sig)

#katram xi bus sava sigma

t4<-c()
t4<-sigMBT(x)

lines(x,t4,col="red") #uzzimejam novertetas disperijas grafiku

#####
#####4. metode
#####uzskata sigma par konstanti (rice metode)

# Nadaraya-Watsona kodolu regresija
NW.reg<-function(x,x.dat,y.dat,h)
{
sum(dnorm((x-x.dat)/h)*y.dat)/sum(dnorm((x-x.dat)/h))
}
NW.reg<-Vectorize(NW.reg,vectorize.arg="x")

library(KernSmooth)
h<-dpill(x.dat,y.dat)
h
h<-h.xy

#novertesim sigma kaa konstante ar rice
n<-length(y.dat)
sig1<-sum(diff(y.dat)^2)/(n-1) #atrodam dispersijas vertibu kaa konstante

```

```
t<-length(x)
sigCONST<-c()
for (i in 1:length(x))
sigCONST[i]<-sig1

t5<-sigCONST

lines(x,t5,col="blue",type="l")
```

darbs “Neparametriskās regresijas kļūdas dispersijas robusti novērtējumi” 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: Maksims Korotejevs

(paraksts)

(datums)

Rekomendēju darbu aizstāvēšanai.

Vadītājs: doc. Dr.math. Jānis Valeinis

(paraksts)

(datums)

Recenzents: doc. Dr.math. Nadežda Siņenko

(paraksts)

(datums)

Darbs iesniegts Matemātikas nodaļā _____

(datums)

(darbu pieņēma)

Darbs aizstāvēts valsts pārbaudījuma komisijas sēdē

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

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