LATVIJAS UNIVERSITĀTE

FIZIKAS UN MATEMĀTIKAS FAKULTĀTE

MATEMĀTIKAS NODAĻA

BUTSTRAPA METODES ANALĪZE ATKARĪGIEM DATIEM

MAĢISTRA DARBS

Autors: **Mārcis Bratka**

Stud. apl. MaSt020015

Darba vadītājs: Dr. Math. Jānis Valeinis

RĪGA 2009

Anotācija

Darbā ir aprakstīts butstrapa metodes pielietojums atkarīgiem datiem. Ir apskatīta bloku butstrapa metožu konstruēšana, parādīta un salīdzināta to darbības efektivitāte. Tāpat apskatīti īpaši gadījumi, kad novērtējamais parametrs ir izsakāms, kā gluda funkcija no izlases vidējās vērtības, kā kādu vienādojumu atrisinājums un kā diferencējams funkcionālis. Darba nobeigumā ir apskatīts butstrapa metodes pielietojums neparametriskās regresijas gadījumā, kad klasiskās metodes, kas balstītas uz neatkarīgiem novērojumiem, nedarbojas.

Atslēgas vārdi: datu pārkārtošana, bloku butstraps, atkarīgi dati, laikrindas, neparametriskā regresija.

Annotation

This paper is concerned with the application of bootstrap for dependent data. Block bootstrap methods are analyzed and compared to each other. Some general classes of estimators that may be represented as smooth function of sample means, as solution of equations and as differentiable functionals are reviewed. Finally, block bootstrap methods are introduced in the context of nonparametric regression where common methods are not effective or do not work.

Key words: resampling, block bootstrap, dependent data, time series, nonparametric regression.

Saturs

1. Ievads

Līdz ar mūsdienu datoru attīstību, arī statistikas nozarē nozīmīgu vietu ir ieņēmušas metodes, kas balstītas uz pastiprinātu datoru resursu izmantošanu. Viena no tādām ir butstrapa metode. Butstrapa metodi piedāvāja un noformulēja Efron (skatīt [1]). Kopš tās parādīšanās, butstrapa metode ir devusi iespēju risināt statistiskas problēmas ar vienkāršotiem modeļu pieņēmumiem. Butstrapa metodes būtība ir konstruēt attiecību starp populāciju un izlasi. Tā pieņem doto izlasi $X_1, \ldots, X_n$, kā labi reprezentējošu priekš populācijas, un veido butstrapa izlases $X_1^*, \ldots, X_n^*$, kā realizācijas no populācijas. Visbiežāk butstrapa izlases elementi tiek iegūti, kā gadījuma izlase ar atkārtojumiem no sākuma $X_1, \ldots, X_n$ ar vienādām varbūtībām (neparametriskais butstraps). Taču ir iespējams konstruēt parametrisku butstrapa metodi, kad butstrapa izlase tiek veidota pēc kāda zināma sadalījuma likuma. Interesējošais parametrs θ, piemēram, $\theta = EX_1$, tiek novērtēts ar butstrapa izlašu palīdzību. Patiesībā butstrapa metodes galvenā ideja ir apskatīt novērtētā θ sadalījuma likumu, nevis pašu parametra θ novērtējumu. Pie tam ir parādīts, ka butstrapa aproksimācija sadalījuma funkcijai ir precīzāka, kā klasikā Centrālās robežas teorēmas aproksimācija.

Tā kā iespējamo butstrapa izlašu skaits ir n^n, tad tieša butstrapa novērtējuma aprēķins ir sarežģīts, pat neiespējams. Šajā vietā liela loma tiek uzticēta datora aprēķiniem, lietojot Monte-Carlo simulācijas. Ar $X_1^{*(j)}, \ldots, X_n^{*(j)}$ apzīmēsim vienas butstrapa izlases realizāciju $j = 1, \ldots, B$, kur B ir butstrapa izlašu skaits. Tad pēc lielo skaitļu likuma, aprēķinot katras butstrapotās izlases interesējošo parametru, vidējais parametrs dod precīzu parametra novērtējumu.

Singh ([2]) parādīja piemēru, kad, augstāk minētā, butstrapa metode dod aplamus rezultātus atkarīgiem novērojumiem. Taču ir butstrapa metodes, kas dod rezultātus arī datiem ar atkarības struktūru. Šajā gadījumā butstrapa metodes mērķis ir nepazaudēt datu atkarības struktūru, pārkārtojot elementus. Vienas no vienkāršākajām un efektīvākajām ir bloku butstrapa metodes, ko piedāvājuši Hall ([3]), Carlstein ([4]), Kunsch ([5]), Liu un Singh ([6]) un citi. Bloku butstrapa metodes pārkārto blokus ar elementiem tā, lai bloki savā starpā būtu neatkarīgi. Bloku butstrapa metodes balstās uz nosacījumiem par vāji atkarīgiem novērojumiem $X_1, \ldots, X_n$, ko raksturo jaukto procesu koeficienti.

Viens no butstrapa pielietojumiem ir neparametriskās regresijas joslas platuma jeb gludināšanas parametra izvēle. Neparametriskās regresijas gludināšanas parametrs ir atkarīgs no regresijas atlikumu struktūras. Gadījumā, kad atlikumi ir atkarīgi, parastās metodes nav pielietojamas. Nelieli šo metožu uzlabojumi ar bloku butstrapu dod apmierinošus rezultātus.

Turpmākais darbs ir veltīts bloku butstrapa metodēm ar atkarīgiem datiem. 2. un 3. nodaļā tiks dotas pamatdefinīcijas un formulējumus, kā arī butstrapa metodes aprakstu neatkarīgiem un vienādi sadalītiem datiem. 4. nodaļā tiks parādīts piemērs, kad butstrapa metode dod aplamus rezultātus atkarīgiem datiem, un definētas metodes, kas paredzētas atkarīgiem datiem. Tālāk 5. nodaļā tiks apskatītas bloku butstrapa metodes īpašības izlases vidējai vērtībai un 6. nodaļā tiks vispārinātas bloka butstrapa metodes ar konkrētiem modeļiem. Praktiski tiks parādītas divas butstrapa pieejas un salīdzināti iegūtie rezultāti. 7. nodaļā teorētiski tiks salīdzinātas četras bloku butstrapa metodes un veikts salīdzinājums ar simulētiem datiem. Pēdējā nodaļa tiks veltīta butstrapa pielietojumam neparametriskajā regresijā. Secinājumu daļā apkoposim galvenās idejas un iespējamās tālākās darbības. Pielikumā ir dotas praktisko uzdevumu programmas.

Šī darba teorijas izklāsts tiks ņemts pamatā no [7], kura šobrīd ir viena no retajām grāmatām par butstrapa metodēm atkarīgiem datiem. Praktiskā daļa tiks veikta ar programmas R palīdzību.

2. Apzīmējumi un definīcijas

Šajā nodaļā dosim apzīmējumu un definīciju formulējumus, kuri tiks izmantoti darbā. Sākumā definēsim dažas virkņu īpašības.

Definīcija 1. Gadījumu vektoru virkne $\{X_n\}_{n\in\mathbf{Z}}$ ir stacionāra, ja katriem $i_1 < \ldots < i_k,\ k \in \mathbf{N}$ un $m \in \mathbf{Z}$ virkņu $\{X_{i_1}, \ldots, X_{i_k}\}$ un $\{X_{i_1+m}, \ldots, X_{i_k+m}\}$ sadalījumi ir vienādi.

Definīcija 2. Gadījumu vektoru virkne $\{X_n\}_{n\in\mathbf{Z}}$ ir m-atkarīga kādam $m \geq 0$, ja virknes $\{\ldots, X_k\}$ un $\{X_{k+m+1}, \ldots\}$ ir neatkarīgas visiem $k \in \mathbf{Z}$.

Tālāk dosim definīcijas dažām konverģencēm.

Definīcija 3. Pieņemsim, ka $X,\ \{X_n\}_{n\geq 1}$ ir gadījuma lielumi un $a \in \mathbf{R}$ ir konstante.

- Saka, ka $\{X_n\}_{n\geq 1}$ konverģē uz a pēc varbūtības, pieraksta $X_n \xrightarrow{p} a$, ja katram $\epsilon > 0$
$$\lim_{n\to\infty} P(|X_n - a| > \epsilon) = 0.$$
- Pieņemsim, ka $\{X_n\}_{n\geq 1}$ un X ir definēti vienā un tajā pašā varbūtību telpā $(\Omega, \mathcal{F}, P)$. Tad saka, ka $\{X_n\}_{n\geq 1}$ konverģē uz X pēc varbūtības, pieraksta $X_n \xrightarrow{p} X$, ja katram $\epsilon > 0$
$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0.$$

Definīcija 4. Pieņemsim, ka $\{X_n\}_{n\geq 1}$ un X ir definēti vienā un tajā pašā varbūtību telpā $(\Omega, \mathcal{F}, P)$. Tad saka, ka $\{X_n\}_{n\geq 1}$ konverģē uz X gandrīz droši (pret P), ja eksistē kopa $A \in \mathcal{F}$ tāda, ka $P(A) = 0$ un
$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \text{ visiem } x \in A^c.$$

Definīcija 5. Pieņemsim, ka $\{X_n\}_{n\geq 1}$ un X ir $\mathbf{R}^d$ vērtību gadījuma vektori. Saka, ka $\{X_n\}_{n\geq 1}$ konverģē uz X pēc sadalījuma, pieraksta $X_n \xrightarrow{d} X$, ja
$$\lim_{n\to\infty} P(X_n \in A) = P(X \in A)$$
katrai $A \in \mathcal{B}(\mathbf{R}^d)$ ar $P(X \in \partial A) = 0$, kur ∂A ir A robeža.

Apzīmēsim dažas virkņu attiecības.

Definīcija 6. Virknēm $\{a_n\}_{n\geq 1} \subset \mathbf{R},\ \{b_n\}_{n\geq 1} \subset (0, \infty)$ teiksim, ka
$$a_n = o(b_n),\ \text{kad } n \to \infty, \quad \text{ja} \quad a_n/b_n \to 0,\ \text{kad } n \to \infty,$$
$$a_n = O(b_n),\ \text{kad } n \to \infty, \quad \text{ja} \quad \lim_{n\to\infty} \sup |a_n|/b_n < \infty.$$

Līdzīgi gadījumu lielumu virknei $\{X_n\}_{n\geq 1}$ un virknei $\{b_n\}_{n\geq 1} \subset (0, \infty)$ teiksim, ka

$$X_n = o_p(b_n), \text{ kad } n \to \infty, \quad \text{ja} \quad X_n/b_n \xrightarrow{p} 0, \text{ kad } n \to \infty,$$
$$X_n = O_p(b_n), \text{ kad } n \to \infty, \quad \text{ja} \quad \{X_n/b_n\}_{n\geq 1} \text{ ir blīva virkne,}$$

tas ir, katram $\epsilon > 0$ eksistē $M \in (0, \infty)$ tāds, ka $\sup_{n\geq 1}(P(X_n/b_n) > M) < \epsilon$.

Pārējie nepieciešamie apzīmējumi tiks doti teorijas izklāsta laikā.

3. Butstrapa metode neatkarīgiem un vienādi sadalītiem datiem

Šajā nodaļā īsi aprakstīsim butstrapa metodi neatkarīgiem un vienādi sadalītiem datiem. Pieņemsim, ka $X_1, X_2, \ldots$ ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar sadalījumu F. $\mathcal{X}_n = (X_1, \ldots, X_n)$ ir dotā izlase un $T_n = t_n(\mathcal{X}_n, F)$ ir funkcionālis, kurš atkarīgs no datiem un sadalījuma F. Vienkāršākais piemērs ir $T_n = \sqrt{n}(\bar{X} - \mu)/\sigma$, kur $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$, $\mu = EX_1$ un $\sigma^2 = DX_1$. Statistikā ir svarīgi noteikt T_n sadalījumu vai kādu citu tā raksturlielumu, piemēram, standartnovirzi. Ar G_n apzīmēsim T_n sadalījumu, $G_n(x) \equiv P(T_n \leq x)$. Butstrapa metode ļauj risināt šāda veida problēmas bez nosacījumiem uz sadalījumu F. Šo metodi piedāvāja Efron ([1]).

Dotai izlasei $\mathcal{X}_n = (X_1, \ldots, X_n)$ veidosim jaunu izlasi $\mathcal{X}_m^* = (X_1^*, \ldots, X_m^*)$ ar atkārtojumiem, kur $m \leq n$ (parasti $m = n$). Tā kā $\mathcal{X}_n = (X_1, \ldots, X_n)$ ir neatkarīgi un vienādi sadalīti, tad $\mathcal{X}_m^* = (X_1^*, \ldots, X_m^*)$ arī ir neatkarīgi un vienādi sadalīti ar

$$P_*(X_1^* = X_j) = n^{-1},\ 1 \leq j \leq n,$$

kur P_* apzīmē varbūtību butstrapotai $\mathcal{X}_n$. Tādejādi, X_i^* ir dots ar sadalījumu

$$F_n = n^{-1}\sum_{i=1}^{n} \delta_{X_i},$$

kur δ_y apzīmē Diraka varbūtības mēru ar vērtību 1 punktā y un 0 citur.

Tālāk definēsim butstrapa versiju $T_{m,n}^*$ no T_n, aizvietojot $\mathcal{X}_n$ ar $\mathcal{X}_m^*$ un F ar F_n,

$$T_{m,n}^* = t_m(\mathcal{X}_m^*, F_n).$$

Apzīmēsim $T_{m,n}^*$ sadalījumu, dotam $\mathcal{X}_n$, ar $\hat{G}_{m,n}$, kas ir butstrapa metodes novērtējums nezināmam T_n sadalījumam G_n. Ja mēs interesēsimies par kādu T_n sadalījuma funkcionāli $\varphi(G_n)$, tad atbilstošais butstrapa novērtējums ir iegūstams, *ievietojot* G_n vietā $\hat{G}_{m,n}$, tas ir, $\varphi(G_n)$ butstrapa novērtējums ir $\varphi(\hat{G}_{m,n})$. Piemēram, ja $\varphi(G_n) = DT_n = \int x^2 dG_n(x) - (\int x dG_n(x))^2$, tad DT_n butstrapa novērtējums ir $\varphi(\hat{G}_{m,n}) = D(T_{m,n}^* | \mathcal{X}_n) = \int x^2 d\hat{G}_{m,n}(x) - (\int x d\hat{G}_{m,n}(x))^2$.

Parādīsim galvenās idejas statistikai $T_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$. $\mu = EX_1$ ir parametrs, kuru vēlamies aprakstīt. Pēc iepriekš minētās procedūras, definēsim butstrapa $T_{m,n}^*$ no butstrapa izlases ar apjomu m

$$T_{m,n}^* = \sqrt{m}(\bar{X}_m^* - E_* X_1^*)/(D_* X_1^*)^{1/2},$$

kur $\bar{X}_m^* = m^{-1}\sum_{i=1}^m X_i^*$ ir vidējā vērtība no butstrapa izlases $X_1^*, \ldots, X_m^*$, E_* un D_* ir nosacītā matemātiskā cerība un nosacītā dispersija butstrapa izlasēm, dotam $\mathcal{X}_n$. Katram $k \geq 1$

$$E_*(X_1^*)^k = \int x^k dF_n(x) = n^{-1}\sum_{i=1}^n X_i^k.$$

Tātad $E_* X_1^* = \bar{X}_n$, $D_* X_1^* \equiv s^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ un tādejādi

$$T_{m,n}^* = \sqrt{m}(\bar{X}_m^* - \bar{X}_n)/s_n.$$

Pie nosacījuma $m = n$ ir spēkā sekojoša teorēma.

Teorēma 1. *Ja $X_1, X_2, \ldots$ ir neatkarīgi un vienādi sadalīti, $EX_1^2 < \infty$, $\sigma^2 = DX_1 \in (0, \infty)$, tad*

$$\sup_n |P_*(T_{n,n}^* \leq x) - \Phi(x)| = o(1) \text{ gandrīz droši, kad } n \to \infty,$$

kur $\Phi(\cdot)$ ir $N(0,1)$ sadalījuma funkcija.

Teorēmas pierādījums atrodams [7].

Pēc Centrālrobežas teorēmas T_n konverģē uz sadalījumu $N(0,1)$ pēc sadalījuma. No tā seko, ka

$$\tilde{\Delta}_n \equiv \sup_x |P_*(T_{n,n}^* \leq x) - P(T_n \leq x)| = o(1) \text{ gandrīz droši, kad } n \to \infty,$$

tas ir, nosacītais $T_{n,n}^*$ sadalījums $\hat{G}_{n,n}$ dod labu aproksimāciju T_n sadalījumam G_n.

Pie papildus nosacījumiem Singh ([2]) parādīja, ka

$$\tilde{\Delta}_n = O(n^{-1}(\log\log n)^{1/2}) \text{ gandrīz droši, kad } n \to \infty.$$

Tātad butstrapa aproksimācija $P(T_n \leq \cdot)$ ir precīzāka kā klasiskā aproksimācija ar Centrālo robežteorēmu, kurai kļūda ir $O(n^{-1/2})$ (skatīt Berry-Esseen teorēmu no [7]).

4. Butstrapa metodes atkarīgiem datiem

4.1. Butstrapa neatbilstība atkarīgiem datiem

3. nodaļā aprakstītā butstrapa metode ir pietiekami vienkārša un viegli realizējama, taču tās pielietojums nav universāls. Singh ([2]) parādīja, ka butstrapa metodei ir trūkumi, lietojot to atkarīgiem datiem. Šajā nodaļā parādīsim rezultātus un piemērus, kad butstrapa metode dod aplamus rezultātus. Pieņemsim, ka $X_1, X_2, \ldots$ ir m-atkarīga gadījumu lielumu virkne ar $EX_1 = \mu$ un $EX_1^2 < \infty$. Atgādināsim, ka virkni sauc par m-atkarīgu, ja divas virknes $\{X_1, \ldots, X_k\}$ un $\{X_{k+m+1}, \ldots\}$ ir neatkarīgas kādam $m \geq 0$ un visiem $k \geq 1$.

Piemērs 1. Neatkarīgi un vienādi sadalīti gadījuma lielumi $\{\epsilon_n\}_{n\geq 1}$ ir 0-atkarīgi. Ja apskatām procesu $X_n = \epsilon_n + 0.5\epsilon_{n-1}$ ar neatkarīgiem un vienādi sadalītiem $\{\epsilon_n\}_{n\geq 1}$, tad virkne $\{X_n\}_{n\geq 1}$ ir 1-atkarīga.

Piemērs no ņemts no [7].

Apzīmēsim

$$\begin{aligned} \sigma_m^2 &= DX_1 + 2\sum_{i=1}^{m-1} \mathrm{Cov}(X_1, X_{1+i}), \\ \bar{X}_n &= n^{-1}\sum_{i=1}^{n} X_i. \end{aligned} \tag{1}$$

Teorēma 2 (Centrālrobežas teorēma m-atkarīgām virknēm)**.** *Pieņemsim, ka gadījumu lielumu virkne $\{X_i\}_{i\in\mathbf{Z}}$ ir stacionāra, m-atkarīga kādam $m \geq 0$. Ja $EX_1^2 < \infty$ un $\sigma_m^2 > 0$, tad*

$$n^{-1/2}\sum_{i=1}^{n}(X_i - EX_1) \xrightarrow{d} N(0, \sigma_m^2). \tag{2}$$

Teorēmas pierādījums atrodams [7].

Novērtēsim $T_n = \sqrt{n}(\bar{X}_n - \mu)$, lietojot butstrapa procedūru neatkarīgiem un vienādi sadalītiem datiem. Pieņemsim, ka $m = n$. T_n butstrapa versija ir

$$T_{n,n}^* = \sqrt{n}(\bar{X}_n^* - \bar{X}_n),$$

kur $\bar{X}_n^* = n^{-1}\sum_{i=1}^{n} X_i^*$.

Teorēma 3. *Pieņemsim, ka gadījumu lielumu virkne $\{X_i\}_{i\geq 1}$ ir stacionāra, m-atkarīga kādam $m \geq 0$ ar $EX_1 = \mu$ un $\sigma^2 = DX_1 \in (0, \infty)$, tad*

$$\sup_x P_*(T_{n,n}^* \leq x) - \Phi(x/\sigma) = o(1) \textit{ gandrīz droši, kad } n \to \infty.$$

Teorēmas pierādījums atrodams [7].

Sekas 1. *Ja ir spēkā Teorēmas 3 nosacījumi un* $\sum_{i=1}^{m} \mathrm{Cov}(X_1, X_{1+i}) \neq 0$ *un* $\sigma_\infty^2 \neq 0$*, tad katram* $x \neq 0$

$$\lim_{n\to\infty} (P_*(T_{n,n}^* \leq x) - P(T_n \leq x)) = \Phi(x/\sigma) - \Phi(x/\sigma_\infty) \neq 0 \textit{ gandrīz droši.}$$

Pierādījums seko no Teorēmas 3 un (2).

Sekas 2. $T_{n,n}^*$ *sadalījums tiecas uz normālo sadalījumu, bet ar nepareizu dispersiju. Visiem* $x \neq 0$*, sadalījuma* $P(T_n \leq x)$ *butstrapa novērtējuma* $P_*(T_{n,n}^* \leq x)$ *vidējā kvadrātiskā kļūda netiecas uz 0.*

Tātad butstrapa metode atkarīgiem datiem nav piemērota, tā dod aplamus rezultātus, jo ignorē $\mathcal{X}_n$ atkarības struktūru.

Piemērs 2. Parādīsim šos teorētiskos rezultātus ar datu simulāciju palīdzību. Apskatīsim ARMA$(1,1)$ modeli

$$X_i = 0.2X_{i-1} + \epsilon_i - 0.4\epsilon_{i-1}, \tag{3}$$

kur ϵ_i ir neatkarīgi un vienādi sadalīti ar $N(0,1)$ sadalījumu. Aprēķināsim teorētisko σ_m^2, kur kovariācijas ņemsim formā

$$\begin{aligned} \gamma(0) &= \sigma_\epsilon^2(1+\theta\phi+\theta^2)/(1-\phi^2), \\ \gamma(h) &= \sigma_\epsilon^2(1+\theta\phi)(\phi+\theta)\phi^{h-1}/(1-\phi^2), \end{aligned} \tag{4}$$

kur $h \geq 1$ un σ_ϵ^2 ir ϵ_i dispersija (rezultāti no [8]). Mūsu gadījumā $\theta = -0.4$, $\phi = 0.2$ un $\sigma_\epsilon^2 = 1$. σ_m^2 aprēķinam atmetīsim kovariācijas, kuras ir mazākas par 0.001. Mūsu modelim (3) pēc (1), (4) teorētiskā $\sigma_m^2 = 0.645$. Pēc (2) ir sagaidāms, ka statistika $T_{n,n}^* = \sqrt{n}(\bar{X}_n^* - \mu)$ tiecas uz $N(0, 0.645)$. Simulēsim datus ar apjomu $n = 50, 100, 200, 500, 1000, 2000, 5000$, izmantojot programmas R procedūru `arima.sim`. Katrai no izlasēm pielietosim butstrapa metodi neatkarīgiem un vienādi sadalītiem datiem. Veidosim otru butstrapa procedūru, kurā esošo izlasi sadalīsim blokos, katrā pa 10 elementiem. Šos blokus nejaušā secībā izvēlēsimies ar atkārtojumiem $n/10$ reizes. Jauno bloku elementi būs viena butstrapa izlase. Abām metodēm veidosim 10000 butstrapa izlases. Par centrējošo vērtību μ izvēlēsimies sākotnējās izlases vidējo vērtību $\bar{X}_n$. Aprēķināsim butstrapa izlašu vidējo vērtību un dispersiju. Butstrapa rezultāti parādīti 1. tabulā.

1. tabula: Statistikas $T^*_{n,n}$ butstrapotie rezultāti ARMA(1, 1) procesam (3). 1. metodē ir butstrapoti izlases elementi, 2. metodē ir butstrapoti elementi, sadalīti blokos.

	1.metode		2.metode	
Izlases apjoms	Vidējā vērtība	Dispersija	Vidējā vērtība	Dispersija
50	-0.003	1.334	-0.003	0.544
100	-0.016	0.966	0.042	0.358
200	0.024	0.913	0.014	0.389
500	-0.019	1.087	0.003	0.625
1 000	-0.033	1.035	0.040	0.637
2 000	0.016	0.972	0.004	0.672
5 000	-0.028	0.965	0.013	0.679

Kā redzams, tad 1. metodes rezultāti vidējai vērtībai ir *tuvi* 0, kas arī sagaidāms, taču dispersijas novērtējumi ir *tālu* no sagaidāmās vērtības 0.645. To arī teorētiski sagaidām no Sekām 1 un 2. Butstraps, kas lietots neatkarīgiem un vienādi sadalītiem datiem, nav piemērots atkarīgu datu analīzei. Apskatot 2. metodi, secinām, ka tās vidējās vērtības ir tuvas 0 un dispersijas novērtējumi pie lielākiem n ir tuvi teorētiskajam rezultātam. Šī ir viena no bloku butstrapa metodēm, kuras tiks apskatītas nākamajā nodaļā. Vēlāk parādīsim, ka pie dažādiem izlašu apjomiem n, ir dažādi *optimālie* bloku garumi l, pie kuriem šīs metode dod precīzākos rezultātus, kas izskaidro rezultātu svārstības arī 2. metodei.

Šī uzdevuma programmas R kods ir atrodams pielikumā.

4.2. Slīdošo bloku butstraps

Iepriekšējā nodaļā apskatītā butstrapa metode ir lietojama gadījumos, kad apstiprināta hipotēze par datu neatkarību. Gadījumos, kad pētniekam nav pietiekamas informācijas par datu struktūru, metodes, kas balstītas uz neatkarīgu un vienādi sadalītu datu butstrapu metodēm, nav piemērotas. Kunsch ([5]) un Liu un Singh ([6]) piedāvāja jaunu metodi-*slīdošo bloku butstrapu* (no angļu valodas ,,moving block bootstrap", turpmāk tekstā lietosim saīsinājumu MBB), kura ir lietojama atkarīgiem datiem bez parametriska modeļa pieņēmumiem. Pretēji iepriekš aprakstītajai procedūrai, kad tika pārkārtoti izlases elementi, MBB metode pārkārto izlases elementus, sakārtotus blokos. Rezultātā, bloka iekšienē datu atkarība ir saglabāta. Bloka garums ir tieši atkarīgs no izlases apjoma, pieaugot izlases apjomam, bloka garums pieaug.

Pieņemsim, ka $X_1, X_2, \ldots$ ir stacionāra gadījuma lielumu virkne un $\mathcal{X}_n = (X_1, \ldots, X_n)$

ir dotā izlase. Definēsim MBB novērtējumus formā $\hat{\theta}_n = T(F_n)$, kur F_n apzīmē empīrisko sadalījumu funkciju dotiem $X_1, \ldots, X_n$ un $T(\cdot)$ ir funkcionālis no F_n. $l \equiv l_n \in [1, n]$ ir vesels skaitlis. $\mathcal{B}_i = (X_i, \ldots, X_{i+l-1})$ ir bloks garumā l sākot ar elementu X_i, $1 \leq i \leq N$, kur $N = n - l + 1$. Lai iegūtu MBB izlasi, izvēlamies vajadzīgo bloku skaitu no $\mathcal{B}_1, \ldots, \mathcal{B}_N$ un pārkārtojam gadījuma izlasē $\mathcal{B}_1^*, \ldots, \mathcal{B}_k^*$, kur katrs no izvēlētajiem blokiem satur l elementus. Apzīmēsim $\mathcal{B}_i^*$ elementus ar $X_{(i-1)l+1}^*, \ldots, X_{il}^*$, $i = 1, \ldots, k$. Tādejādi $X_1^*, \ldots, X_m^*$ sastāda MBB izlasi ar apjomu $m \equiv kl$. MBB versija novērtējumam $\hat{\theta}_n$ ir

$$\theta_{m,n}^* = T(F_{m,n}^*),$$

kur $F_{m,n}^*$ ir empīriskā sadalījuma funkcija $X_1^*, \ldots, X_m^*$.

Dosim ekvivalentu formulējumu MBB metodei. Atzīmēsim, ka izvēlēties blokus $\mathcal{B}_i^*$ no $\mathcal{B}_1, \ldots, \mathcal{B}_N$ ir tas pats, kā izvēlēties k indeksus no kopas $1, \ldots, N$. Pieņemsim, ka $I_1, \ldots, I_k$ ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar diskrēto vienmērīgo sadalījumu no $1, \ldots, N$. Ja apzīmējam $\mathcal{B}_i = \mathcal{B}_{I_i}$, kur $i = 1, \ldots, k$, tad $\mathcal{B}_1^*, \ldots, \mathcal{B}_k^*$ raksturo gadījuma izlasi no $\mathcal{B}_1, \ldots, \mathcal{B}_N$. Butstrapa izlase $X_1^*, \ldots, X_m^*$ no $\mathcal{B}_1^*, \ldots, \mathcal{B}_k^*$ tiek definēta, kā iepriekš. Piezīmēsim, ka izlases bloki $(X_1^*, \ldots, X_l^*), \ldots, (X_{(k-1)l+1}^*, \ldots, X_{kl}^*)$ ir neatkarīgi un vienādi sadalīti l-dimensionāli gadījuma vektori ar

$$P_*((X_1^*, \ldots, X_l^*) = (X_j, \ldots, X_{j+l-1})) = P_*(I_1 = j) = N^{-1}, \tag{5}$$

kur $1 \leq j \leq N$ un P_* apzīmē varbūtību, dotai $\mathcal{X}_n$. Gadījumā, kad $l = 1$ jeb bloks sastāv no viena elementa, MBB reducējas uz butstrapa metodi neatkarīgiem un vienādi sadalītiem gadījuma lielumiem, kas apskatīta 3. nodaļā.

Līdzīgi, kā butstrapa metodei ar neatkarīgiem un vienādi sadalītiem gadījuma lielumiem, MBB izlases apjoms parasti ir izvēlēts ar tādu pašu kārtu, kā sākotnējās izlases apjoms. Ja b ir mazākais veselais skaitlis, kurš apmierina $bl \geq n$, tad iespējams izvēlēties b, kā MBB izlases apjomu, un lietot tikai pirmos n elementus, lai noteiktu butstrapa T_n. Tomēr dažkārt sastopami gadījumi, kad mazāka izlase dod labākus rezultātus (skatīt [7]).

Lai gan novērtējumi formā $\hat{\theta}_n = T(F_n)$ ietver bieži lietotus novērtējumus, piemēram, izlases vidējā vērtība, tomēr tie nav pietiekami plaši, lai būtu lietojami laikrindu analīzē, jo $\hat{\theta}_n$ ir atkarīgs tikai no viendimensionāla empīriskā sadalījuma F_n. Tas neiekļauj bieži lietotas statistikas, piemēram, izlases korelācijas koeficientus. Definēsim vispārinātu MBB, kas ļauj apskatīt šāda veida problēmas.

Dotiem novērojumiem $\mathcal{X}_n$, definēsim p-dimensionālu empīrisko mēru

$$F_{p,n} = (n-p+1)^{-1} \sum_{j=1}^{n-p+1} \delta_{Y_j},$$

kur $Y_j = (X_j, \ldots, X_{j+p-1})$ un katram $y \in \mathbf{R}^p$, δ_y nosaka varbūtības mēru uz $\mathbf{R}^p$ ar vērtību 1 punktā y. Vispārinātais MBB veids ietver novērtējumus formā

$$\hat{\theta}_n = T(F_{p,n}), \tag{6}$$

kur $T(\cdot)$ ir funkcionālis, kurš definēts visu varbūtību mēru $\mathbf{R}^p$ kopas apakškopā.

Piemērs 3. Izlases kovariācijas novērtējums ar kārtu $k \geq 0$ ir

$$\hat{\gamma}_n(k) = (n-k)^{-1} \sum_{j=1}^{n-k} (X_{j+k} - \bar{X}_{n,k})(X_j - \bar{X}_{n,k}),$$

kur $\bar{X}_{n,k} = (n-k)^{-1} \sum_{j=1}^{n-k} X_j$. Novērtējums $\hat{\gamma}_n(k)$ ir formā (6) ar $p = k+1$.

Piemērs no [7].

Lai definētu MBB jauno $\hat{\theta}_n$ versiju, fiksēsim bloku garumu l, $1 \leq l \leq n-p+1$ un definēsim blokus attiecībā pret Y_i, kā

$$\tilde{\mathcal{B}}_j = (Y_j, \ldots, Y_{j+l-1}),\ 1 \leq j \leq n-p-l+2.$$

Izvēlēsimies nejauši $k \geq 1$ blokus no $\{\tilde{\mathcal{B}}_j : 1 \leq i \leq n-p-l+2\}$, lai veidotu MBB izlasi $Y_1^*, \ldots, Y_m^*,\ m = kl$. MBB (6) versija ir definēta

$$\theta_{m,n}^* = T(\tilde{F}_{m,n}^*), \tag{7}$$

kur $\tilde{F}_{m,n}^* \equiv m^{-1} \sum_{j=1}^{m} \delta_{Y_j^*}$ ir $Y_1^*, \ldots, Y_m^*$ empīriskā sadalījuma funkcija. Tātad novērtējumiem formā (6) MBB versija ir definēta, kā Y vērtību bloku pārkārtošana, nevis X vērtību. Definētais (7) attiecas uz gadījumiem, kad p ir fiksēts, gan kad p ar n tiecas uz bezgalību.

Tā kā novērtējumu $\hat{\theta}_n$ vienmēr var izteikt kā funkciju no dotiem $X_1, \ldots, X_n$, tad ir iespējams dot alternatīvu definīciju $\hat{\theta}_n$, pārkārtojot tieši $X_1, \ldots, X_n$. Pieņemsim, ka bloka butstrapa izlase $X_1^*, \ldots, X_m^*$ ir iegūta, pārkārtojot blokus $\mathcal{B}_i = (X_i, \ldots, X_{i+l-1})$, $i = 1, \ldots, N$. Definēsim butstrapa p-dimensionālu lielumu $Y_i \equiv (X_i, \ldots, X_{i+p-1})$, kā $Y_i^{**} \equiv (X_i^*, \ldots, X_{i+p-1}^*)$, $i = 1, \ldots, m-p+1$ no $X_1^*, \ldots, X_m^*$. Tad butstrapa $\hat{\theta}_n$ alternatīva versija ir

$$\theta_{m,n}^{**} = T(\tilde{F}_{m,n}^{**}), \tag{8}$$

kur $\tilde{F}_{m,n}^{**} \equiv (m-p+1)^{-1} \sum_{j=1}^{m-p+1} \delta_{Y_j^{**}}$. Šo pieeju MBB versijai $\hat{\theta}_n$ sauksim par ,,naivo" pieeju, bet sākotnējo pieeju (7) sauksim par ,,vienkāršo" pieeju.

Lai salīdzinātu ,,naivo" un ,,vienkāršo" pieeju, pieņemsim, ka $\{X_n\}_{n\geq 1}$ ir stacionāra gadījumu lielumu virkne. Katram i, gadījuma lieluma vektors $Y_i = (X_i, \ldots, X_{i+p-1})$ ir ar tādu pašu sadalījumu, kā $X_1, \ldots, X_p$. Pārkārtotie vektori Y_i^* ar „vienkāršo" pieeju vienmēr saglabā $X_1, \ldots, X_p$ atkarības struktūru. Kad lietojam „naivo" pieeju, butstrapa izlases elementi X_i^*, kuri viens no otra ir mazāk kā p vērtību attālumā un ir tuvu pie robežas diviem blokiem $\mathcal{B}_j^*$ un $\mathcal{B}_{j+1}^*$, ir *neatkarīgi*. Tāpēc Y_i^{**} komponentes, iegūtas ar „naivo" pieeju, nesaglabā $X_1, \ldots, X_p$ atkarības struktūru. Rezultātā „naivā" pieeja dod papildus novirzi $\hat{\theta}_n$ butstrapa versijai $\theta_{m,n}^{**}$. Abu pieeju praktiskais pielietojums tiks parādīts 6.1. nodaļā.

MBB rezultāti ir būtiski atkarīgi no bloku garuma l izvēles. Lai saglabātu $\{X_n\}_{n\geq 1}$ atkarības struktūru, tad l jātiecas uz bezgalību, kad izlases apjoms n tiecas uz bezgalību. Bieži l izvēle ir formā $l = Cn^\delta$, kur C ir konstante un $\delta \in (0, 1/2)$. 7.3.2. nodaļā teorētiski tiks parādīti bloku garumi, pie kuriem MBB metode dod labākos rezultātus.

4.3. Nešķeļošo bloku butstraps

Šajā nodaļā apskatīsim jaunu bloku butstrapa veidu. Carlstein ([4]) piedāvāja *nešķeļošo bloku butstrapa* metodi (no angļu valodas „nonoverlapping block butstrap" (NBB)). Parādīsim rezultātus situācijā, kad $p = 1$.

Galvenā NBB ideja ir lietot blokus, kuri nešķeļas. Pieņemsim, ka $l \equiv l_n \in [1, n]$ ir vesels skaitlis un $b \geq 1$ ir lielākais veselais skaitlis, kurš apmierina $lb \leq n$. Definēsim blokus

$$\mathcal{B}_n^{(2)} = (X_{(i-1)l+1}, \ldots, X_{il}), \ i = 1, \ldots, b,$$

kur apzīmējumu „(2)" lietosim saistībā ar NBB metodi, līdzīgi, turpmāk lietosim indeksu 1, apzīmējot MBB metodi, un 3, 4 lietosim vēlāk apskatītajām metodēm. Piezīmēsim, ka, ja MBB izvēlētie bloki šķeļas, tad NBB jeb $\mathcal{B}_n^{(2)}$ nešķeļas. Rezultātā, iespējamais atšķirīgo bloku skaits ir mazāks, nekā MBB metodei.

Tālākā procedūra ir identiska MBB metodei. Veidojam gadījuma izlasi ar atkārtojumiem $\mathcal{B}_1^{*(2)}, \ldots, \mathcal{B}_k^{*(2)}$ no $\mathcal{B}_1^{(2)}, \ldots, \mathcal{B}_b^{(2)}$, kur $k \geq 1$. Apzīmēsim butstrapa izlases $X_{2,1}^*, \ldots, X_{2,m}^*$, kura iegūta no blokiem $\mathcal{B}_1^{*(2)}, \ldots, \mathcal{B}_k^{*(2)}$ un $m = kl$, empīrisko sadalījumu ar $F_{m,n}^{*(2)}$. Butstrapa versija novērtējumam $\hat{\theta}_n = T(F_n)$ ir

$$\hat{\theta}_{m,n}^{*(2)} = T(F_{m,n}^{*(2)}). \tag{9}$$

Lai gan butstrapa novērtējumu definīcijas MBB un NBB ir *līdzīgas*, tomēr $\theta_{m,n}^*$ un $\theta_{m,n}^{*(2)}$ sadalījumu īpašības ir dažādas. Parādīsim vienkāršāko gadījumu, kad $\hat{\theta}_n = n^{-1}\sum_{j=1}^n X_j$

ir izlases vidējā vērtība. $\hat{\theta}_n$ butstrapa versijas abām metodēm ir attiecīgi

$$\theta^*_{m,n} = m^{-1}\sum_{j=1}^{m} X^*_j \text{ un } \theta^{*(2)}_{m,n} = m^{-1}\sum_{j=1}^{m} X^*_{2,j}.$$

No (5) iegūsim, ka

$$\begin{aligned} E_*\theta^*_{m,n} &= E_*\left(l^{-1}\sum_{i=1}^{l} X^*_i\right) \\ &= N^{-1}\sum_{j=1}^{N}\left(l^{-1}\sum_{i=1}^{l} X_{j+i-1}\right) \\ &= N^{-1}\left(n\bar{X}_n - l^{-1}\sum_{j=1}^{l-1}(l-j)(X_j + X_{n-j+1})\right). \end{aligned} \tag{10}$$

Lai iegūtu $E_*\theta^{*(2)}_{m,n}$, ievērosim, ka NBB metodei $(X^*_{2,1},\ldots,X^*_{2,l}),\ldots,(X^*_{2,(m-l+1)},\ldots,X^*_{2,m})$ ir neatkarīgi un vienādi sadalīti ar

$$P_*((X^*_{2,1},\ldots,X^*_{2,l}) = (X_{(j-1)l+1},\ldots,X_{jl})) = b^{-1},$$

kur $j = 1,\ldots,b$. No tā seko, ka

$$\begin{aligned} E_*\theta^{*(2)}_{m,n} &= E_*\left(l^{-1}\sum_{i=1}^{l} X^*_{2,i}\right) \\ &= b^{-1}\sum_{j=1}^{b}\left(l^{-1}\sum_{i=1}^{l} X_{(j-1)l+i}\right) \\ &= (bl)^{-1}\left(n\bar{X}_n - \sum_{i=bl+1}^{n} X_i\right), \end{aligned} \tag{11}$$

kas ir vienāds ar $\bar{X}_n$, ja n dalās ar l bez atlikuma. Tātad butstrapa novērtējumiem ir dažādas vidējās vērtības. Tomēr var parādīt, ka, ja process $\{X_n\}_{n\geq 1}$ apmierina dažus momentu un jauktos nosacījumus, tad $E(E_*\theta^*_{m,n} - E_*\theta^{*(2)}_{m,n}) = O(l/n^2)$. Tas nozīmē, ka pie lieliem izlašu apjomiem, atšķirība ir nenozīmīga.

4.4. Vispārinātais bloku butstraps

Nodefinēsim vispārinātā bloku butstrapa ideju un metodi, ar kura palīdzību var vispārināti nodefinēt bloku butstrapus, ieskaitot MBB un NBB. Dotiem $\mathcal{X}_n = (X_1,\ldots,X_n)$, definēsim periodisku attiecību $Y_{n,i},\ i \geq 1$. Piezīmēsim, ka katram $i \geq 1$ eksistē veseli skaitļi $k_i \geq 0$ un $j_i \in [1,n]$ tādi, ka $i = k_i n + j_i$. Tad $i = j_i (mod\ n)$. Definēsim elementus $Y_{n,i},\ i \geq 1$ ar attiecību $Y_{n,i} = X_{j_i}$. Tas ir tāpat, kā novietot $X_1,\ldots,X_n$ periodiski uz bezgalīgas taisnes un katru elementu apzīmēt ar $Y_{n,i}$. Definēsim blokus

$\mathcal{B}(i,j) = (Y_{n,i}, \ldots, Y_{n,i+j-1})$, kur $i \geq 1,\ j \geq 1$. Ar Γ_n apzīmēsim pārejas varbūtības funkciju kopā $\mathbf{R}^n \times \bigotimes_{t=1}^{\infty}(\{1,\ldots,n\} \times \mathbf{N})$, tas ir, katram $x \in \mathbf{R}^n$, Γ_n ir varbūtības mērs uz $\bigotimes_{t=1}^{\infty}(\{1,\ldots,n\} \times \mathbf{N}) \equiv \{\{i_t, l_t\}_{t=1}^{\infty} : 1 \leq i_t \leq n,\ 1 \leq l_t < \infty$ katram $t \geq 1\}$, un katrai kopai $A \subset \bigotimes_{t=1}^{\infty}(\{1,\ldots,n\} \times \mathbf{N})$, Γ_n ir Boreļa mērojama funkcija no $\mathbf{R}^n$ uz $[0,1]$. Tad vispārinātais bloku butstraps pārkārto blokus no $\{\mathcal{B}(i,j) : i \geq 1,\ j \geq 1\}$ pēc pārejas varbūtības funkcijas Γ_n sekojoši. Pieņemsim, ka $(I_1, J_1), (I_2, J_2), \ldots$ ir gadījumu vektoru virkne ar nosacīto kopējo sadalījumu $\Gamma_n(\mathcal{X}_n, \cdot)$, dotam $\mathcal{X}_n$. Bloki, kuri izvēlēti ar vispārināto bloku butstrapa metodi, ir $\mathcal{B}(I_1, J_1), \mathcal{B}(I_2, J_2), \ldots$, kuri var nebūt neatkarīgi. Ar $X^*_{G,1}, X^*_{G,2}, \ldots$ apzīmēsim pārkārtoto bloku elementus. Butstrapa versija novērtējumam $\hat{\theta}_n = T(F_n)$ ar vispārināto bloku butstrapa metodi ir $\theta^{*(G)}_{m,n} = T(F^{*(G)}_{m,n})$ ar atbilstošu $m \geq 1$ izvēli, kur $F^{*(G)}_{m,n}$ ir empīriskais sadalījums $X^*_{G,1}, \ldots, X^*_{G,m}$.

Gandrīz visas literatūrā sastopamās bloku butstrapa metodes var tikt izteiktas, kā speciāls gadījums no vispārinātā bloku butstrapa. Piemēram, MBB ar bloku garumu l pārejas varbūtības funkcija Γ_n ir

$$\Gamma_n(x, \cdot) = \bigotimes_{i=1}^{\infty} \left((N^{-1} \sum_{j=1}^{N} \delta_j) \times \delta_l \right),\ x \in \mathbf{R}^n,$$

kur $N = n - l + 1$ un δ_y ir varbūtību mērs ar vērtību 1 punktā y. Šajā gadījumā $\Gamma_n(x, \cdot)$ nav atkarīgs no $x \in \mathbf{R}^n$ un gadījuma vektori $(I_1, J_1), (I_2, J_2), \ldots$ ir *nosacīti neatkarīgi un vienādi sadalīti* gadījuma vektori ar nosacīto sadalījumu

$$P_*(I_1 = j, J_1 = k) = \begin{cases} N^{-1}, & 1 \leq j \leq N \text{ un } k = l; \\ 0, & \text{citādi.} \end{cases}$$

Rezultātā, pārkārtotie bloki $\mathcal{B}(I_1, J_1), \mathcal{B}(I_2, J_2), \ldots$ ir no $\{\mathcal{B}(i,j) : 1 \leq i \leq N,\ j = l\}$, kas vienāds ar $\mathcal{B}_1, \ldots, \mathcal{B}_N$, kas definēts 4.2. nodaļā. Līdzīgi varam definēt arī NBB metodi ar vispārināto bloku butstrapu.

Pēc MBB metodes definīcijas seko, ka MBB ir problēmas ar elementiem, kuri atrodas izlases sākumā un beigās, tātad pie robežām. MBB sākuma un beigu izlases elementiem piešķir mazāku varbūtību tikt izvēlētiem butstrapa izlasē, nekā elementiem, kuri atrodas izlases vidū. $X_j,\ l \leq j \leq n-l$ elements tieši l reizes ir izlasē $\mathcal{B}_1, \ldots, \mathcal{B}_N$, savukārt elementi X_j un $X_{n-j+1},\ 1 \leq j \leq l-1$ parādās tikai j reizes. Tā kā parasti mēs nevaram palielināt izlasi pirms X_1 un pēc X_n elementa, tad nav iespējams definēt jaunus blokus, lai izvairītos no šī robežu efekta. Līdzīga problēma eksistē NBB metodei. NBB neiekļauj butstrapa izlasē pēdējos elementus, ja izlases apjoms n, dalīts ar bloku garumu l, nav vesels skaitlis.

Politis un Romano ([9]) piedāvāja vienkāršu risinājumu robežu problēmām. Viņu ideja bija izveidot ,,riņķi", uz kura ir novietoti izlases elementi, un veidot blokus pa šo ,,riņķi". Viņi attīstīja 2 jaunas bloku pārkārtošanas metodes, balstītas uz riņķveida blokiem: *riņķveida bloku butstraps* un *stacionārais butstraps.*

4.4.1. Riņķveida bloku butstraps

Riņķveida bloku butstrapa (no angļu valodas ,,circular block bootstrap" (CBB)) metode pārkārto šķeļošus un periodiski sakārtotus blokus ar dotu garumu l no $\mathcal{B}(i,l),\ldots,\mathcal{B}(n,l)$. Pārejas funkcija Γ_n CBB ir

$$\Gamma_n(x,\cdot) = \bigotimes_{i=1}^{\infty}\left((n^{-1}\sum_{j=1}^{n}\delta_j)\times\delta_l\right),\ x\in\mathbf{R}^n. \qquad (12)$$

Apzīmēsim CBB pārkārtoto bloku indeksus ar $I_{3,1}, I_{3,2},\ldots$. Tad pēc (12) seko, ka $I_{3,1}, I_{3,2},\ldots$ ir *nosacīti neatkarīgi un vienādi sadalīti* ar $P_*(I_{3,1}=i)=n^{-1}$ un $P_*(J_i=l)=1$ visiem $i=1,\ldots,n$. Tā kā katrs no X_i ir iekļauts $\mathcal{B}(1,l),\ldots,\mathcal{B}(n,l)$ tieši l reizes un CBB metode pārkārto visus blokus ar vienādu varbūtību, tad visi elementi no sākotnējās izlases $X_1,\ldots,X_n$ ir ar vienādu svaru. Šī īpašība atšķir CBB no MBB un NBB metodēm, kurām ir problēmas pie izlases robežām. Šī CBB īpašība izriet arī no sekojoša rezultāta. Ar $X^*_{3,1},\ldots,X^*_{3,m}$ apzīmēsim CBB elementus, iegūtus no pārkārtotajiem blokiem $\{\mathcal{B}(I_{3,i},l): i\geq 1\}$, un ar $\bar{X}^{*(3)}_m$ apzīmēsim CBB izlases vidējo vērtību, iegūtu no m butstrapa novērojumiem, kur $m=kl$, kādam $k\geq 1$. Tad pēc (12) seko, ka

$$\begin{aligned}
E_*\theta^{*(3)}_m &= E_*\left(m^{-1}\sum_{i=1}^{m}X^*_{3,i}\right)\\
&= l^{-1}E_*\left(\sum_{i=1}^{l}X^*_{3,i}\right)\\
&= l^{-1}\left(n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{l}Y_{n,j+i-1}\right)\\
&= l^{-1}(l\bar{X}_n)\\
&= \bar{X}_n.
\end{aligned}$$

Tātad CBB metodes butstrapa izlases sagaidāmā vidējā vērtība ir vienāda ar izlases $\mathcal{X}_n$ vidējo vērtību. Šī īpašība nav spēkā MBB un NBB gadījumos. Politis un Romano ([9]) atzīmēja, ka CBB gadījumā ir vieglāk definēt butstrapa versiju formā $T_n = t_n(\bar{X}_n,\mu)$, kur $\mu = EX_1$. CBB gadījumā, $T^*_{m,n} = t_m(\bar{X}^{*(3)}_m,\bar{X}_n)$ ir atbilstoša T_n butstrapa versija. Tomēr, aizvietojot parametru μ ar $\bar{X}_n$, lai definētu butstrapa versiju T_n MBB un NBB metodēm, tiks izveidota papildus novirze.

4.4.2. Stacionārais butstraps

Stacionārais (bloku) butstraps (no angļu valodas ,,stationary bootstrap" (SB)), atšķirībā no iepriekš apskatītajām MBB, NBB un CBB metodēm, neizmanto fiksētu bloka garumu l, bet nosaka to ar gadījuma lielumu palīdzību. $p \equiv p_n \in (0,1)$ tāds, ka $p \to 0$ un $np \to \infty$, kad $n \to \infty$. Tad SB pārkārto blokus $\mathcal{B}(I_{4,1}, J_{4,1}), \mathcal{B}(I_{4,2}, J_{4,2}), \ldots$, kur indeksu vektori $(I_{4,1}, J_{4,1}), (I_{4,2}, J_{4,2}), \ldots$ ir *nosacīti neatkarīgi un vienādi sadalīti*, kur $I_{4,1}$ ir ar diskrētu vienmērīgo sadalījumu kopā $1, \ldots, n$ un $J_{4,1}$ ir ar ģeometrisko sadalījumu ν_n ar parametru p, tas ir,

$$P_*(J_{4,1} = j) \equiv \nu_n(j) = p(1-p)^{j-1}, \ j = 1, 2, \ldots,$$

turklāt $I_{4,1}$ un $J_{4,1}$ ir neatkarīgi. Tātad SB atbilst vispārinātā bloku butstrapa nosacījumiem ar pārejas funkciju $\Gamma_n(\cdot, \cdot)$ tādu, ka

$$\Gamma_n(x, \cdot) = \bigotimes_{i=1}^{\infty} \left((n^{-1} \sum_{j=1}^{n} \delta_j) \times \nu_n \right), \ x \in \mathbf{R}^n.$$

Piezīmēsim, ka $\Gamma_n(x, \cdot)$ nav atkarīgs no $x \in \mathbf{R}^n$.

SB metodi var formulēt savādāk. Pieņemsim, ka $X^*_{4,1}, X^*_{4,2}, \ldots$ ir SB metodes izlases elementi, sakārtoti virknē un iegūti no pārkārtotiem blokiem $\mathcal{B}(I_{4,1}, J_{4,1}), \mathcal{B}(I_{4,2}, J_{4,1}), \ldots$. Virkne $\{X^*_{4,i}\}_{i \in \mathbf{N}}$ ir iegūstama no sekojošas procedūras. Izvēlēsimies $X^*_{4,1}$ elementu nejauši no $X_1, \ldots, X_n$, tas ir, $X^*_{4,1} = Y_{n,I_{4,1}}$, kur $I_{4,1}$ ir definēts, kā iepriekš. Lai iegūtu nākamo elementu $X^*_{4,2}$, mēs veidosim bināru procedūru, kur ,,veiksmīgs" iznākums ir ar varbūtību p. Ja šīs procedūras rezultāts ir ,,veiksmīgs", tad izvēlēsimies $X^*_{4,2}$ atkal nejauši no $X_1, \ldots, X_n$. Ja pretēji, tad izvēlēsimies $X^*_{4,2} = Y_{n,I_{4,1}+1}$, nākamo elementu pēc $X^*_{4,1} \equiv Y_{n,I_{4,1}}$ no periodiskas virknes $\{Y_{n,i}\}_{i \geq 1}$. Vispārējā gadījumā, ja $X^*_{4,i}$ ir izvēlēts un ir vienāds ar kādu Y_{n,i_0} kādam $i_0 \geq 1$, tad nākamais SB izlases elements $X^*_{4,i+1}$ tiek izvēlēts, kā Y_{n,i_0+1} ar varbūtību $1-p$, bet ar varbūtību p $X^*_{4,i+1}$ tiek nejauši izvēlēts no $\mathcal{X}_n$.

Parādīsim, ka abi formulējumi ir ekvivalenti. Ar W_i apzīmēsim binārās procedūras rezultātu, lai izvēlētos elementu $X^*_{4,i}, \ i \geq 2$. Tad attiecībā uz $\mathcal{X}_n$, $W_i, \ i \geq 2$ ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar $P_*(W_i = 1) = p = 1 - P_*(W_i = 0)$, un virkne $\{W_i : i \geq 2\}$ nav atkarīga no virknes $\{I_{4,i} : i \geq 1\}$. Tālāk definēsim mainīgo $M_j, \ j \geq 0$ tādu, ka $M_0 \equiv 1$ un $M_j = \inf\{i \geq M_{j-1} + 1 : W_i = 1\}, \ j \geq 1$. Tātad $M_j, \ j \in \mathbf{N}$ nosaka kārtas numuru virknē $\{W_i : i \geq 2\}$, kur rezultāts ir bijis ,,veiksmīgs". Pie tam $M_j, \ j \in \mathbf{N}$ ir ar negatīvo binomiālo sadalījumu ar parametriem j un p. Tad atbilstošie SB metodes izlases elementi X^*_{4,M_j} ir nejauši izvēlēti no $X_1, \ldots, X_n$, kā $X^*_{4,M_j} = Y_{n,I_{4,j+1}}, \ j \geq 1$. No otras puses, katram i starp $M_{j-1}+1$ un M_j-1, binārās procedūras rezultāts ir bijis „neveiksmīgs" un atbilstošie SB metodes izlases elementi ir izvēlēti, ņemot $M_j - M_{j-1} - 1$ elementu no virknes $\{Y_{n,i}\}_{i\in\mathbf{N}}$ pēc $Y_{n,I_{4,j}}$. Binārie iznākumi $\{W_i : i = M_{j-1}, \ldots, M_j - 1\}$ dod SB bloka elementus $(X^*_{4,M_{j-1}}, \ldots, X^*_{4,M_j-1}) = (Y_{n,I_{4,j}}, \ldots, Y_{n,I_{4,j}+M_j-M_{j-1}-1})$, $j \geq 1$. Tad, definējot $J_{4,j} = M_j - M_{j-1}$, $j \geq 1$ un lietojot negatīvā binomiālā sadalījuma īpašības, varam secināt, ka $J_{4,1}, J_{4,2}, \ldots$ ir neatkarīgi un vienādi sadalīti ar ģeometrisko sadalījumu ar parametru p. Esam parādījuši, ka abi SB metodes formulējumi ir ekvivalenti.

Svarīga SB metodes īpašība ir tā, ka butstrapa izlases elementi $\{X^*_{4,i}\}_{i\in\mathbf{R}}$ ir *stacionāri*, no kā radies metodes nosaukums. Pierādīsim šo īpašību, izmantojot otro SB metodes formulējumu. Pieņemsim, ka $\{Z_i\}_{i\in\mathbf{N}}$ ir Markova ķēde ar stāvokļiem $\{1, \ldots, n\}$ tāda, ka sākuma stāvoklis ir ar varbūtībām $\pi \equiv (n^{-1}, \ldots, n^{-1})'$ un $\{Z_i\}_{i\in\mathbf{N}}$ pārejas varbūtību matrica ir $Q \equiv ((q_{ij}))$, kur

$$q_{ij} = \begin{cases} p + n^{-1}(1-p), & 1 \leq i < n,\ j = i+1; \\ n^{-1}(1-p), & 1 \leq i < n,\ j \neq i+1; \\ n^{-1}(1-p), & i = n,\ 2 \leq j \leq n; \\ p + n^{-1}(1-p), & i = n,\ j = 1. \end{cases} \tag{13}$$

Tad Z_i pieņem vērtības $1, \ldots, n$, katru ar varbūtību n^{-1}. Pie tam katram $k \geq 1$ un dotam $Z_k = i$, $1 \leq i \leq n$, nākamais Z_{k+1} pieņem vērtību $i+1(mod\ n)$ ar varbūtību $p+n^{-1}(1-p)$ un pieņem katru no palikušajām $n-1$ vērtībām ar varbūtību $n^{-1}(1-p)$. Tad pēc otrā SB metodes formulējuma izriet, ka SB elementi $\{X^*_{4,i}\}_{i\in\mathbf{N}}$ var tikt veidoti no $\{Z_i\}_{i\in\mathbf{N}}$ kā

$$X^*_{4,i} = X_{Z_i},\ i \geq 1. \tag{14}$$

Lai parādītu, ka $\{X^*_{4,i}\}_{i\in\mathbf{N}}$ ir stacionāra, atzīmēsim, ka pēc definīcijas matrica Q ir divkārt stohastiska un tā apmierina $\pi' Q = \pi'$. Tādēļ π ir *stacionārs* $\{Z_i\}_{i\in\mathbf{N}}$ sadalījums un $\{Z_i\}_{i\in\mathbf{N}}$ ir stacionāra Markova ķēde. Esam pierādījuši sekojošu teorēmu. Gadījumu vektoriem $\{X_i : i \in I\}$ no varbūtību telpas $(\Omega, \mathcal{F}, P)$ ar $\sigma\langle\{X_i : i \in I\}\rangle$ apzīmēsim σ-algebru no $\mathcal{F}$, ģenerētu ar $\{X_i : i \in I\}$. Gadījuma vektoram X un σ-algebrai $\mathcal{G}$ ar $\mathcal{L}(X)$ un $\mathcal{L}(X|\mathcal{G})$ apzīmēsim X varbūtību sadalījumu un nosacīto X varbūtību sadalījumu ar dotu $\mathcal{G}$.

Teorēma 4. *Pieņemsim, ka $\mathcal{F}_{in}$ ir σ-algebra, dota ar Z_i, un $\mathcal{X}_n$, $i \geq 1$. Tad $\{X^*_{4,i}, \mathcal{F}_{in}\}_{i\in\mathbf{N}}$*

ir stacionāra Markova ķēde katram $n \geq 1$, tas ir,

$$\mathcal{L}(X^*_{4,i}|\mathcal{X}_n) = \mathcal{L}(X^*_{4,1}|\mathcal{X}_n), \text{ katram } i \geq 1$$

un

$$\mathcal{L}(X^*_{4,i+1}|Z_i, \ldots, Z_i, \mathcal{X}_n) = \mathcal{L}(X^*_{4,i+1}|Z_i, \mathcal{X}_n), \text{ katram } i \geq 1.$$

Tālāk no (13) un (14) seko, ka, dotam pārkārtotās izlases apjomam m, nosacītā sagaidāmā SB metodes izlases vidējā vērtība $\bar{X}^{*(4)}_m \equiv m^{-1} \sum_{i=1}^m X^*_{4,i}$ ir

$$E_* \bar{X}^{*(4)}_m = E_* X^*_{4,1} = \bar{X}_n.$$

5. Bloku butstrapa metodes izlases vidējai vērtībai

Šajā sadaļā parādīsim bloku butstrapa īpašības izlases vidējai vērtībai. Sākumā ieviesīsim dažus atkarības mērus, kuri noderēs turpmākam teorijas izklāstam.

Pieņemsim, ka $(\Omega, \mathcal{F}, P)$ ir varbūtību telpa un $\mathcal{A}$ un $\mathcal{B}$ ir σ-algebras no $\mathcal{F}$. Ja $\mathcal{A}$ un $\mathcal{B}$ ir neatkarīgi, tad visiem $A \in \mathcal{A}$ un $B \in \mathcal{B}$ eksistē sakarība $\Delta_1 \equiv P(A \cap B) - P(A)P(B) = 0$ un $\Delta_2 \equiv P(A|B) - P(B) = 0$, pieņemot, ka $P(A) \neq 0$. Ja $\mathcal{A}$ un $\mathcal{B}$ nav neatkarīgi, tad varam novērtēt atkarības mēru starp $\mathcal{A}$ un $\mathcal{B}$, kā maksimālās vērtības no Δ_1 un Δ_2 vai arī lietot līdzīgus mērus. Apskatīsim atkarības jauktos koeficientus, kurus ieviesa Rosenblatt ([10]).

- α-jauktais koeficients:
$$\alpha(\mathcal{A}, \mathcal{B}) = \{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{A},\ B \in \mathcal{B}\}; \tag{15}$$
- ϕ-jauktais koeficients:
$$\phi(\mathcal{A}, \mathcal{B}) = \left\{ \left| \frac{P(A \cap B)}{P(A)} - P(B) \right| : A \in \mathcal{A},\ P(A) \neq 0,\ B \in \mathcal{B} \right\}; \tag{16}$$
- ψ-jauktais koeficients:
$$\psi(\mathcal{A}, \mathcal{B}) = \left\{ \left| \frac{P(A \cap B)}{P(A)P(B)} - 1 \right| : A \in \mathcal{A},\ P(A) \neq 0,\ B \in \mathcal{B},\ P(B) \neq 0 \right\}; \tag{17}$$
- ρ-jauktais koeficients:
$$\rho(\mathcal{A}, \mathcal{B}) = \left\{ \frac{|\mathrm{Cov}(X, Y)|}{\sqrt{DX}\sqrt{DY}} : X \in L^2(\mathcal{A}),\ Y \in L^2(\mathcal{B}) \right\}, \tag{18}$$
kur $\mathrm{Cov}(X, Y) = EXY - EXEY$, $DX = \mathrm{Cov}(X, X)$ un $L^2(\mathcal{A}) = \{X : X$ ir gadījuma lielums telpā $(\Omega, \mathcal{A}, P)$ ar $EX^2 < \infty\}$.

Vispārīgā gadījumā eksistē attiecība.

Teorēma 5.

$$\begin{gathered} 4\alpha(\mathcal{A}, \mathcal{B}) \leq \rho(\mathcal{A}, \mathcal{B}) \leq 2\phi^{1/2}(\mathcal{A}, \mathcal{B})\phi^{1/2}(\mathcal{B}, \mathcal{A}) \\ \textit{un} \\ \rho(\mathcal{A}, \mathcal{B}) \leq \psi(\mathcal{A}, \mathcal{B}). \end{gathered} \tag{19}$$

Pierādījumu skatīt [11].

Indeksu kopai $I \subset \mathbf{Z}$, $I \neq \emptyset$ laikrindas $\{X_i\}_{i \in I}$ pie kārtas $m \geq 1$ jauktie koeficienti ir definēti, ņemot koeficientu maksimālās vērtības pār σ-algebrām $\mathcal{A} = \sigma\langle\{X_i : i \leq k,\ i \in I\}\rangle$ un $\mathcal{B} = \sigma\langle\{X_i : i \geq k + m,\ i \in I\}\rangle$ katram $k \in I$. α-jauktajos un ρ-jauktajos gadījumos ir sekojoša definīcija.

Definīcija 7. Pieņemsim, ka $\{X_n\}_{n\in\mathbf{N}}$ ir gadījuma lielumu virkne telpā $(\Omega, \mathcal{A}, P)$ un $\mathcal{F}_a^b = \sigma\langle\{X_i : a \le i \le b\}\rangle,\ 1 \le a \le b \le \infty$.

- α-jauktais koeficients virknei $\{X_i\}_{i\ge1}$ ir definēts
$$\alpha(m) = \sup\left\{\alpha(\mathcal{F}_1^{k+1}, \mathcal{F}_{k+m+1}^{\infty}) : k \in \mathbf{N}\right\},\ m \ge 1, \tag{20}$$
kur $\alpha(\cdot,\cdot)$ ir definēts kā (15). Procesu $\{X_i\}_{i\ge1}$ sauc par α-jauktu, ja $\alpha(m) \to 0$, kad $m \to \infty$;

- ρ-jauktais koeficients virknei $\{X_i\}_{i\ge1}$ ir definēts
$$\rho(m) = \sup\left\{\rho(\mathcal{F}_1^{k+1}, \mathcal{F}_{k+m+1}^{\infty}) : k \in \mathbf{N}\right\},\ m \ge 1, \tag{21}$$
kur $\rho(\cdot,\cdot)$ ir definēts kā (18). Procesu $\{X_i\}_{i\ge1}$ sauc par ρ-jaukto procesu, ja $\rho(m) \to 0$, kad $m \to \infty$.

Abpusēji bezgalīgai laikrindai $\{X\}_{i\in\mathbf{Z}}$ koeficienti $\alpha(m)$ un $\rho(m)$, $m \ge 1$ ir definēti, (20) un (21) aizvietojot σ-algebru $\mathcal{F}_1^{k+1}$ ar $\mathcal{F}_{-\infty}^{k+1} \equiv \sigma\langle\{X_i : i \le k\}\rangle$ un $\mathbf{N}$ aizvietojot ar $\mathbf{Z}$. $\{X_i\}_{i\in\mathbf{Z}}$ ϕ-jauktos un ψ-jauktos koeficientus definē līdzīgi. Katrs no četriem jauktajiem koeficientiem saka, ka procesa $\{X_i\}_{i\in\mathbf{Z}}$ atkarība krītas, ja attālums m starp segmentiem $\{X_i : i \le k\}$ un $\{X_i : i \ge k+m+1\}$ palielinās. Piebildīsim, ka, ja process ir m_0-atkarīgs, kādam $m_0 \ge 0$, tad visi jauktie koeficienti ir vienādi ar 0 visiem $m > m_0$. Tātad virknes m-atkarība ir stingrāks nosacījums, nekā tikko raksturotie atkarības mēri, pie tam no (19) seko, ka α-jauktais ir visvājākais. Sīkāk par jauktajiem procesiem skatīt [11].

Gadījuma lielumam W un $p \in [1,\infty]$ definēsim p-normu
$$\|W\|_p = \begin{cases} (E|W|^p)^{1/p}, & p \in [1,\infty); \\ \inf\{x : P(|W| > x) = 0\}, & p = \infty. \end{cases}$$

Apgalvojums 1. *Pieņemsim, ka X un Y ir gadījuma lielumi no varbūtību telpas $(\Omega, \mathcal{F}, P)$.*

- *Ja $P(|X| \le a_1) = 1$, $P(|Y| \le a_2) = 1$ kādiem $a_1, a_2 \in (0,\infty)$, tad*
$$|\mathrm{Cov}(X,Y)| \le 4a_1a_2\alpha(\sigma\langle\{X\}\rangle, \sigma\langle\{Y\}\rangle);$$

- *pieņemsim, ka $p, q, r \in (1,\infty)$ ir reāli skaitļi, kuri apmierina $p^{-1} + q^{-1} + r^{-1} = 1$. Tad*
$$|\mathrm{Cov}(X,Y)| \le 8(\alpha(\sigma\langle\{X\}\rangle, \sigma\langle\{Y\}\rangle))^{1/r}\|X\|_p\|Y\|_q;$$

- $$|\mathrm{Cov}(X,Y)| \le \rho(\sigma\langle\{X\}\rangle, \sigma\langle\{Y\}\rangle)\|X\|_2\|Y\|_2.$$

Apgalvojuma pierādījums atrodams [7].

5.1. MBB, NBB, CBB metožu konsistence

Pieņemsim, ka $\{X_i\}_{i\in\mathbf{Z}}$ ir virkne no gadījuma vektoriem, kuri pieņem vērtības no $\mathbf{R}^n$. Ar T_n apzīmēsim centrētu un normētu izlases vidējo vērtību

$$T_n = \sqrt{n}(\bar{X}_n - \mu),$$

kur $\mu = EX_1$ un $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$. Šajā nodaļā parādīsim MBB, NBB un CBB T_n asimptotiskas kovariāciju matricas

$$DT_n \equiv ET_nT_n',$$

un T_n sadalījuma

$$G_n(x) \equiv P(T_n \le x),\ x \in \mathbf{R}^n$$

īpašības. Vienkāršības dēļ, pieņemsim, ka katrā no metodēm $b \equiv \lfloor n/l \rfloor$ bloki ir pārkārtoti un no tā seko, ka pārkārtotās izlases apjoms ir $n_1 = bl$. Apzīmēsim ar $\bar{X}_n^{*(1)}$, $\bar{X}_n^{*(2)}$ un $\bar{X}_n^{*(3)}$, attiecīgi MBB, NBB un CBB metožu, izlases vidējās vērtības. Tad butstrapa T_n versija ir

$$T_n^{*(j)} \equiv \sqrt{n_1}(\bar{X}_n^{*(j)} - E_*\bar{X}_n^{*(j)}),\ j = 1, 2, 3.$$

Butstrapa DT_n novērtējums ir $D_*T_n^{*(j)}$, $j = 1, 2, 3$, kur E_* un D_* apzīmē sagaidāmo nosacīto matemātisko cerību un nosacīto dispersiju, dotam $\mathcal{X}_n$. Līdzīgi, butstrapa $G_n(\cdot)$ novērtējums ir uzdots ar $T_n^{*(j)}$ nosacīto sadalījumu pie dota $\mathcal{X}_n$, $j = 1, 2, 3$.

5.1.1. Dispersijas novērtējums

Butstrapa dispersijas novērtējumam $D_*T_n^{*(j)}$, $j = 1, 2, 3$ piemīt īpašība tāda, ka to var izteikt vienkāršā formā, bez Monte-Carlo simulāciju palīdzības. Šī īpašība ir spēkā, pateicoties butstrapa izlases vidējās vērtības linearitātei pārkārtotajos elementos. Ar $U_i = (X_i + \ldots + X_{i+l-1})/l$ apzīmēsim bloka $X_i, \ldots, X_{i+l-1}$, $i \ge 1$ vidējo vērtību, ar $U_i^{(2)} \equiv U_{(i-1)l+1} = (X_{(i-1)l+1} + \ldots + X_{il})/l$, $i \ge 1$ apzīmēsim nešķeļošo bloku vidējo vērtību un ar $U_i^{(3)} = (Y_{n,i} + \ldots + Y_{n,i+l-1})/l$, $i \ge 1$ apzīmēsim periodiskas virknes $\{Y_{n,i}\}_{i\ge 1}$ bloku vidējo vērtību. Lietojot pārkārtoto bloku neatkarības īpašību, iegūsim

$$\begin{aligned}
D_*T_n^{*(1)} &= l\left(N^{-1}\sum_{i=1}^{N} U_iU_i' - \hat{\mu}_n\hat{\mu}_n'\right), \\
D_*T_n^{*(2)} &= l\left(b^{-1}\sum_{i=1}^{b} U_i^{(2)}U_i^{(2)\prime} - \hat{\mu}_{n,2}\hat{\mu}_{n,2}'\right) \qquad (22) \\
&\text{un} \\
D_*T_n^{*(3)} &= l\left(n^{-1}\sum_{i=1}^{n} U_i^{(3)}U_i^{(3)\prime} - \bar{X}_n\bar{X}_n'\right),
\end{aligned}$$

kur $N = n - l + 1$, $\hat{\mu}_n = N^{-1}\sum_{i=1}^{N} U_i$ un $\hat{\mu}_{n,2} = b^{-1}\sum_{i=1}^{b} U_i^{(2)}$. Šeit $\hat{\mu}_n$ ir labots [7] rezultāts, kur bija neprecizitāte.

Ja process $\{X_i\}_{i\in\mathbf{Z}}$ apmierina konkrētus momentu un α-jauktos nosacījumus tādus, kā Teorēmā 6 (skatīt zemāk), tad T_n asimptotiskā kovariāciju matrica ir uzdota ar bezgalīgu summu

$$\Sigma_\infty \equiv \lim_{n\to\infty} DT_n = \sum_{i=-\infty}^{\infty} EZ_1 Z_{1+i}',$$

kur $Z_i = X_i - \mu$, $i \in \mathbf{Z}$. Tādejādi butstrapa novērtējumu $D_* T_n^{*(j)}$, $j = 1, 2, 3$ var apskatīt, kā populācijas parametra Σ_∞ novērtējumu. Sekojošais rezultāts pierāda gan parametra DT_n, gan Σ_∞ butstrapa novērtējuma konsistenci.

Teorēma 6. *Pieņemsim, ka eksistē tāds $\delta > 0$, ka $E\|X_1\|^{2+\delta} < \infty$ un ka $\sum_{n=1}^{\infty} \alpha(n)^{\delta/(2+\delta)} < \infty$. Pie tam, ja $l^{-1} + n^{-n} l = o(1)$, kad $n \to \infty$, tad katram $j = 1, 2, 3$*

$$D_* T_n^{*(j)} \xrightarrow{p} \Sigma_\infty, \text{ kad } n \to \infty. \tag{23}$$

Teorēmas pierādījums atrodams [7]. Teorēmas nosacījumos ir labota grāmatā [7] pieļautā neprecizitāte.

Teorēma rāda, ka pie samērā vājiem momentu un α-jauktajiem nosacījumiem, procesam $\{X_i\}_{i\geq\mathbf{Z}}$ butstrapa dispersijas novērtējums $D_* T_n^{*(j)}$, $j = 1, 2, 3$ ir konsistents plašām bloku garuma l izvēlēm tādām, kur l tiecas uz bezgalību kopā ar n, bet ar kārtu lēnāk, kā n. Tātad bloku garumu izvēles $l = \log\log n$ vai $l = n^{1-\epsilon}$, $0 < \epsilon < 1$ ir atbilstošas, lai novērtējums $D_* T_n^{*(j)}$, $j = 1, 2, 3$ būtu konsistents. Vēlāk apskatīsim, ka *optimālais* bloka garums, kurš asimptotiski minimizē vidējo kvadrātisko kļūdu, ir formā $l = C_j n^{1/3}(1 + o(1))$, kad $n \to \infty$, kur $C_j > 0$ ir atbilstoša konstante, kura atkarīga no populācijas parametriem.

5.1.2. Sadalījuma funkcijas novērtējums

Šajā nodaļā parādīsim butstrapa metožu novērtējumus T_n sadalījuma funkcijai G_n. Ar $\hat{G}_n^{(j)}$, $j = 1, 2, 3$ apzīmēsim $T_n^{*(j)}$ nosacīto sadalījumu. Mēs varam interpretēt $\{G_n\}_{n\geq 1}$, kā punktu virkni visu varbūtības mēru, uzdotu uz $\mathbf{R}^n$, telpā. Ja G_n vāji konverģē uz robežsadalījumu G_∞, tad klasiskā pieeja ir G_n aproksimēt ar G_∞. Savukārt butstrapa metodes aproksimē G_n, ģenerējot *gadījuma* varbūtību mērus $\hat{G}_n^{(j)}$, kas atkarīgi no n. Sekojošā teorēma parāda butstrapa metodes konsistenci gadījuma mēriem $\hat{G}_n^{(j)}$, kad T_n sadalījums ir asimptotiski normālais. Lai formulētu rezultātu, teiksim, ka $x = (x_1, \ldots, x_d)$, $y = (y_1, \ldots, y_d) \in \mathbf{R}$, $x \leq y$, ja $x_i \leq y_i$ katram $i = 1, \ldots, d$.

Teorēma 7. *Pieņemsim, ka eksistē $\delta > 0$ tādi, ka $E\|X_1\|^{2+\delta} < \infty$ un $\sum_{n=1}^{\infty} \alpha(n)^{\delta(2+\delta)} < \infty$. Pie tam pieņemsim, ka $\Sigma_\infty = \sum_{i\in\mathbf{Z}} \mathrm{Cov}(X_1, X_{1+i})$ ir nesingulāra matrica un ka $l^{-1} + n^{-1}l = o(1)$, kad $n \to \infty$. Tad katram $j = 1, 2, 3$*

$$\sup_{x\in\mathbf{R}^d} \left| P_*(T_n^{*(j)} \le x) - P(T_n \le x) \right| \xrightarrow{p} 0, \text{ kad } n \to \infty.$$

Pierādījums atrodams [7].

Teorēma 7 rāda, ka tāpat, kā dispersijas novērtējumam, sadalījuma funkcijas novērtējums $\hat{G}_n^{(j)}$ ir konsistents pie plašas bloku garuma l izvēles. Nosacījumi uz l Teorēmās 6 un 7 ir nepieciešami butstrapa novērtējumu konsistencei. Ja parametrs l ir ierobežots, tad bloku butstrapa metodes nav korektas, jo pazaudē sākotnējās izlases atkarības struktūru. Tādēļ butstrapa novērtējumi konverģē uz *nepareizu* sadalījumu. To parādījām 4.1. nodaļā. No otras puses, ja l tiecas uz bezgalību ar tādu pašu kārtu, kā n, tas ir, neievēro nosacījumu $n^{-1}l = o(1)$, kad $n \to \infty$, tad ir nepietiekams bloku skaits, lai, pārkārtojot tos, tie *labi* reprezentētu populāciju. Var parādīt, ka šajā gadījumā novērtējumi $\hat{G}_n^{(j)}$ konverģē uz kādiem zināmiem *gadījuma* varbūtības mēriem. [7] atrodami formulējumi daudz plašākai novērtējumu klasei, kā Teorēmā 7.

5.2. SB metodes konsistence

Šajā nodaļā parādīsim SB metodes īpašības izlases vidējās vērtības dispersijas un sadalījuma funkcijas novērtējumam. Tāpat, kā iepriekš, $\{X_i\}_{i\in\mathbf{Z}}$ ir stacionāra gadījumu vektoru no $\mathbf{R}^d$ virkne ar μ. Katram $n \ge 1$ $\{Y_{n,i}\}_{i\ge 1}$ ir periodiska laikrinda, definēta $Y_{n,i} = X_j$, ja $i = j$ $(mod\ n)$. Atgādināsim, ka SB metode pārkārto blokus ar garumu l, kas ir gadījuma skaitlis. Kā parādījām 4.4.2. nodaļā, $E^*\bar{X}_n^{*(4)} = \bar{X}_n$ un statistikas $T_n = \sqrt{n}(\bar{X}_n - \mu)$ SB metodes variants ir $T_n^{*(4)} = \sqrt{n}(\bar{X}_n^{*(4)} - \bar{X}_n)$.

5.2.1. Dispersijas novērtējums

Centrētai un normētai izlases vidējai vērtībai ir iespējams bloku butstrapa dispersijas novērtējumu izteikt bez datu pārkārtošanas.

Apgalvojums 2. $\hat{\Gamma}_n(k) = n^{-1}\sum_{i=1}^{n-k} X_i X_{i+k}' - \bar{X}_n\bar{X}_n'$, $0 \le k < n$, $q = 1 - p$, $q_{n0} = 1/2$ *un* $q_{nk} = (1 - n^{-1}k)q^k + (n^{-1}k)q^{n-k}$, $1 \le k < n$. *Ja* $0 < p < 1$, *tad*

$$D_*T_n^{*(4)} = \sum_{k=0}^{n-1} q_{nk}(\hat{\Gamma}_n(k) + \hat{\Gamma}_n(k)'). \tag{24}$$

Tālāk parādīsim izlases vidējās vērtības dispersijas novērtējumu, saistībā ar populācijas parametru.

Teorēma 8. *Pieņemsim, ka* $E\|X_1\|^{4+\delta} < \infty$ *un* $\sum_{n=1}^{\infty} n^3\alpha(n)^{\delta/(4+\delta)} < \infty$ *kādam* $\delta > 0$, *turklāt* $p + (n^{1/2}p)^{-1} \to \infty$, *kad* $n \to \infty$. *Tad*

$$D_* T_n^{*(4)} \xrightarrow{p} \Sigma_\infty, \ \textit{kad } n \to \infty. \tag{25}$$

Gan Apgalvojuma 2, gan Teorēmas 8 pierādījums ir atrodams [7].

5.2.2. Sadalījuma funkcijas novērtējums

Tālāk parādīsim SB metodes izlases vidējās vērtības sadalījuma funkcijas konsistenci. Apskatīsim variantu, kad visi butstrapotās izlases elementi no pēdējā bloka tiek saglabāti, nevis atmesti pēc n. $K \equiv \inf\{i \geq 1 : L_1 + \ldots + L_i \geq n\}$ ir minimālais bloku skaits, lai ģenerētu SB izlasi apjomā n, un $N_1 = L_1 + \ldots + L_K$ ir kopējais butstrapa izlases elementu skaits pirmajos K blokos. Definēsim SB versiju centrētai un normētai izlases vidējai vērtībai T_n ar pārkārtotās izlases apjomu N_1

$$\tilde{T}_n^{*(4)} = N_1^{-1/2} \sum_{i=1}^{N_1} (X_i^{*(4)} - \bar{X}_n).$$

$\tilde{T}_n^{*(4)}$ no $T_n^{*(4)}$ atšķiras ar papildus iekļautiem elementiem butstrapa izlasēs, ja eksistē tādi elementi, kuri butstrapa izlasē atrodas aiz n-tā elementa. Redzams, ka $N_1 - L_k \leq n \leq N_1$, tātad starpība starp n un N_1 ir vismaz L_k. Tā kā pieņēmums par bloka garumu l ir tāds, ka l ir *nenozīmīgs* pret n, tad starpība starp abām statistikām arī ir *nenozīmīga* pie sekojošās teorēmas nosacījumiem.

Teorēma 9. *Pieņemsim, ka* $E\|X_1\|^{6+\delta} < \infty$, Σ_∞ *ir nesingulāra un* $\sum_{n=1}^{\infty} n^5\alpha(n)^{\delta/(6+\delta)} < \infty$ *kādam* $\delta > 0$, *pie tam* $p + (n^{1/2}p)^{-1} \to 0$, *kad* $n \to \infty$. *Tad*

$$\sup_{x \in \mathbf{R}^n} \left| P_*(\tilde{T}_n^{*(4)} \leq x) - P(T_n \leq x) \right| \xrightarrow{p} 0, \ \textit{kad } n \to \infty.$$

Teorēmas pierādījums un tās vispārinājums ir atrodams [7].

6. Bloku butstrapa novērtējumu vispārinājumi

Šajā nodaļā parādīsim vispārinātu bloku butstrapa novērtējumu īpašības un veiksim teorētisku salīdzinājumu ar simulētiem datiem. Apskatīsim novērtējumus, kurus var izteikt, kā gludas funkcija no izlases vidējās vērtības, M-novērtējumus, kuri iekļauj parametru vislielākās ticamības novērtējumus. Kā arī tiks parādīta diferencējamu statistisko funkcionāļu novērtējumu konsistence bloku butstrapam.

6.1. Gludas funkcijas no vidējās vērtības

Šajā nodaļā apskatīsim novērtējumus, kurus varam iekļaut *Gludās funkcijas modeļa* klasē (no angļu valodas ,,Smooth function model"), ko attīstījuši Bhattacharya un Ghosh ([12]). Pieņemsim, ka $\{X_{0i}\}_{i\in\mathbf{Z}}$ ir $\mathbf{R}^{d_0}$ vērtību stacionārs process, $f : \mathbf{R}^{d_0} \to \mathbf{R}$ ir Boreļa mērojama funkcija un $H : \mathbf{R}^d \to \mathbf{R}$ ir gluda funkcija. Pieņemsim, ka mūsu interesējošais parametrs ir dots ar $\theta = H(Ef(X_{0i}))$. Dabisks θ novērtējums ir, piemēram, $\hat{\theta}_n = H(n^{-1}\sum_{i=1}^n f(X_{0i}))$. Tādejādi θ un $\hat{\theta}_n$ ir gludas funkcijas no pārveidotās virknes $\{f(X_{0i})\}_{i\in\mathbf{Z}}$, attiecīgi, populācijas un izlases vidējās vērtības. Daudzus parametrus un to novērtējumus var izteikt, kā gludas funkcija no vidējās vērtības.

Piemērs 4. Pieņemsim, ka $\{X_{0i}\}_{i\in\mathbf{Z}}$ ir reālu vērtību laikrinda ar autokovariāciju funkciju $\gamma(k) = \mathrm{Cov}(X_{0i}, X_{0(i+k)}),\ i, k \in \mathbf{Z}$. $\gamma(k)$ novērtējums izlasei $X_{0i}, \ldots, X_{0n}$ ar apjomu n ir

$$\hat{\gamma}_n(k) = (n-k)^{-1}\sum_{i=1}^{n-k} X_{0i}X_{0(i+k)} - \bar{X}^2_{0(n-k)}, \tag{26}$$

kur $\bar{X}_{0(n-k)} = (n-k)^{-1}\sum_{i=1}^{n-k} X_{0i}$. Pieminēsim, ka novērtējums $\hat{\gamma}_n(k)$ ir versija no izlases autokovariācijas ar kārtu k. Parādīsim, ka novērtējums $\hat{\gamma}_n(k)$ un parametrs $\gamma(k)$ atbilst Gludas funkcijas modelim. Definēsim jaunu virkni no gadījuma vektoriem

$$X_i = (X_{0i}, X_{0i}X_{0(i+k)})',\ i \in \mathbf{Z}.$$

Tad parametrs $\theta \equiv \gamma(k)$ ir

$$\theta = EX_{01}X_{0(1+k)} - (EX_{01})^2 = H(EX_1), \tag{27}$$

kur $H : \mathbf{R}^2 \to \mathbf{R}$ ir $H((x, y)) = y - x^2$. Līdzīgi, novērtējums $\hat{\theta}_n \equiv \hat{\gamma}_n(k)$ ir

$$\hat{\theta}_n = \hat{\gamma}_n(k) = H(\bar{X}_{n-k}),$$

kur $\bar{X}_{n-k} = (n-k)^{-1}\sum_{i=1}^{n-k} X_i$. Šis piemērs rāda, ka varam autokovariācijas parametru un novērtējumu izteikt, kā Gludas funkcijas modeli.

Piemērs no [7].

Piemērus, kā autokorelācijas koeficientus un Jula-Volkera novērtējumus izteikt, kā Gludas funkcijas modeli, skatīt [7].

Tālāk parādīsim Gludās funkcijas modeļa konsistenci dažādām bloku butstrapa metodēm, kur novērtējumi ir centrēti un normēti

$$T_{1n} = \sqrt{n}(\hat{\theta}_n - \theta),$$

kur $\theta = H(Ef(X_{01}))$ un $\hat{\theta}_n = H(n^{-1}\sum_{i=1}^{n} f(X_{0i}))$. T_{1n} bloku butstrapa versiju var definēt, izmantojot 4. un 5. nodaļās izklāstītās idejas. Pieņemsim, ka $X_i = f(X_{0i})$, $i \in \mathbf{Z}$, tātad $\theta = H(EX_1)$ un $\hat{\theta}_n = H(\bar{X}_n)$. Tālāk, pieņemsim, ka $\mathcal{X}_n^{*(j)}$, $j = 1, 2, 3$ ir butstrapa izlases ar apjomu n_1, bloku garumu l un bloku skaitu b no transformētajiem mainīgajiem $\mathcal{X}_n = (X_1, \ldots, X_n)$, attiecīgi MBB, NBB un CBB metodēm, bet $\mathcal{X}_n^{*(4)}$ ir izlase ar N_1 novērojumiem, iegūta ar SB metodi ar sagaidāmo bloka garumu $l^{-1} \in (1, n)$, kur $b = \lfloor n/l \rfloor$, $n_1 = bl$ un $N_1 = L_1 + \ldots + L_K$. Ar $\bar{X}_n^{*(j)}$, $j = 1, 2, 3, 4$ apzīmēsim atbilstošās butstrapa izlases vidējo vērtību. Līdzīgi, kā 5.2.2. nodaļā, n var nesakrist ar n_1 un N_1, taču vidējo vērtību atšķirība ir *niecīga*. T_{1n} bloku butstrapa versijas ir

$$\begin{aligned} T_{1n}^{*(j)} &= \sqrt{n_1}(\theta^{*(j)} - \tilde{\theta}_{n,j}),\ j = 1, 2, 3 \\ T_{1n}^{*(j)} &= \sqrt{N_1}(\theta^{*(j)} - \tilde{\theta}_{n,j}),\ j = 4, \end{aligned}$$

kur $\theta_n^{*(j)} = H(\bar{X}_n^{*(j)})$ un $\tilde{\theta}_{n,j} = H(E_*\bar{X}_n^{*(j)})$, $1 \leq j \leq 4$. Varam formulēt sekojošu rezultātu. Ja $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbf{Z}_+^d$, tad teiksim, ka $|\alpha| = \alpha_1 + \ldots + \alpha_d$, $\alpha! = \prod_{j=1}^{d} \alpha_j!$ un $D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}$.

Teorēma 10. *Pieņemsim, ka funkcija H ir diferencējama EX_1 apkārtnē $N_H \equiv \{x \in \mathbf{R}^d : \|x - EX_1\| < 2\eta\}$, kādam $\eta > 0$, $\sum_{|\alpha|=1} |D^\alpha H(EX_1)| \neq 0$ un H pirmās kārtas atvasinājumi apmierina Lipšica nosacījumus ar kārtu $k > 0$ apkārtnē N_H. Turklāt Teorēmas 7 un 9 nosacījumi izpildās, attiecīgi gadījumos $j = 1, 2, 3$ un $j = 4$. Tad*

$$\sup_{x \in \mathbf{R}^d} \left| P_*(T_{1n}^{*(j)} \leq x) - P(T_{1n} \leq x) \right| \xrightarrow{p} 0, \text{ kad } n \to 0$$

katram $j = 1, 2, 3, 4$.

Pierādījumu skatīt [7].

Piezīme 1. *Daudzos pielietojumos mūs var interesēt parametra $\theta = H(EX_1)$ novērtējums formā*

$$\bar{\theta}_n = H\left((n-k_1)^{-1}\sum_{i=1}^{n-k_1} f_1(X_{0i}), \ldots, (n-k_d)^{-1}\sum_{i=1}^{n-k_d} f_d(X_{0i})\right),$$

kādiem fiksētiem $k_1, \ldots, k_d$, neatkarīgiem no n, kur $f_1, \ldots, f_d$ ir funkcijas $f : \mathbf{R}^{d0} \to \mathbf{R}^d$ komponentes. Var parādīt, ka Teorēma 10 ir spēkā novērtējumam $\bar{\theta}_n$, ja aizvietojam $\hat{\theta}_n$ ar $\bar{\theta}_n$ un $\theta_n^{*(j)}$ ar attiecīgo bloka butstrapa versiju. Pie tam, var parādīt, ka bloku butstrapa konsistence sadalījuma funkcijai saglabājas, ja funkcija f pieņem vektoru vērtības, kur katra komponente no H saglabā teorēmas nosacījumus.

Piezīme 2. *Tāpat, kā pēc Teorēmām 6 un 8, bloku butstrapa metodes dod konsistentu asimptotiskās dispersijas novērtējumu parametram $\hat{\theta}_n$, kurš apskatīts Teorēmā 10, taču ir nepieciešami stingrāki momentu un jauktie nosacījumi, nekā Teorēmā 10. Šie nosacījumi ir doti [7].*

Piemērs 5. Turpināsim Piemēru 4. Apskatīsim laikrindas modeli

$$X_i = -0.5X_{i-1} + \epsilon_i + 0.8\epsilon_{i-1}, \tag{28}$$

kur ϵ_i ir neatkarīgi un vienādi sadalīti ar sadalījumu $N(0,1)$. Pieņemsim, ka izlases apjoms $n = 102$, un apskatīsim otrās kārtas kovariācijas, $k = 2$. Pēc Teorēmas 10 un Piezīmes 2 visas bloku butstrapu metodes dod konsistentus novērtējumus

$$T_{1n} \equiv \sqrt{n-k}(\hat{\theta}_n - \theta) = \sqrt{n-k}(\hat{\gamma}_n(k) - \gamma(k))$$

otrās kārtas momentam, kur $\hat{\gamma}_n(k)$ un $\gamma(k)$ ir tādi, kā Piemērā 4. Novērtēsim parametru

$$\varphi_n \equiv ET_{1n}^2 = (n-k)\mathrm{MSE}(\hat{\gamma}_n(k)),$$

jo $ET_{1n} = 0$.

Apskatīsim MBB butstrapa metodes ,,vienkāršo" pieeju. Definēsim jaunus vektorus $X_i = (X_{0i}, X_{0i}X_{0(i+2)})$, $i = 1, \ldots, 100$ ar, piemēram, $l = 15$. No jaunajiem X_i veidosim blokus $\mathcal{B}_i = (X_i, \ldots, X_{i+14})$, $i = 1, \ldots, 86$. Tālāk veidosim butstrapa izlases no $k_0 = \lfloor 100/15 \rfloor + 1 = 7$ blokiem, iegūstot $\mathcal{B}_1^*, \ldots, \mathcal{B}_7^*$ un butstrapa izlasi $X_1^*, \ldots, X_{105}^*$. Butstrapa ,,vienkāršās" pieejas T_{1n} versija ir

$$T_{1n}^* = \sqrt{100}\left(H(\bar{X}_{100}^*) - H(\hat{\mu}_{100})\right),$$

kur $\bar{X}_{100}^*$ ir butstrapa izlases vidējā vērtība no pirmajiem 100 elementiem, $\hat{\mu}_{100} \equiv E_*\bar{X}_{100}^*$ un $H((x,y)) = y - x^2$. Centrējošo vērtību iegūsim no [7],

$$\hat{\mu}_{100} = (100)^{-1}\left((7-1)(86)^{-1}\sum_{i=1}^{86} V_i(15) + (86)^{-1}\sum_{i=1}^{86} V_i(a)\right),$$

kur $a = n - k - (k_0 - 1)l = 10$ un $V_i(0) = 0$, $V_i(m) = X_i + \ldots + X_m$ ir pirmo m elementu summa no bloka $\mathcal{B}_i$, $1 \leq m \leq 15$. Tas seko no tā, ka $X_1^*, \ldots, X_{100}^*$ sastāv no pilniem 6 blokiem un pirmajiem 10 elementiem no 7-tā bloka. Protams, $\hat{\mu}_{100}$ visām butstrapa izlasēm ir viens un tas pats. MBB φ_n novērtējums ir

$$\hat{\varphi}_n = E_* \left(T_{1n}^*\right)^2.$$

Tā kā slēgtā formā izteiksme $\hat{\varphi}_n$ nav pieejama, dēļ $\hat{\theta}_n = H(\bar{X}_{100})$ nelinearitātes, tad iegūsim $\hat{\varphi}_n$ ar Monte-Carlo simulāciju palīdzību, kā

$$\hat{\varphi}_n^B = B^{-1} \sum_{r=1}^{B} \left(T_{1n,r}^*\right)^2,$$

kur B ir butstrapa izlašu skaits un $T_{1n,r}^*$ ir vienas butstrapa izlases realizācija. Rezultāti ar $l = 15$ un citiem bloku garumiem ir 2. tabulā.

Tālāk apskatīsim NBB metodi. Sadalīsim elementus blokos tā, ka $\mathcal{B}_i = (X_{(i-1)15+1}, \ldots, X_{15*i})$, $i = 1, \ldots, b$, tas ir, bloki $\mathcal{B}_i$ un $\mathcal{B}_j$ savā starpā nešķeļas, ja $i \neq j$, kur $b = \lfloor (n-k)/l \rfloor = 6$. Atlikusī procedūra ir identiska MBB metodei, tikai ņemsim vēra, ka $\hat{\mu}_{100} = (90)^{-1} \sum_{i=1}^{90} X_i$, jo pēdējie 10 elementi no sākotnējās izlases nekad neparādīsies butstrapotajās izlasēs.

CBB metodei definēsim elementus $Y_{100*p+i} = X_i$, visiem $p \in \mathbf{Z}$ un $i = 1, \ldots, 100$. Iegūsim periodisku virkni $\{Y_{100,i}\}_{i \in \mathbf{Z}}$, no kuras veidosim blokus $\mathcal{B}_j = (Y_j, \ldots, Y_{j+15-1})$, $j = 1, \ldots, 100$. Atšķirība, kā jau iepriekš minēts, no MBB metodes ir tā, ka visiem elementiem X_i, $i = 1, \ldots, 100$ ir vienāda varbūtība, tikt iekļautiem butstrapa izlasē, tāpēc $\hat{\mu}_{100} = \bar{X}_{100}$.

SB metodes atšķirība no iepriekšējām ir tā, ka bloku garums l nav fiksēts. Tāpat, kā CBB metodē, definēsim periodisku virkni $\{Y_{100,i}\}_{i \in \mathbf{Z}}$. Veidosim butstrapa izlasi no $M = \inf\{1 \leq m \leq 100 : L_1 + \ldots + L_m \geq 100\}$ blokiem, kur L_i ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar ģeometrisko sadalījumu ar parametru $1/15$. Katrs L_i nosaka i-tā bloka garumu. Tātad butstrapa izlasēm var būt dažādi apjomi, lielāki vai vienādi par 100. Statistikas T_{1n}^* aprēķinam ņemsim tikai pirmos 100 elementus no butstrapotajām izlasēm. Tāpat, kā CBB metodei, $\hat{\mu}_{100} = \bar{X}_{100}$.

Teorētisko $\gamma(2)$ atradīsim pēc (4). Šajā gadījumā $\gamma(2) = -0.120$. Ar 10000 simulāciju palīdzību atradīsim $\varphi_n = 1.332$. Lietosim MBB, NBB, CBB un SB metodes, lai novērtētu parametru φ_n ar butstrapu. Ņemsim $B = 1000$ butstrapa izlases un veiksim butstrapus ar dažādiem bloku garumiem $l = 5, 10, 15, 20, 25$.

2. tabula Statistikas $\varphi = ET_{1n}^2$,,vienkāršās" pieejas butstrapa novērtējumi.

Bloka garums	MBB	NBB	SBB	SB
5	2.093	1.292	1.978	1.771
10	1.286	1.645	1.515	1.457
15	0.981	1.672	1.291	1.362
20	1.241	1.339	1.433	1.362
25	1.044	1.768	1.448	1.362

Līdzīgi varam veikt ,,naivo" butstrapa pieeju. MBB bloki ir $\mathcal{B}_i = (X_{0i}, \ldots, X_{0(i+14)})$, $i = 1, \ldots, 86$. Šajā gadījumā mūsu statistika T_{1n} ir

$$T_{1n}^* = \sqrt{100}\left(\gamma_n^*(2) - \hat{\gamma}_n(2)\right),$$

kur $\gamma_n^*(2) = (100)^{-1}\sum_{i=1}^{100} X_{0i}^* X_{0(i+2)}^* - (\bar{X}_{100}^*)^2$ un $\bar{X}_{100}^* = (100)^{-1}\sum_{i=1}^{100} X_{0i}$. Tātad ,,naivajā" pieejā nav nepieciešami papildus mainīgie X_i, butstrapošanu veic ar sākotnējiem X_{0i}. $\hat{\gamma}_n(2)$ ir dotās izlases kovariācijas novērtējums. Četru metožu rezultāti ir 3. tabulā.

3. tabula Statistikas $\varphi = ET_{1n}^2$,,naivās" pieejas butstrapa novērtējumi.

Bloka garums	MBB	NBB	SBB	SB
5	2.447	2.634	2.383	1.921
10	1.651	2.367	1.786	1.744
15	1.553	1.632	1.588	1.402
20	1.468	2.014	1.396	1.313
25	1.621	2.234	1.673	1.108

Intuitīvi aprakstīsim atšķirību starp ,,vienkāršo" un ,,naivo" pieeju. Neatkarīgi no bloku butstrapa metodes, pieņemsim, ka eksistē divi butstrapoto izlašu elementi $X_{i_0}^*$ un $X_{i_0+2}^*$, kādam i_0, un ka katrs no viņiem atrodas savā butstrapotajā blokā. Tad ,,naivai" pieejai reizinājums $X_{i_0}^* X_{i_0+2}^*$ ieiet kovariāciju aprēķinā. Mūsu pieņēmums ir par bloku neatkarību, tātad elementi $X_{i_0}^*$ un $X_{i_0+2}^*$ ir neatkarīgi, kas nav *pareizi*. Savukārt ,,vienkāršās" pieejas butstraps bez X_{j_0} pārkārto arī atbilstošo reizinājumu $X_{j_0}X_{j_0+2}$, kādam j_0, no dotās izlases. Tātad atkarības struktūra starp šiem elementiem ir saglabāta. Teorētiski šo atšķirību jau pamatojām 4.2. nodaļā.

No rezultātiem varam secināt, ka ,,vienkāršā" butstrapa pieeja dod stabilākus rezultātus pie dažādiem bloku garumiem l, tas ir, rezultāti ir *tuvāki* teorētiskajam rezultātam. ,,Naivās" pieejas rezultātiem, salīdzinot ar teorētisko vērtību $\varphi_n = 1.332$, ir lielākas atšķirības pie maziem bloku garumiem, jo ,,naivā" pieeja rada papildus novirzi novērtējumiem.

Tomēr arī šis uzdevums nedod atbildi uz jautājumu, kuru bloku garumu izvēlēties. Par to, kā teorētiski izvēlēties bloka garumu, kurš dod precīzākos rezultātus, apskatīsim 7. nodaļā.

Piemēra ideja realizēta no [7] un programmas R kods ir atrodams pielikumā.

6.2. M-novērtējumi

Pieņemsim, ka $\{X_i\}_{i\in\mathbf{Z}}$ ir stacionārs process, kurš pieņem $\mathbf{R}^d$ vērtības. Mūsu interesējošais parametrs θ ir vienādojuma

$$E\Psi(X_1,\dots,X_m,\theta)=0 \tag{29}$$

atrisinājums, kādai funkcijai $\Psi:\mathbf{R}^{dm+s}\to\mathbf{R}^s,\ m,s\in\mathbf{N}$. M-novērtējums $\hat{\theta}_n$ parametram θ ir definēts, kā atrisinājums *novērtētajam* vienādojumam

$$(n-m+1)\sum_{i=1}^{n-m+1}\Psi(X_i,\dots,X_{i+m-1},\hat{\theta}_n)=0. \tag{30}$$

Novērtējumi, kuri definēti ar novērtēto vienādojumu formā (30), tiek saukti par *vispārinātajiem M-novērtējumiem*. Šī novērtējumu klase satur vislielākās ticamības novērtējumus un zināmus robustus parametra novērtējumus daudzos pazīstamos laikrindu modeļos, ieklaujot autoregresīvos slīdošā vidējā modeļus ar parametriem $p,q\in\mathbf{Z}_+$ (ARMA(p,q)).

Lai definētu $\hat{\theta}_n$ butstrapa versiju, pieņemsim, ka $Y_i=(X_i,\dots,X_{i+m-1}),\ 1\le i\le n_0$, kur $n_0=n-m+1$. Ar $Y_1^{*(j)},\dots,Y_{n_0}^{*(j)}$ apzīmēsim ,,vienkāršā" bloku butstrapa izlasi ar apjomu n_0, iegūtu no vektoriem Y ar j-to metodi, $j=1,2,3,4$. Dēļ (29) strukturāliem ierobežojumiem, ir iespējami vairāki veidi, lai definētu butstrapa versiju vispārinātam M-novērtējumam un tā centrētai un normētai versijai

$$T_{2n}\equiv\sqrt{n}(\hat{\theta}_n-\theta).$$

Varam definēt $\hat{\theta}_n$ butstrapa versiju $\theta_n^{*(j)}$, kā atrisinājumu vienādojumam

$$n_0^{-1}\sum_{i=1}^{n_0}\Psi(Y_i^{*(j)},\theta_n^{*(j)})=0,\ j=1,2,3,4. \tag{31}$$

T_{2n} butstrapa versija ir dota

$$T_{2n}^{*(j)}=\sqrt{n_0}(\theta_n^{*(j)}-\tilde{\theta}_n^{(j)}), \tag{32}$$

kur vērtība $\tilde{\theta}_n^{(j)}$ ir vienādojuma

$$n_0^{-1}\sum_{i=1}^{n_0}E_*\Psi(Y_i^{*(j)},\tilde{\theta}_n^{(j)})=0 \tag{33}$$

atrisinājums. CBB un SB metodes, lietotas datiem $Y_1, \ldots, Y_n$, reducē vienādojumu (33) uz (30) un $\tilde{\theta}_n^{(j)} = \hat{\theta}_n$, $j = 3, 4$. Tātad sākotnējās izlases $\hat{\theta}_n$ tiek izmantots, lai centrētu butstrapa versiju $\theta_n^{*(j)}$. Savukārt MBB un NBB metodēm nav teikts, ka $\tilde{\theta}_n^{(j)}$ ir vienāds ar $\hat{\theta}_n$. MBB un NBB gadījumos nepieciešams risināt papildus vienādojumus, lai aprēķinātu butstrapa versiju $T_{2n}^{*(j)}$ un *pareizi* centrētu $\theta_n^{*(j)}$. Protams, būtu ērti aizvietot $\tilde{\theta}_n^{(j)}$ ar $\hat{\theta}_n$ un definēt

$$\check{T}_{2n}^{*(j)} = \sqrt{n_0}(\theta_n^{*(j)} - \hat{\theta}_n), \tag{34}$$

kā T_{2n} butstrapa versiju, kad $j = 1, 2$. Tomēr, centrējot $\tilde{\theta}_n^{(j)}$ ar $\hat{\theta}_n$, MBB un NBB metodēs tas rada *papildus* novirzi, kas parasti noved pie sliktākas kārtas $\mathcal{L}(T_{2n})$ aproksimācijas ar $\mathcal{L}(\check{T}_{2n}^{*(j)}|\mathcal{X}_n)$, salīdzinot ar klasisko normālo aproksimāciju.

Cita pieeja, definējot T_{2n} butstrapa versiju, ir M-novērtējuma butstrapa versijas definīcijā veidot strukturālu attiecību starp (29) un (30). Ja (30) aizvietosim $\hat{\theta}_n$ ar θ, tad sagaidāmā vērtība kreisajā pusē būs 0. Tātad, novērtētā funkcija ir *nenovirzīta* attiecībā pret θ. Tomēr butstrapotā M-novērtējuma definīcijā (31), šī nenovirzītības īpašība, novērtētajai funkcijai, ne vienmēr izpildās. Vienkāršu risinājumu piedāvāja Shaorack (1982), butstrapojot lineārās regresijas ar neatkarīgiem un vienādi sadalītiem atlikumiem M-novērtējumus. Definēsim alternatīvu $\hat{\theta}_n$ butstrapa versiju $\theta_n^{**(j)}$, kā atrisinājumu pārveidotam vienādojumam

$$n_0^{-1} \sum_{i=1}^{n_0} \left(\Psi(Y_i^{*(j)}, \theta_n^{**(j)}) - \hat{\psi}_j \right) = 0, \tag{35}$$

kur $\hat{\psi}_j = n_0^{-1} E_* \{\sum_{i=1}^{n_0} \Psi(Y_i^{*(j)}, \hat{\theta}_n)\}$. Pie tam, visiem $j = 1, 2, 3, 4$, novērtējot funkciju $\sum_{i=1}^{n_0} \left(\Psi(Y_i^{*(j)}, t) - \hat{\psi}_j \right)$, sagaidāmā vērtība ir 0 pie $t = \hat{\theta}_n$. Tātad $\hat{\psi}_j$ ir atbilstoša konstante, kas dod to, ka novērtētā funkcija iekš (35) ir nenovirzīta, ja mēs centrējam M-novērtējumus ar $\hat{\theta}_n$. Tad T_{2n} butstrapa versija alternatīvajai pieejai ir

$$T_{2n}^{**(j)} = \sqrt{n_0}(\theta_n^{**(j)} - \hat{\theta}_n). \tag{36}$$

Salīdzinot abas alternatīvas (36) un (32), (36) priekšrocība ir tā, ka, lietojot MBB un NBB metodes, ir nepieciešams atrisināt tikai vienu vienādojumu sistēmu (35), nevis divas-(31) un (33). Tā kā CBB un SB metodēm $\hat{\psi}_j = n_0^{-1} \sum_{i=1}^{n_0} \Psi(Y_i, \hat{\theta}_n) = 0$, $j = 1, 2$, tad abas pieejas dod vienādus T_{2n} butstrapa novērtējumus.

Sekojošais rezultāts pamato bloku butstrapa aproksimācijas abām pieejām $T_{2n}^{*(j)}$ un $T_{2n}^{**(j)}$, $j = 1, 2, 3, 4$. Pieņemsim, $\Sigma_\Psi = \lim_{n\to\infty} D(n_0^{-1/2} \sum_{i=1}^{n_0} \Psi(Y_i, \theta))$ un D_Ψ ir $s \times s$ matrica ar (i, j)-to elementu $E\left(\frac{\partial}{\partial \theta_j} \Psi_i(Y_1, \theta)\right)$, $1 \le i, \ j \le s$. Pie tam, pieņemsim, ka novērtēto vienādojumu atrisinājumi ir mērojami un vieni vienīgi. Apzīmēsim $\Delta(r,\delta) = 1+\sum_{n=1}^{\infty} n^{2r-1}\alpha^{\delta/(2r+\delta)}(n)$, kur $\alpha(\cdot)$ ir α-jauktais koeficients procesam $\{X_i\}_{i\in\mathbf{Z}}$. r ir vesels, pozitīvs skaitlis.

Teorēma 11. *Pieņemsim, ka*

- *$\Psi(y,t)$ ir diferencējama pret t gandrīz visiem y (pie F_Y) un Ψ pirmās kārtas parciālie atvasinājumi apmierina Lipšica nosacījumus ar kārtu $k \in (0,1]$, gandrīz droši (F_Y), kur F_Y ir Y_1 varbūtību sadalījums;*
- *$E\Psi(Y_1,\theta) = 0$ un Σ_Ψ un D_Ψ ir nesingulāras;*
- *eksistē $\delta > 0$ tāds, ka $E\|D^\alpha\Psi(Y_1,\theta)\|^{2r_j+\delta} < \infty$ visiem $\alpha \in \mathbf{Z}_+^s$ ar $|\alpha| = 0,1$, un $\Delta(r_j,\delta) < \infty$, kur $r_j = 1$ visiem $j = 1,2,3$ un $r_j = 3$, kad $j = 4$;*
- *$l^{-1} + n^{-1/2}l = o(1)$ un $p + (n^2p)^{-1} = o(1)$, kad $n \to \infty$.*

Tad

1. *$\{\hat\theta_n\}_{n\ge1}$ ir konsistents priekš θ un*

$$\sqrt{n}(\hat\theta_n - \theta) \xrightarrow{d} N(0, D_\Psi\Sigma_\Psi D'_\Psi);$$

2. *$j = 1,2,3,4$,*

$$\sup_{x\in\mathbf{R}^s}\left|P_*(T_{2n}^{**(j)} \le x) - P(T_{2n} \le x)\right| \xrightarrow{p} 0, \text{ kad } n \to \infty;$$

3. *2 ir patiess, ja $T_{2n}^{**(j)}$ aizvietosim ar $T_{2n}^{*(j)}$ no (32).*

Teorēmas pierādījums ir atrodams [7].

6.3. Diferencējami funkcionāļi

Šajā nodaļā parādīsim MBB metodes konsistenci, aproksimējot novērtējumus, kuri ir empīriska procesa gludas funkcijas. Pieņemsim, ka $\{X_i\}_{i\in\mathbf{Z}}$ ir virkne no R^d telpas gadījuma vektoriem. Definēsim m-dimensionālas $X_1,\ldots,X_n$ apakšvirknes empīrisko sadalījumu, kā

$$F_n^{(m)} = (n-m+1)\sum_{i=1}^{n-m+1}\delta_{Y_i},$$

kur $Y_i \equiv (X_i,\ldots,X_{i+m-1})$, $i \ge 1$, un kur δ_y ir varbūtības mērs ar vērtību 1 punktā y. Varbūtības mērs $F_n^{(m)}$ darbojas, kā novērtējums robežsadalījumam $F^{(m)}$ m-dimensionālai apakšvirknei $X_1,\ldots,X_m$.

Pieņemsim, ka interesējošais parametrs θ ir s-dimensionāls funkcionālis no $F^{(m)}$

$$\theta = T(F^{(m)}). \tag{37}$$

Tad dabisks θ novērtējums ir

$$\hat{\theta} = T(F_n^{(m)}). \tag{38}$$

Daudzus, bieži lietotus, novērtējumus var izteikt ar (38), piemēram, vispārinātos M-novērtējumus, apskatītus 6.2. nodaļā. Vispārinātajiem M-novērtējumiem atbilstošais funkcionālis $T(\cdot)$ ir *netieši* definēts ar attiecību

$$\int \Psi(x_1, \ldots, x_m, T(G^{(m)}))dG^{(m)} = 0,\ G^{(m)} \subset \mathcal{G}^{(m)},$$

kādai atbilstošai saimei $\mathcal{G}^{(m)}$ no varbūtību mēriem, kas uzdoti $\mathbf{R}^{dm}$, atkarībā no funkcijas Ψ. Raksturosim vēl vienu nozīmīgu robustu novērtējumu klasi, ko var izteikt šādā formā.

Piemērs 6. Pieņemsim, ka process $\{X_i\}_{i\in\mathbf{Z}}$ pieņem reālas vērtības. Tad visiem $0 < \alpha < 1/2$, α-neiekļautās vidējā vērtība ir

$$\hat{\theta}_n = (n(1-2\alpha))^{-1} \sum_{i=\lfloor n\alpha \rfloor+1}^{\lfloor n(1-\alpha)\rfloor} X_{i:n}, \tag{39}$$

kur $X_{1:n} \leq \ldots \leq X_{n:n}$ ir sakārtota $X_1, \ldots, X_n$. Šeit ir labota [7] dotā definīcija. Apzīmēsim $F_n^{(1)} = F_n$ un $F^{(1)} = F$. Tad novērtējums $\hat{\theta}_n$ no (39) ir ekvivalents viendimensionāla empīriskā sadalījuma F_n funkcionālim $T(\cdot)$

$$T(F_n) \equiv (1-2\alpha)^{-1} \int_\alpha^{1-\alpha} F_n^{-1}(u)du,$$

kur katram varbūtību sadalījumam G uz $\mathbf{R}$, $G^{-1}(u) = \inf\{x \in \mathbf{R} : G((-\infty, x]) \geq u\}$, $0 < u < 1$ ir G kvantīles pārveidojums. Novērtējumu $\hat{\theta}_n$ jeb $T(F_n)$ lieto, lai novērtētu parametru

$$\theta = T(F) \equiv (1-2\alpha)^{-1} \int_\alpha^{1-\alpha} F^{-1}(u)du.$$

α-neiekļautās vidējā vērtība $\hat{\theta}_n$ ir robusts θ novērtējums, kurš izslēdz izlecēju ietekmi. Pie tam, ja $\alpha \to (1/2)^-$, tad parametrs θ ir populācijas mediāna $F^{-1}(1/2)$ (pieņemot, ka $F^{-1}(u)$ ir nepārtraukta punktā $u = 1/2$), kamēr, ja $\alpha = 0$, tad iegūsim $\theta = EX_1$ ir populācijas vidējā vērtība. Tātad, dažādām α vērtībām, θ ir kāda no bieži lietotām statistikām.

Piemērs ņemts no [7].

Atgriezīsimies pie vispārēja novērtējuma (38), kurš ir funkcionālis no $F_n^{(m)}$. Ja funkcionālis $T(\cdot)$ ir gluds, tad $\hat{\theta}_n$ asimptotiskais sadalījums ir iegūstams no empīriskā procesa $\left\{\sqrt{n}\left(F_n^{(m)}(\cdot) - F^{(m)}(\cdot)\right)\right\}$ asimptotiskā sadalījuma, atbilstošā metriskā telpā, lietojot Delta metodi. Lai parādītu bloku butstrapu konsistenci šādu novērtējumu sadalījumiem, mums būtu jāzina butstrapotā empīriskā procesa asimptotiskā uzvedība.

6.3.1. Empīriska procesa butstrapošana

Empīriska procesa asimptotiskā sadalījuma rezultātus atkarīgiem gadījuma lielumiem ir parādījuši dažādi autori (piemēram, Billingsley ([13])). Līdzīgi rezultāti, butstrapotam empīriskam procesam, ir zināmi tikai MBB un CBB gadījumos. Ja ņemsim $Y_i = (X_i, \ldots, X_{i+m-1}),\ i \in \mathbf{Z}$ un $n_0 = n - m + 1$, tad, attiecībā pret Y,

$$F_n^{(m)} = n_0 \sum_{i=1}^{n_0} \delta_{Y_i},$$

un tas reducē vispārīgo problēmu uz gadījumu, kad $m = 1$, dm-dimensionāliem gadījuma vektoriem $Y_1, \ldots, Y_{n_0}$. Tāpēc šajā nodaļā pieņemsim, ka $m = 1$. Pie tam, apzīmēsim $F_n^{(1)} = F_n$ un $F^{(1)} = F$.

Ar $\mathbf{D}_d$ apzīmēsim telpu no visām reālu vērtību funkcijām, definētām $[-\infty, \infty]^d$, kuras ir nepārtrauktas no augšas un tām eksistē robežas no apakšas. Pievienosim $\mathbf{D}_d$ (paplašināto) Skorohoda metriku. $F_n(x)$ un $F(x)$ ir sadalījuma funkcijas varbūtību mēriem F_n un F. Tad

$$\begin{aligned} F_n(x) &= n^{-1} \sum_{i=1}^{n} \mathbf{1}(X_i \leq x), \\ F(x) &= P(X_1 \leq x), \end{aligned}$$

kur $x \in [-\infty, \infty]^d$. Definēsim empīrisku procesu

$$W_n(x) = \sqrt{n}(F_n(x) - F(x)),\ x \in [-\infty, \infty]^d.$$

Tad pie dažiem likumsakarīgiem nosacījumiem, empīrisks process W_n konverģē vāji (kā $\mathbf{D}_d$ vērtību gadījuma elementi) uz Gausa procesu W no $[-\infty, \infty]^d$, kas apmierina

$$\begin{aligned} &EW(x) = 0, \\ &EW(x)W(y) = \sum_{k=-\infty}^{\infty} \mathrm{Cov}(\mathbf{1}(X_1 \leq x), \mathbf{1}(X_{1+k} \leq y)), \\ &P(W \in \mathbf{C}_d^0) = 1 \end{aligned}$$

visiem $x, y \in [\infty, \infty]^d$, kur $\mathbf{C}_d^0$ ir nepārtrauktu funkciju klase no $[\infty, \infty]^d$, kuras tiecas uz 0 pie $(-\infty, \ldots, -\infty)$ un $(\infty, \ldots, \infty)$. Nākamā teorēma parāda, ka līdzīgi rezultāti ir spēkā butstrapotam empīriskam procesam. Ar $F_n^*(x)$ apzīmēsim empīrisku procesu no MBB izlasēm, apjomā n_1 ar bloku garumu l, un pieņemsim, ka $W_n^*(x) = \sqrt{n_1}(F_n^*(x) - E_* F_n^*(x)),\ x \in \mathbf{R}^d$.

Teorēma 12. *Pieņemsim, ka $\{X_n\}_{n\geq 1}$ ir stacionāru α-jauktu $\mathbf{R}^d$ vērtību gadījuma vektoru virkne ar $\sum_{i=1}^{\infty} i^{8d+7}[\alpha(i)]^{1/2} < \infty$ un ka X_1 ir nepārtraukts. Turklāt, $l \to \infty$ un $l = O(n^{1/2-\epsilon})$, kad $n \to \infty$, kādiem $0 < \epsilon < 1/2$. Tad*

$$W_n^* \xrightarrow{d} W \text{ gandrīz droši, kad } n \to \infty.$$

Teorēmas pierādījums un tās vispārinājums, kā arī SB metodes rezultāti, atrodami [7].

6.3.2. MBB konsistence diferencējamiem statistiskiem funkcionāļiem

Pieņemsim, ka $\hat{\theta}_n$ ir novērtējums, ko var reprezentēt ar *gludu* funkcionāli no m-dimensionāla empīriska sadalījuma $F_n^{(m)}$ (kā (38)), ar kādu $m \geq 1$. Vispārīga pieeja, lai iegūtu šādu statistisku funkcionāļu asimptotiskos sadalījumus, lietojot diferencējamību, ieviesa von Mises ([14]). Šeit apskatīsim statistiskos funkcionāļus, kuri ir *Frešē diferencējami.* Pieņemsim, ka $\mathbf{P}_k$ ir kopa no visiem varbūtības mēriem uz $\mathbf{R}^k$ un ka $\mathbf{S}_k$ ir kopa no visiem galīgiem zīmju mēriem uz $\mathbf{R}^k$, $k \geq 1$. Ar $\|\cdot\|_{(k)}$ apzīmēsim normu $\mathbf{S}_k$ un $\|\cdot\|$ apzīmēsim Eiklīda normu $\mathbf{R}^k$. Tad teiksim, ka funkcionālis $T : \mathbf{P}_k \to \mathbf{R}^s$ ir Frešē diferencējams punktā $F \in \mathbf{P}_k$ pie normas $\|\cdot\|_{(k)}$, ja eksistē funkcija $T^{(1)}(F,\cdot) : \mathbf{S}_k \to \mathbf{R}^s$ tāda, ka

- $T^{(1)}(F, a\nu_1 + b\nu_2) = aT^{(1)}(F,\nu_1) + bT^{(1)}(F,\nu_2)$ visiem $a, b \in \mathbf{R}$ un $\nu_1, \nu_2 \in \mathbf{S}_k$;
- $G \in \mathbf{P}_k$,
$$\frac{\|T(G) - T(F) - T^{(1)}(F, G-F)\|}{\|G-F\|_{(k)}} \to 0,$$
kad $\|G - F\|_{(k)} \to 0$.

Lineārs funkcionālis $T^{(1)}(F,\cdot)$ tiek saukts par T Frešē atvasinājumu punktā F. Šī Frešē diferencējamības definīcija ir atšķirīga no standarta funkcionālanalīzes definīcijas, kur pieņem, ka funkcionālis T ir definēts *normētā vektoru telpā*, kā $\mathbf{S}_k$, nekā tikai $\mathbf{P}_k$. Tomēr, dotā definīcija ir piemērota mūsu gadījumam, jo mēs esam ieinteresēti apskatīt funkcionāļa vērtības tikai pie varbūtību mēriem.

Pieņemsim, ka parametrs θ un tā novērtējums $\hat{\theta}_n$ ir doti, kā (37) un (38), tas ir, $\theta = T(F^{(m)})$ un $\hat{\theta}_n = T(F_n^{(m)})$, kādam funkcionālim $T : \mathbf{P}_{dm} \to \mathbf{R}^s$, kādiem $m, s \in \mathbf{N}$. Ja

T ir Frešē diferencējams pie $F^{(m)}$ ar Frešē atvasinājumu $T^{(1)}(F^{(m)}, \cdot)$, tad pēc $T^{(1)}(F^{(m)}, \cdot)$ linearitātes īpašības

$$
\begin{aligned}
\hat{\theta}_n - \theta &= T(F_n^{(m)} - T(F^{(m)})) \\
&= T^{(1)}(F^{(m)}, F_n^{(m)} - F^{(m)}) + R_n \\
&= n_0^{-1} \sum_{i=1}^{n_0} T^{(1)}(F^{(m)}, \delta_{Y_i} - F^{(m)}) + R_n \qquad (40) \\
&\equiv n_0^{-1} \sum_{i=1}^{n_0} h(Y_i) + R_n,
\end{aligned}
$$

kur $n_0 = (n - m + 1)$, $h(y) = T^{(1)}(F^{(m)}, \delta_y - F^{(m)})$, $y \in \mathbf{R}^{dm}$ un R_n ir atlikuma loceklis, kurš apmierina $\|R_n\| = o(\|F_n^{(m)} - F^{(m)}\|_{(dm)})$, kad $\|F_n^{(m)} - F^{(m)}\|_{dm} \to 0$. Tad no (40) varam iegūt $\sqrt{n}(\hat{\theta}_n - \theta)$ asimptotisko sadalījumu, paredzot, ka $R_n = o_p(n^{-1/2})$. Tālākie nosacījumi izpildīsies, ja $\sqrt{n}\|F_n^{(m)} - F^{(m)}\|_{(dm)}$ ir stohastiski ierobežots pie normas $\|\cdot\|_{(dm)}$. Ņemsim $\|\cdot\|_{(dm)}$, kā Kolmogorova pustelpas normu, definētu

$$\|\nu\|_\infty = \sup\{|\nu((-\infty, x])| : x \in \mathbf{R}^{\mathbf{dm}}\}, \ \nu \in \mathbf{S}_{dm}.$$

Teorēma 13. *Pieņemsim, ka $T(\cdot)$ ir Frešē diferencējams pie $F^{(m)}$ normā $\|\cdot\|_\infty$ un ka $E\|h(Y_1)\|^3 < \infty$, $Eh(Y_1) = 0$ un $\Sigma_\infty^{(m)} \equiv \lim_{n\to\infty} D(n^{-1/2}\sum_{i=1}^n h(Y_i))$ ir nesingulāra. Pie tam, Teorēmas 12 nosacījumi izpildās. Tad*

- $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \Sigma_\infty^{(m)})$, *kad* $n \to \infty$,
- *Ja* $\theta_n^* = T(F_n^{(m)*})$ *un* $\tilde{\theta}_n = T(E_* F_n^{(m)*})$, *kur* $F_n^{(m)*}$ *ir empīriskais sadalījums MBB izlasei, balstītai uz* $b_0 = \lfloor n_0/l \rfloor$ *pārkārtotiem blokiem ar garumu* l. *Tad*

$$\sup_{x \in \mathbf{R}^s} \left| P_*\left(\sqrt{n_0}(\theta_n^* - \tilde{\theta}_n) \le x\right) - P\left(\sqrt{n}(\hat{\theta}_n - \theta) \le x\right)\right| \xrightarrow{p} 0, \ \text{kad } n \to \infty. \qquad (41)$$

Pierādījumu un vājākus nosacījumus, kad šī teorēma arī izpildās, skatīt [7].

Tātad MBB aproksimācija $\sqrt{n}(\hat{\theta}_n - \theta)$ pārkārtošanas sadalījumam ir konsistenta pie Teorēmas 13 nosacījumiem.

7. Bloku butstrapa metožu salīdzinājums

7.1. Teorētiskie modeļi

Pieņemsim, ka $\{X_i\}_{i\in\mathbf{Z}}$ ir $\mathbf{R}^d$ vērtību stacionārs process ar $EX_1 = \mu$. Teorētiskajā daļā strādāsim ar Gludas funkcijas modeli, kas tika definēts 6.1. nodaļā. Pieņemsim, ka parametrs θ un tā novērtējums $\hat{\theta}$ var tikt izteikts, kā gluda funkcija no populācijas un izlases vidējās vērtības, μ un $\bar{X}_n = \sum_{i=1}^{n} X_i$. Kā jau iepriekš parādījām, tad šādā veidā var definēt plašu novērtējumu klasi. Piemēram, izlases kovariāciju novērtējumus, autokorelāciju novērtējumus un Jula-Volkera novērtējumus autoregresīviem procesiem. Tālāk, pieņemsim, ka eksistē gluda funkcija $H : \mathbf{R}^d \to \mathbf{R}$ tāda, ka $\theta = H(\mu)$ un $\hat{\theta}_n = H(\bar{X}_n)$. φ_n ir parametrs, kas ir funkcionālis no $\hat{\theta}_n$ sadalījuma. Pie tam $\hat{\varphi}_n(j,l)$ ir φ_n butstrapa novērtējums, lietojot j-to bloku butstrapa metodi ar (sagaidāmo) bloku garumu l, $j = 1, 2, 3, 4$. Skaidrs, ka $\hat{\varphi}_n(\cdot, l)$ precizitāte ir atkarīga no $\hat{\theta}_n$ sadalījuma funkcionāļa φ_n. Pievērsīsim uzmanību diviem parametriem

$$\varphi_{1n} \equiv \text{Bias}(\hat{\theta}_n) = E\hat{\theta}_n - \theta, \tag{42}$$

$$\varphi_{2n} \equiv D\hat{\theta}_n = E(\hat{\theta}_n - E\hat{\theta}_n)^2, \tag{43}$$

un salīdzināsim bloku butstrapus, novērtējot šos parametrus, lietojot vidējās kvadrātiskās kļūdas jeb MSE kritēriju. Līdzīgus rezultātus var pierādīt sadalījuma funkcijas butstrapa novērtējumiem un citiem zināmiem funkcionāļiem no $\hat{\theta}_n$ sadalījuma, piemēram, kvantīles, tomēr šajos gadījumos būtu nepieciešami citi nosacījumi un argumenti.

Ar $X_{j,1}^*, X_{j,2}^*, \ldots$ apzīmēsim butstrapa izlases, iegūtas ar j-to metodi. Dotai vērtībai l, pieņemsim, ka $b = \lfloor n/l \rfloor$ bloki ir pārkārtoti metodēm MBB, NBB un CBB, iegūstot butstrapa izlases ar apjomu $n_1 \equiv bl$. Tad butstrapa versija, centrētam lielumam $T_n = \hat{\theta}_n - \theta$ MBB, NBB un CBB metodēm ir

$$T_{n,l}^{*(j)} = H(\bar{X}_{n,l}^{*(j)}) - H(E_*\bar{X}_{n,l}^{*(j)}),\ j = 1, 2, 3,$$

kur $\bar{X}_{n,l}^{*(j)} = n_1^{-1} \sum_{i=1}^{n_1} X_{j,i}^*$.

Tālāk parādīsim atbilstošus rezultātus SB metodei. Tā kā l ir sagaidāmais bloka garums, tad bloku garumi $L_1, L_2, \ldots$ ir neatkarīgi un vienādi sadalīti gadījuma lielumi ar ģeometrisko sadalījumu ar parametru $p = 1/l$. Dotam l, $K = \inf\{1 \le k \le n, L_1 + \ldots + L_k \ge n\}$ bloki tiek pārkārtoti SB metodē, iegūstot izlasi ar apjomu $N_1 = L_1 + \ldots + L_k$. Pieņemsim, ka $\bar{X}_{n,l}^{*(4)} \equiv n^{-1} \sum_{i=1}^{n} X_{j,i}^*$ ir vidējā vērtība no pirmajiem n novērojumiem. Kā parādīts iepriekš, $E_* \bar{X}^{*(4)}_{n,l} = \bar{X}_n$ visiem l. SB metodes $T_n = \hat{\theta}_n - \theta$ butstrapa versija ir

$$T^{*(4)}_{n,l} = H(\bar{X}^{*(4)}_{n,l}) - H(\bar{X}_n).$$

Piezīmēsim, ka parametri (42) un (43) ir pirmie divi T_n momenti, tas ir, $\varphi_{1n} = \text{Bias}(\hat{\theta}_n) = ET_n$ un $\varphi_{2n} = D\hat{\theta}_n = DT_n$. Tādejādi, φ_{1n} un φ_{2n} butstrapa novērtējumi ir definēti

$$\begin{aligned} \hat{\varphi}_{1n}(j,l) &\equiv \widehat{\text{BIAS}}_j(l) = E_* T^{*(j)}_{n,l}, \ j = 1,2,3,4, \\ \hat{\varphi}_{2n}(j,l) &\equiv \widehat{\text{VAR}}_j(l) = D_* T^{*(j)}_{n,l}, \ j = 1,2,3,4. \end{aligned}$$

Nākošajā nodaļā parādīsim MSE izvirzījumus bloku butstrapa novērtējumiem $\widehat{\text{BIAS}}_j(l)$ un $\widehat{\text{VAR}}_j(l), \ j = 1,2,3,4$.

7.2. MSE izvirzījumi

Lai iegūtu butstrapa MSE izvirzījumus, piezīmēsim, ka gadījuma lielumam Y, $\text{MSE}(Y) = (\text{Bias}(Y))^2 + DY$, tātad MSE varam iegūt no novirzes un dispersijas daļas. Turpmāk apskatīsim novirzes un dispersijas daļas atsevišķi un kombinēsim, lai iegūtu kopējo MSE.

Pieņēmums 1. Funkcija $H : \mathbf{R}^d \to \mathbf{R}$ ir r reizes nepārtraukti diferencējama un $\max\{|D^\nu H(x)| : |\nu| = r\} \le C(1 + \|x\|^{a_0}), \ x \in \mathbf{R}^d$ kādam veselam skaitlim $a_0 \ge 1$.

Pieņēmums 2. $E\|X_1\|^{2r+\delta} < \infty$ un $\Delta(r,\delta) < \infty$ kādam $\delta > 0$.

Tad $\hat{\varphi}_{1n}(j,l)$ un $\hat{\varphi}_{2n}(j,l), \ j = 1,2,3,4$ novirzes daļai mums ir sekojošs rezultāts.

Teorēma 14. *Pieņemsim, ka l ir tāds, ka $l^{-1} + n^{-1/2} l = o(1)$, kad $n \to \infty$.*

1. *Ja Pieņēmums 1 ir spēkā ar $r = 3$ un Pieņēmums 2 ir spēkā ar $r = 3 + a_0$, kur a_0 ir tāds, kā Pieņēmumā 1, tad*

$$\text{Bias}(\widehat{BIAS}_j(l)) = n^{-1} l^{-1} A_1 + o(n^{-1} l^{-1}), \ j = 1,2,3,4,$$

   kur $A_1 = -\sum_{|\alpha|=1} \sum_{|\beta|=1} c_{\alpha+\beta} (\sum_{j=-\infty}^{\infty} |j| E X_1^\alpha X_{1+j}^\beta)$ un $c_\alpha = D^\alpha H(\mu)/\alpha!, \ \alpha \in \mathbf{Z}^d_+$.

2. *Ja Pieņēmums 1 ir spēkā ar $r = 2$ un Pieņēmums 2 ir spēkā ar $r = 4 + 2a_0$, kur a_0 ir tāds, kā Pieņēmumā 1, tad*

$$\text{Bias}(\widehat{VAR}_j(l)) = n^{-1} l^{-1} A_2 + o(n^{-1} l^{-1}), \ j = 1,2,3,4,$$

   kur $A_2 = -\sum_{j=-\infty}^{\infty} |j| E\tilde{Z}_1 \tilde{Z}_{1+j}, \ \tilde{Z}_i = \sum_{|\alpha|=1} c_\alpha (X_i - \mu)^\alpha, \ i \ge 1$ un c_α ir tāds, kā šīs teorēmas 1. daļā.

Pierādījumu skatīt [7].

Tātad no Teorēmas 14 seko, ka φ_{1n} un φ_{2n} butstrapa novērtējumiem novirzes visām četrām bloku butstrapa metodēm ir vienādas līdz konstantēm. Pretēji gaidītajam, SB metode nesamazina butstrapa novērtējumu novirzi, pie tam, asimptotiski novirzes ir vienādas, neatkarīgi no šķeļošu vai nešķeļošu bloku izmantošanas. Tā kā novirze bloku butstrapu novērtējumiem rodas no sākotnējās izlases $X_1, \ldots, X_n$ aizvietošanas ar neatkarīgām blokiem, tad visas metodes dod līdzīgus rezultātus, kamēr bloku garumi l ir asimptotiski vienādi.

Tālāk salīdzināsim φ_{1n} un φ_{2n} dispersijas.

Teorēma 15. *Pieņemsim, ka Teorēmas 14 nosacījumi par bloku garumu l un Pieņēmumu 1 un 2 nosacījumi, attiecīgajām teorēmas daļām, ir spēkā.*

1. *Tad eksistē simetriskas, nenegatīvas reālu vērtību funkcijas g_1, g_2 tādas, ka*

$$\begin{aligned} D(\widehat{BIAS}_j(l)) &= \{4\pi^2 g_1(0)/3\}n^{-3}l + o(n^{-3}l), \ j = 1,3, \\ D(\widehat{BIAS}_j(l)) &= \{2\pi^2 g_1(0)\}n^{-3}l + o(n^{-3}l), \ j = 2 \end{aligned}$$

2. *un*

$$\begin{aligned} D(\widehat{VAR}_j(l)) &= \{4\pi^2 g_2(0)/3\}n^{-3}l + o(n^{-3}l), \ j = 1,3, \\ D(\widehat{VAR}_j(l)) &= \{2\pi^2 g_2(0)\}n^{-3}l + o(n^{-3}l), \ j = 2. \end{aligned}$$

Funkcijas definīciju $g_k(\cdot)$, $k = 1,2$ un teorēmas pierādījumu skatīt [7].

Tomēr arī bez detalizētas $g_k(\cdot)$, $k = 1,2$ definīcijas, varam salīdzināt dispersijas daļas dažādām bloku butstrapa metodēm. No Teorēmas 15 1. un 2. daļas redzams, ka MBB un CBB $\varphi_{1n} = \mathrm{Bias}(\hat{\theta}_n)$ un $\varphi_{2n} = D\hat{\theta}_n$ novērtējumiem ir 2/3 reizes mazākas dispersijas kā NBB novērtējumam. Tā kā MBB un CBB metodēs bloki pārklājas, tad variācija starp pārkārtotajiem blokiem ir mazāka, kas dod mazāku dispersiju šiem novērtējumiem. Šo MBB priekšrocību pār NBB pirmais parādīja Kunsch ([5]). SB metodes gadījumā, Nordman ([15]) parādīja, ka [7] dotais rezultāts nav korekts. Nordman parādīja, ka asimptotiski SB un NBB metožu rezultātu precizitāte ir vienāda. Tāpēc labosim [7] rezultātus, tas ir, atstāsim nepublicētus SB metodes rezultātus no [7], bet teiksim, ka SB rezultāti sakrīt ar NBB. Tātad pretēji Lahiri paustajam, ka SB metodei novērtējumu dispersija ir asimptotiski vislielākā starp četrām metodēm, Nordman parādīja, ka SB novērtējumu dispersija un NBB novērtējumu dispersija ir asimptotiski vienādas.

7.3. Teorētiskie salīdzinājumi

7.3.1. Asimptotiskā efektivitāte

Šajā nodaļā apkoposim Teorēmu 14 un 15 rezultātus. Divām novērtējumu virknēm $\{\hat{\psi}_{1n}\}_{n\geq 1}$ un $\{\hat{\psi}_{2n}\}_{n\geq 1}$ definēsim *asimptotiski relatīvo efektivitāti*

$$\mathrm{ARE}(\hat{\psi}_{1n}, \hat{\psi}_{2n}) = \lim_{n\to\infty} \Big(\mathrm{MSE}(\hat{\psi}_{2n})/\mathrm{MSE}(\hat{\psi}_{1n})\Big).$$

Tātad, ja $\mathrm{ARE}(\hat{\psi}_{1n}, \hat{\psi}_{2n}) < 1$, tad virkne $\{\hat{\psi}_{1n}\}_{n\geq 1}$ ir mazāk *efektīva*, kā $\{\hat{\psi}_{2n}\}_{n\geq 1}$, proti, $\{\hat{\psi}_{1n}\}_{n\geq 1}$ novērtējumiem ir lielāka MSE, nekā $\{\hat{\psi}_{2n}\}_{n\geq 1}$ pie lieliem n.

Teorēma 16. *Pieņemsim, ka Teorēmu 14 un 15 nosacījumi izpildās un ka* $A_k \neq 0$, $g_k \neq 0$, $k = 1, 2$.

1. $l^{-1} + n^{-1/3} l = o(1)$, *katram* $i, j = 1, 2, 3$, $k = 1, 2$,

$$\mathrm{ARE}\left(\hat{\varphi}_{kn}(j, l), \hat{\varphi}_{kn}(i, l)\right) = 1;$$

2. $l^{-1} n^{1/3} = o(1)$, $k = 1, 2$,

$$\mathrm{ARE}\left(\hat{\varphi}_{kn}(2, l), \hat{\varphi}_{kn}(j, l)\right) = 2/3, \; j = 1, 3;$$

3. $l = Cn^{1/3}(1 + o(1))$, $C \in (0, \infty)$, $k = 1, 2$,

$$\mathrm{ARE}\left(\hat{\varphi}_{kn}(2, l), \hat{\varphi}_{kn}(j, l)\right) = \frac{3 + 4\pi^2 C^3 A_k^{-2} g_k(0)}{3 + 6\pi^2 C^3 A_k^{-2} k_k(0)} \in (2/3, 1), \; j = 1, 3.$$

Teorēmas pierādījums ir atrodams [7].

Šajā gadījumā atkal labosim [7] rezultātus SB metodei, atstājot tos nepublicētus, bet teiksim, ka NBB metodes rezultāti ir asimptotiski vienādi ar SB metodes rezultātiem, atsaucoties uz [15]. Teorēmas 16 1. un 2. daļas attiecas uz gadījumiem, attiecīgi, kad l vērtība ir maza un MSE, galvenokārt, nosaka novirzes daļa (1. daļa) un kad l vērtība ir liela un MSE, galvenokārt, nosaka dispersijas daļa (2. daļa).

7.3.2. Optimālais bloku garums

No Teorēmām 14 un 15 redzams, ka katrai no butstrapa metodēm, palielinot l, bloku butstrapa novērtējuma novirze samazinās, kamēr to dispersija palielinās. Tā rezultātā, eksistē kritiskā bloka garuma vērtība l, pie kuras MSE sasniedz savu minimumu. Šo minimumu sauksim par MSE-optimālo bloku garumu. Apzīmēsim ar

$$\begin{aligned} l_{1j}^0 &= \operatorname{argmin}\{\mathrm{MSE}(\widehat{\mathrm{BIAS}}_j(l)) : n^\epsilon \leq l \leq n^{(1-\epsilon)/2}\}, \\ l_{2j}^0 &= \operatorname{argmin}\{\mathrm{MSE}(\widehat{\mathrm{VAR}}_j(l)) : n^\epsilon \leq l \leq n^{(1-\epsilon)/2}\}, \end{aligned}$$

$1 \leq j \leq 4$ MSE-optimālā bloku garumus, novērtētajā $\hat{\theta}_n$ novirzei un dispersijai, kur $\epsilon \in (0, 1/3)$ ir dots. Sekojošais rezultāts dod optimālo l^0_{kj}, $k = 1, 2$, $j = 1, 2, 3, 4$, novērtējot φ_{1n} un φ_{2n}.

Teorēma 17. *Pieņemsim, ka Teorēmas 16 nosacījumi ir spēkā. Ja $k = 1, 2$, tad*

$$\begin{aligned} l^0_{kj} &\sim (3A_k^2/(2\pi^2 g_k(0)))^{1/3} n^{1/3},\ j = 1, 3, \\ l^0_{kj} &\sim (A_k^2/(\pi^2 g_k(0)))^{1/3} n^{1/3},\ j = 2. \end{aligned}$$

Teorēmas pierādījums ir atrodams [7].

Nepublicēsim SB rezultātus no [7], bet teiksim, ka tie ir vienādi ar NBB metodes rezultātiem. No teorēmas redzams, ka optimālais bloku garums MBB metodei ir $(3/2)^{1/3}$ reizes lielāks kā NBB metodei.

Pēc l^0_{kl} definīcijas ir skaidrs, ka katra bloku butstrapa metode dod vislabākos parametra φ_{kn} novērtējuma rezultātus, ja tā lietota ar optimālo bloka garumu. Tālāk salīdzināsim bloku butstrapa metodes, ja tās lietotas ar optimālo bloku garumu.

Teorēma 18. *Pieņemsim, ka Teorēmas 16 nosacījumi ir spēkā.*

- *Ja $k = 1, 2$, tad*

$$\begin{aligned} \text{MSE}(\hat{\varphi}_{kn}, l^0_{kj}) &= 3^{1/3}(2\pi^2 g_k(0) A_k)^{2/3} n^{-8/3} + o(n^{-8/3}),\ j = 1, 3; \\ \text{MSE}(\hat{\varphi}_{kn}, l^0_{kj}) &= 3(\pi^2 g_k(0) A_k)^{2/3} n^{-8/3} + o(n^{-8/3}),\ j = 2 \end{aligned}$$

- *un*

$$\text{ARE}\left(\hat{\varphi}_{kn}(2, l^0_{k2}), \hat{\varphi}_{kn}(j, l^0_{kj})\right) = (2/3)^{2/3},\ j = 1, 3.$$

Teorēmas pierādījums ir atrodams [7].

Teiksim, ka SB metodei ir vienādi rezultāti ar NBB metodi. Teorēma 18 rāda, ka, ja katra no metodēm ir veikta ar optimālo bloka garumu, tad MBB un CBB optimālā MSE ir $(2/3)^{2/3}$ reizes mazāka kā optimālā MSE NBB metodei.

7.4. Datu simulācijas

Iepriekšējās nodaļās parādījām teorētiskus bloku butstrapa metožu salīdzinājumus. Šajā nodaļā veiksim praktisku metožu salīdzināšanu ar simulētiem datiem.

Piemērs 7. Izvēlēsimies ARIMA(1,1) modeli

$$X_i = 0.6X_{i-1} + \epsilon_i - 0.1\epsilon_{i-1}, \tag{44}$$

kur ϵ_i ir neatkarīgi un vienādi sadalīti ar $N(0,1)$ sadalījumu. Apskatīsim parametru $\varphi_n = nD\bar{X}_n$. Mēģināsim novērtēt šī parametra MSE dažādām bloku butstrapu metodēm ar dažādiem bloku garumiem l. Butstrapotā statistika būs formā

$$T_n^* = \sqrt{n}(\bar{X}_n^* - E_*\bar{X}_n^*), \tag{45}$$

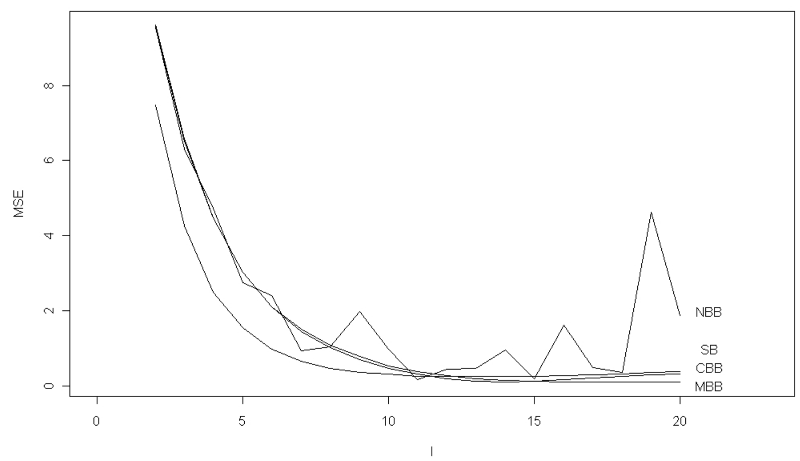
kur centrējošās vērtības $E_*\bar{X}_n^*$ atradīsim pēc (10) un (11) metodēm MBB un NBB, bet CBB un SBB $E_*\bar{X}_n^* = \bar{X}_n$. Apskatīsim vidējo kvadrātisko kļūdu bloku garumiem $l = 2, \ldots, 20$ ar soli 1. Simulētos datus ņemsim ar apjomu $n = 400$. Katrā no soļiem veiksim 2000 simulācijas, kur katrā no simulācijām veidosim 500 butstrapa izlases. Pieņemsim, ka

$$T_{i,j,n}^* = \sqrt{n}(\bar{X}_{i,j,n}^* - E_*\bar{X}_n^*),$$

ir j-tās butstrapa izlases realizācija i-tai simulācijai. MSE novērtējums ir

$$\begin{aligned}\widehat{\mathrm{MSE}}(l) &= (2000)^{-1}\sum_{i=1}^{2000}(\hat{\varphi}_n - \varphi_n)^2 \\ &= (2000)^{-1}\sum_{i=1}^{2000}\left((500)^{-1}\sum_{j=1}^{500}(T_{i,j,n}^*)^2 - \varphi_n\right)^2,\end{aligned}$$

kur teorētiskais $\varphi_n = nD\bar{X}_n = DX_1 + 2\sum_{i=1}^{n-1}\mathrm{Cov}(X_1, X_{1+i})$. Kovariācijas iegūsim no (4), atmetot tās, kuras mazākās par 0.001. Tad šajā gadījumā iegūsim $\varphi_n = 5.153$, ko izmantosim MSE aprēķinam. Iegūtie rezultāti parādīti 1. attēlā.



1. att. Parametra $\varphi_n = nD\bar{X}_n$ MSE novērtējumi modelim (44).

Redzams, ka MBB, NBB un CBB metodēm ir vienādi rezultāti pie maziem bloku garumiem l, kas arī ir sagaidāms no teorijas. Tātad no Teorēmas 14 zinām, ka novirzes daļas visām metodēm ir vienādas un ka MSE pie maziem l tiek, galvenokārt, noteiks ar novirzes daļu. SB metodes uzvedība pie maziem l ir skaidrojama ar to, ka šī metode ir mazāk jūtīga pret bloku garumu izvēli nekā pārējās trīs metodes. Tāpēc, iespējams, SB metode šajā piemērā darbojas labāk ar maziem bloku garumiem. Pie lielākiem bloku garumiem NBB metode kļūst nestabila, kas izskaidrojams ar sākotnējās simulētās izlases *īpatnībām* un to, ka NBB bloku izvēles skaits ir mazs. Pie $l > 20$ mēs sagaidītu, ka MBB un CBB metodes darbojas labāk nekā NBB un SB (skatīt Teorēmas 15 un 16), kas daļēji manāms jau pie $l = 20$.

No šī piemēra varam secināt par modeļa (44) ar $n = 400$ optimālo bloku garumu parametra φ_n novērtēšanai. Pēc MSE grafikiem SB metodes optimālais bloku garums ir aptuveni $l = 10$, bet pārējām metodēm $l = 15$. Šo rezultātu varam salīdzināt ar metodi, ko piedāvāja Politis, White un Patton ([16] un [17]). Autori deva R programmu, lai noteiktu optimālu bloku garumu MBB un SB metodēm. Programma atrodama `http://www.economics.ox.ac.uk/members/andrew.patton/code.html`. Iegūsim, ka MBB metodes optimālais bloka garums ir 11, bet SB metodes optimālais bloka garums ir 10, kas sakrīt ar mūsu atrasto optimālo bloka garumu. Protams, MSE izpratnē iegūt optimālo bloka garumu ir iespējams gadījumos, kad teorētiskā parametra vērtība ir zināma, kas praksē ir reti sastopams. Literatūrā eksistē metodes, piemēram, jau minētā Politis, White un Patton metode, kuras dod optimālo bloka garumu bez teorētiskās vērtības zināšanas. Citas metodes skatīt [7] un [18].

Vēl piebilstams, ka gadījumā, kad $\varphi_n = nD\bar{X}_n$, ir iespējams MSE novērtējumus iegūt bez butstrapa procedūras, jo katras metodes vidējās vērtības dispersijas novērtējumi pie katra l ir iegūstami no (22) un (24).

Piemēra ideja ņemta no [7] un atbilstošais programmas R kods atrodams pielikumā.

Iepriekšējo nodaļu rezultāti rāda, ka MSE izpratnē MBB un CBB novērtējumi ir precīzāki, kā NBB un SB novērtējumi. Šis apgalvojums ir spēkā, kamēr bloku garums aug ar ātrumu ne zemāku, kā optimālais bloku garums. Lai novērtētu funkcionāļu φ_{1n} un φ_{2n} novirzi un dispersiju, šis optimālais ātrums ir $Cn^{1/3}$, kur n ir izlases apjoms. Gadījumā, ja bloka garums aug ar zemāku ātrumu nekā optimālais ātrums, tad MSE galvenokārt nosaka novirzes daļa. Šajā gadījumā starp visām metodēm ARE ir 1. Tās ir sekas no fakta, ka asimptotiski butstrapa novērtējumu novirzes visām metodēm ir vienādas. Tātad, ja bloka garums ir tuvs optimālajam bloku garumam, tad MBB un CBB metodes dod visprecīzākos rezultātus butstrapa novērtējumiem.

Ir iespējams salīdzināt bloku butstrapa metodes daudz komplicētākiem $\hat{\theta}_n$ sadalījuma funkcionāļiem. Lai novērtētu sadalījuma funkciju un kvantīles $\hat{\theta}_n$ stjūdentizētai versijai, optimālais bloka garums ir formā $Cn^{1/4}$ visām četrām bloku butstrapa metodēm. Arī šajā gadījumā precīzākās izrādās MBB un CBB metodes.

Iepriekšējie rezultāti parāda salīdzinājumu starp četrām bloku butstrapa metodēm. Carlstein, Do, Hall, Hesterberg un Kunsch ([19]) piedāvāja saskaņoto bloku butstrapa metodi (no angļu valodas ,,matched block bootstrap" (MaBB)), kur bloki tiek pārkārtoti, izmantojot Markova ķēdes. Pretēji zināmajām četrām metodēm, MaBB pārkārtotie bloki ir atkarīgi. Pie zināmiem procesa strukturāliem nosacījumiem, tas ir, AR(p) vai Markova, Carlstein parādīja, ka MaBB izlases vidējās vērtības dispersijas novērtējumam dispersija ir salīdzināma ar NBB dispersijas novērtējumu, bet novirze ir ar mazāku kārtu. Tātad MaBB MSE ir ar mazāku kārtu, kā četrām apskatītajām metodēm. Rezultātā, procesos ar Markova struktūru MaBB darbojas labāk pie optimālajiem bloku garumiem. Vispārīgākām laikrindām, kurām nav obligāta Markova struktūra, Paparoditis un Politis ([20]) piedāvāja konusveida bloku butstrapu (no angļu valodas ,,tapered block bootstrap" (TaBB)) un parādīja, ka TaBB darbojas labāk kā MBB un CBB.

8. Butstrapa pielietojums neparametriskajā regresijā

Neparametriskā regresija ir plaši pielietota metode statistikā, kur nav nepieciešama parametriska modeļa izvēle. Metodes galvenā ideja ir novērtēt $m(x) = E(Y|X = x)$, kur X var būt gan viendimensionāls, gan vairāk dimensionāls lielums. Šajā nodaļā dosim īsu ieskatu regresijas konstruēšanā un divās gludināšanas parametra izvēles metodēs. Apskatīsim regresijas atlikumu struktūru atkarībā no regresijas modeļa. Tālāk parādīsim ar piemēru, ka īpašos gadījumos bieži lietotas metodes, gludināšanas parametra noteikšanai, dod nepareizus rezultātus.

8.1. Neparametriskās regresijas konstrukcija

Apskatīsim modeli $Y_i = m(X_i) + \epsilon_i$, kur $\{\epsilon_i\}$ ir stacionāra virkne ar vidējo vērtību 0. Parastā pieeja, novērtējot $m(x)$, ir lietot kodolu gludināšanu un izvēlēties atbilstošu svaru funkciju ω_i. Ņemsim

$$\hat{m}(x) = \sum_{i=1}^{n} \omega_i(x) Y_i,$$

kā $m(x)$ novērtējumu, kur n ir dotās izlases apjoms. Viena no pieejām ir ω_i ņemt, kā Nadaraya-Watson kodolu novērtējumus

$$\omega_i(x) = K\left((x - X_i)/h\right) \left(\sum_{j=1}^{n} K\left((x - X_i)/h\right) \right)^{-1},$$

kur $K(x)$ ir kodola funkcija ar īpašībām $\int K(x)dx = 1$, $\int xK(x)dx = 0$, $\int x^2 K(x)dx > 0$, pie tam $K(x)$ ir gluda. h ir gludināšanas parametrs.

Otra pieeja ir lietot lokālās lineārās regresijas gludinātāju, ko piedāvāja Fan (1993),

$$\omega_i(x) = \upsilon_i(x) \left(\sum_{j=1}^{n} \upsilon_i(x) + n^{-2} \right)^{-1},$$

kur

$$\begin{aligned} \upsilon_i(x) &= K\left((x - X_i)/h\right)\left(s_2 - (x - X_i)s_1\right), \\ s_k &= \sum_{j=1}^{n} K\left((x - X_j)/h\right)(x - X_j)^k, \ k = 1, 2. \end{aligned}$$

Abu metožu galvenā problemātika ir parametra h izvēle. Kodolu K izvēle nav izšķiroša. $\hat{m}(x)$ atbilstības kritēriji ir

$$\mathrm{MSE}(x) = E\left(\hat{m}(x) - m(x)\right)^2 \text{ un } \mathrm{MISE} = \int_X E\left(\hat{m}(x) - m(x)\right)^2 dx.$$

Tālāk aplūkosim divas optimālā gludināšanas parametra izvēles metodes. Viena no piedāvātajām metodēm ir krosvalidācijas metode. Definēsim vispārinātu *atstāt* $2l+1$ *ārā* krosvalidācijas metodi ar lokālajiem lineārajiem gludinātājiem. Mērķis ir minimizēt MISE novērtējumu

$$\widehat{\mathrm{MISE}}(h) = n^{-1} \sum_{j=1}^{n} (\hat{m}_j(X_j) - Y_j)^2$$

ar

$$\hat{m}_j(X_j) = \sum_{|i-j|>l} \omega_{ij}(X_j) Y_i,$$

kur, līdzīgi, kā iepriekš,

$$\begin{aligned} \omega_{ij} &= \upsilon_{ij}(x) \left(\sum_{|k-j|>l} \upsilon_{kj}(x) + n^{-2} \right)^{-1}, \\ \upsilon_{ij}(x) &= K\left((x - X_i)/h\right)\left(s_{2j} - (x - X_i)s_{1j}\right), \\ s_{kj} &= \sum_{|i-j|>l} K\left((X_j - X_i)/h\right)(X_j - X_i)^k,\ k = 1, 2. \end{aligned}$$

Tad $h_{opt} = \operatorname{minarg}\{\widehat{\mathrm{MISE}}(h)\}$. Visbiežāk lietotā ir *atstāt vienu ārā* jeb $l = 0$ krosvalidācija. Šī metode dod labus rezultātus lielākajā daļā no gadījumiem. Teorija par neparametrisko regresiju ņemta no [21].

Apskatīsim otru h izvēles metodi, balstītu uz butstrapu, kas aprakstīta [22]. Pieņemsim, kādu sākuma vērtību h_0. Dažreiz h_0 izvēlas pēc augstāk minētās krosvalidācijas metodes. Iegūsim $\hat{m}_{h_0}(x)$ regresijas novērtējumu ar gludināšanas parametru h_0. Katrā X_i punktā iegūsim atbilstošos atlikumus

$$\epsilon_i = Y_i - \hat{m}_{h_0}(X_i),\ i = 1, \dots, n.$$

Šos atlikumus centrēsim

$$\tilde{\epsilon}_i = \epsilon_i - n^{-1} \sum_{i=1}^{n} \epsilon_i,\ i = 1, \dots, n$$

un normēsim

$$\hat{\epsilon}_i = \tilde{\epsilon}_i / (1 - K(0)/nh),\ i = 1, \dots, n,$$

kur $K(x)$ ir kodola funkcija. Tālāk butstraposim iegūtos $\hat{\epsilon}_1, \dots, \hat{\epsilon}_n$ un iegūsim butstrapa izlasi $\epsilon_1^*, \dots, \epsilon_n^*$. No butstrapotajiem atlikumiem iegūsim butstrapa izlasi $Y_1^*, \dots, Y_n^*$

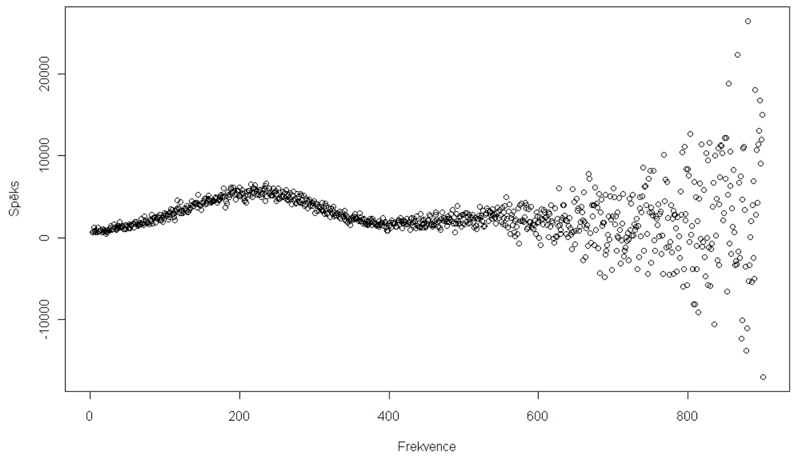
$$Y_i^* = \hat{m}_{h_0}(X_i) + \epsilon_i^*,\ i = 1, \dots, n.$$

Atkārtosim butstrapošanu B reizes un iegūsim B butstrapotās izlases. Katrai butstrapotai izlasei novērtēsim savu $\hat{m}_h^*$. Tad MISE novērtējums ir

$$\widehat{\text{MISE}}(h) = (nB)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{B} \left(\hat{m}_{j,h}^*(X_i) - \hat{m}_{h_0}(X_i) \right), \; i = 1, \ldots .n$$

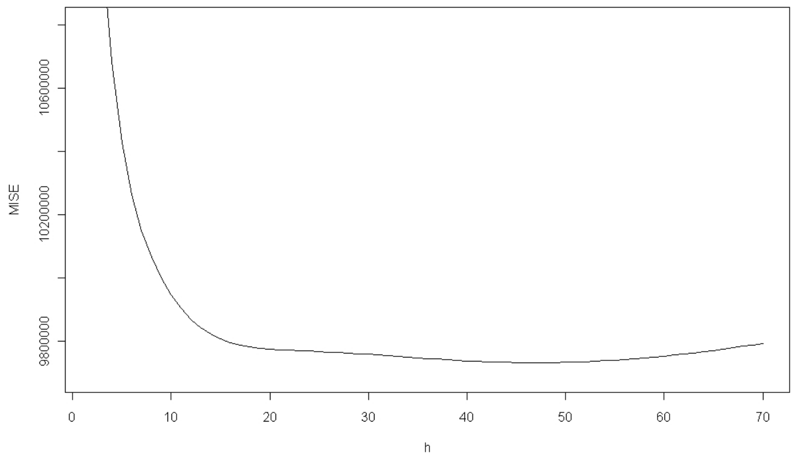
Optimālais $h_{opt} = \text{minarg}\{\widehat{\text{MISE}}(h)\}$. Ir iespējams veidot iteratīvu procedūru, kur h_0 tiek izvēlēts, kā iepriekšējās iterācijas h_{opt}. Kā norādījis Faraway ([22]), tad šī procedūra ir lietojama arī heteroskedastiskiem atlikumiem.

Salīdzināsim šīs metodes praktiski. No [23] iegūsim CMB datus (saīsinājums no angļu valodas ,,cosmic microwave background radiation"). X_i dati attēlo temperatūras fluktuācijas frekvenci un Y_i dati reprezentē fluktuāciju spēku katrā frekvencē. Dati parādīti 2. attēlā.



2. att. CMB dati.

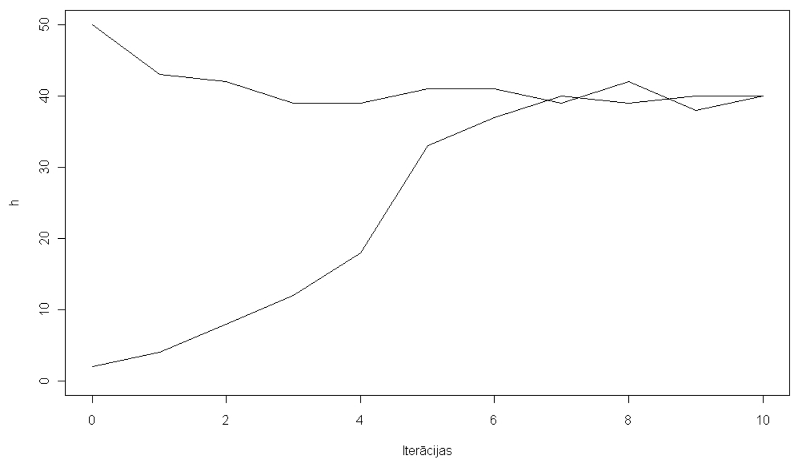
Noteiksim optimālos gludināšanas parametrus ar krosvalidācijas un butstrapa metodēm. Krosvalidācijas *atstāt vienu ārā* metodei meklēsim optimālo h intervālā no 2 līdz 70 ar soli 1. Atradīsim katra soļa novērtēto MISE un izvēlēsimies to h, pie kura $\widehat{\text{MISE}}(h)$ sasniedz savu minimumu.



3. att. CMB datu $\widehat{\mathrm{MISE}}(h)$ krosvalidācijas *atstāt vienu ārā* metodei.

Krosvalidācijas metodes $\widehat{\mathrm{MISE}}(h)$ sasniedz savu minimumu pie $h = 47$. Šis arī ir krosvalidācijas optimālais gludināšanas parametrs.

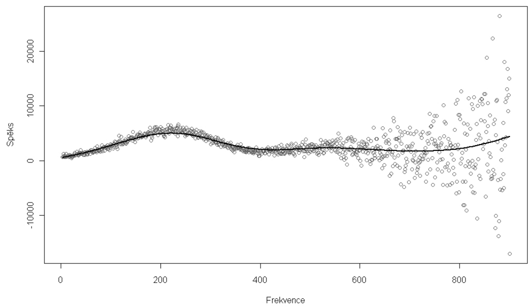
Šiem pašiem datiem aplūkosim butstrapa metodi, optimālā parametra h izvēlei. Veidosim divus iteratīvos procesus, kur pirmajā sākuma $h_0 = 2$, bet otrajā $h_0 = 50$. Veiksim 10 iterāciju soļus. Katrā no iterācijām veidosim 500 butstrapa izlases un h soli ņemsim vienādu ar 1. Rezultāti parādīti 4. attēlā.



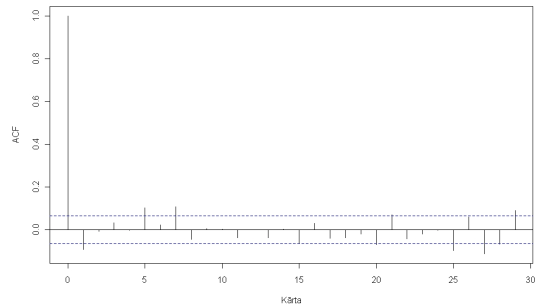
4. att.: CMB datu iterāciju rezultāti butstrapa metodei. Pie iterācijas soļa 0 h vērtības ir vienādas ar sākuma h_0 izvēli.

Kā redzams, tad abas līknes tiecas uz vienu $h = 40$ vērtību. Tātad ņemot jebkuru sākuma vērtību no intervāla $[2, 50]$, iegūsim optimālo gludināšanas parametru 10 iterāciju soļos. Kā redzams, iterāciju rezultāts nav atkarīgs no sākuma h_0 izvēles, taču iterāciju soļu skaitam jābūt *pietiekami* lielam. Ņemot mazāku soli un lielāku butstrapa izlašu skaitu, rezultātus ir iespējams iegūt precīzākus.

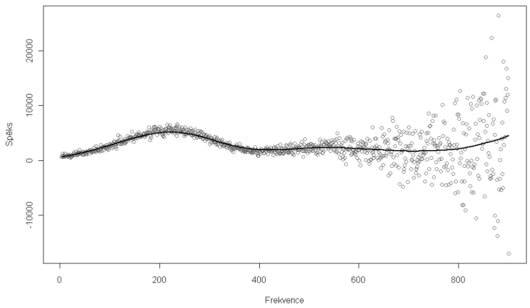
Attēlosim datus ar iegūtajām regresijām un uzzīmēsim atlikumu autokorelāciju grafikus. Šeit un turpmāk izmantosim lokālās lineārās regresijas modeli.



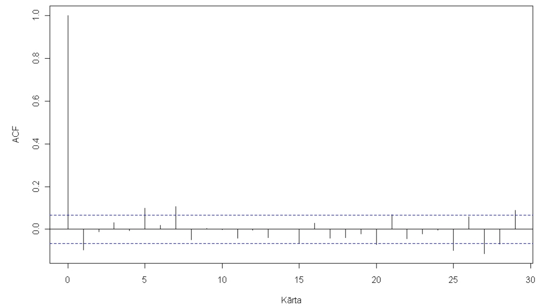
(a)



(b)



(c)



(d)

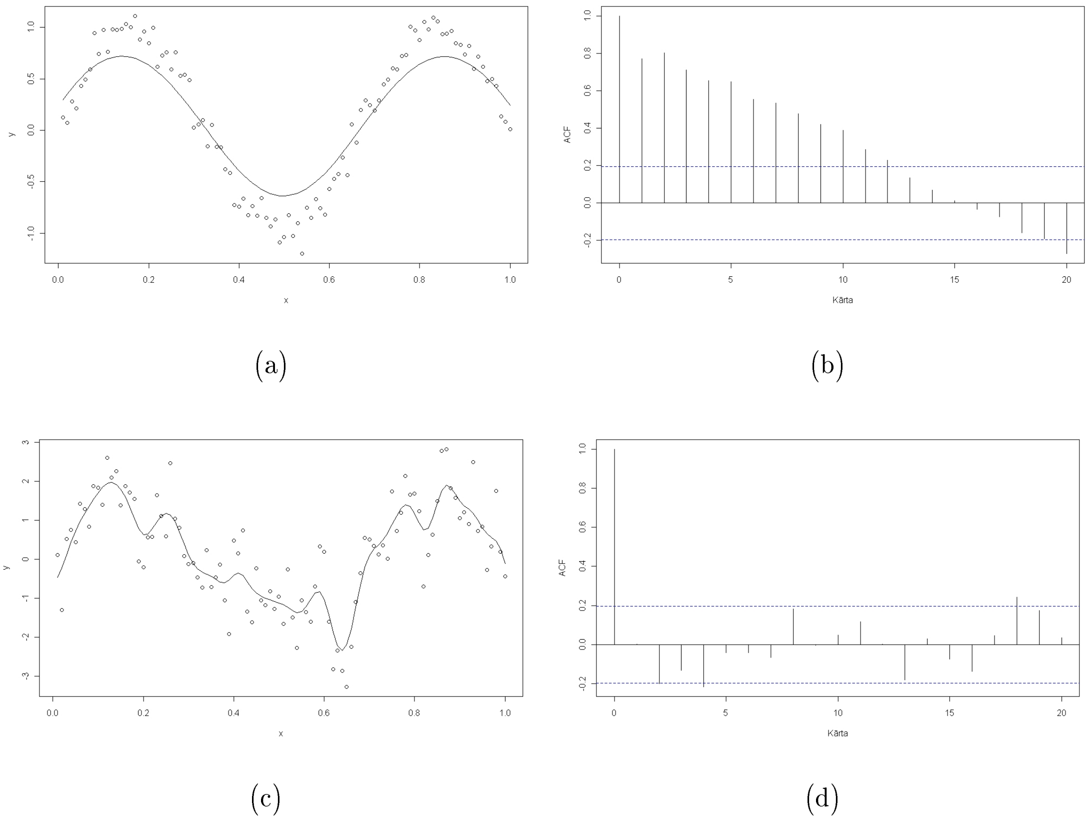
5. att.: CMB dati: (a) ar krosvalidācijas metodi izvēlētais $h = 47$ un (b) atlikumu autokorelāciju grafiks; (c) ar butstrapa metodi izvēlētais $h = 40$ un (d) atlikumu autokorelāciju grafiks.

Abas metodes dod optimālu h izvēli, un regresijas, kas balstītas uz šiem h_{opt}, ir ar nekorelētiem atlikumiem. Pie tam interesanti, ka butstrapa metode strādā arī šajā gadījumā, kad atlikumi ir heteroskedastiski. Būtībā dotā butstrapa procedūra ir konstruēta homoskedastiskiem atlikumiem, bet autors norādījis, ka tai vajadzētu darboties arī heteroskedastisku atlikumu gadījumā.

8.2. Regresijas atlikumu nozīme modeļa izvēlē

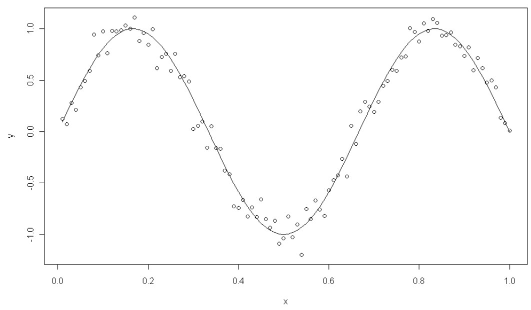
Šajā nodaļā apskatīsim regresijas atlikumus un parādīsim to, ka no atlikumu neatkarības nevar viennozīmīgi noteikt modeļa atbilstību.

Konstruēsim modeli $y_i = \sin(3\pi x_i) + \epsilon_i$ intervālā $x_i \in [0, 1]$ ar $i = 1, \ldots, 100$. Šim modelim izvēlēsimies ϵ_i, kuri ir neatkarīgi un vienādi sadalīti ar vidējo vērtību 0 un dispersiju 0.01. Izmantosim programmas R procedūru `polyloc`, kur gludināšanas parametru izvēlēsimies $h = 0.1$, un konstruēsim novērtēto regresiju. Veidosim otrus ϵ_i tādus, ka $E(\epsilon_i, \epsilon_j) = e^{-70|x_i - x_j|}$, tātad atkarīgus gadījuma lielumus ϵ_i ar vidējo vērtību 0 un dispersiju 1. Šo atkarības struktūru iegūsim ar `GaussRF` procedūru no programmas R. Arī šajā gadījumā lietosim `polyloc`, lai konstruētu regresijas funkciju, pie tam gludināšanas parametru izvēlēsimies $h = 0.02$. Regresijas un to atlikumu autokorelāciju grafiki ir apskatāmi 6. attēlā.

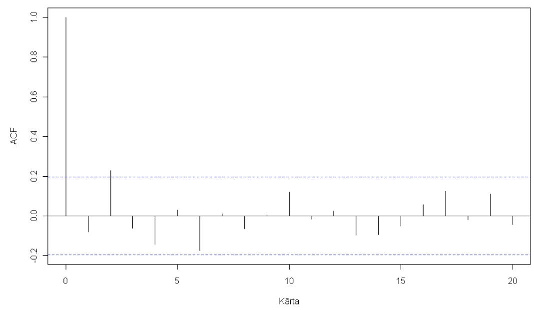


6. att.: Simulētie dati: (a) lokālās lineārās regresijas veidota regresijas līkne ar parametru $h = 0.1$; (b) atlikumu autokorelācijas; (c) lokālās lineārās regresijas veidota regresijas līkne ar parametru $h = 0.02$; (d) atlikumu autokorelācijas.

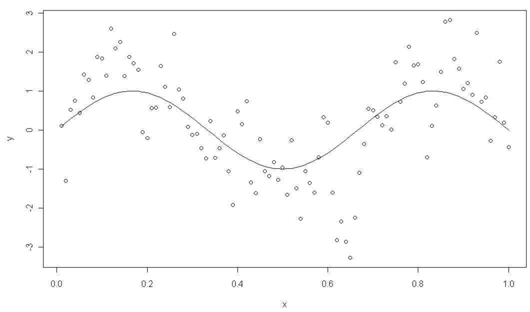
Rezultāti liecina par korelētiem atlikumiem pirmajā gadījumā un nekorelētiem otrajā. Tālāk attēlosim īsto regresijas līkni $y_i = \sin(3\pi x_i)$ un attēlosim atlikumu autokorelācijas grafikus.



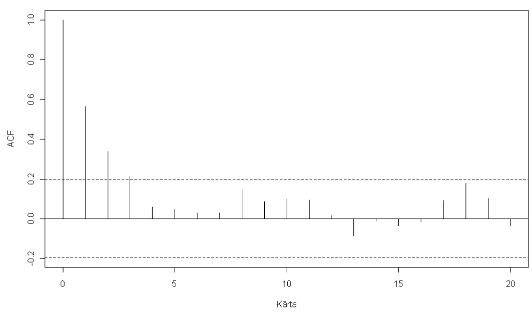
(a)



(b)



(c)



(d)

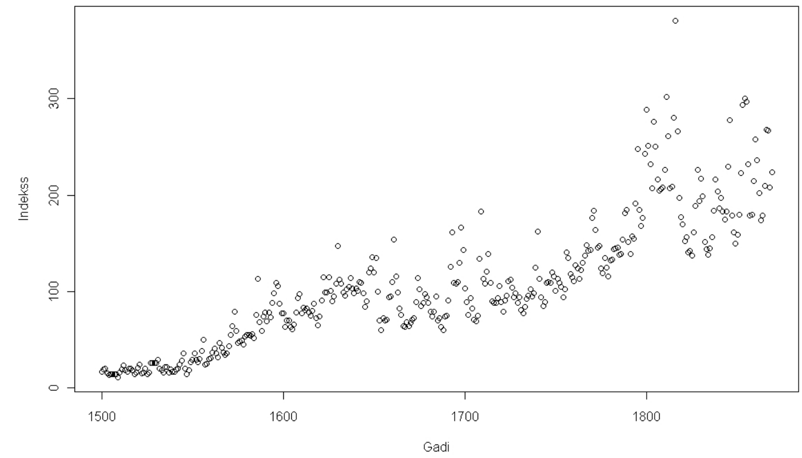
7. att.: Simulētie dati: (a) īstā regresijas funkcija; (b) atlikumu autokorelācijas; (c) īstā regresijas funkcija; (d) atlikumu autokorelācijas.

Tātad redzams, ka nepareizi konstruējot regresijas līknes (6.(a) un 6.(c) attēlos), abos gadījumos slēdzieni par atlikumiem ir bijuši aplami. Pirmajā gadījumā ieguvām atkarīgus atlikumus, jo izvēlētais gludināšanas parametrs $h = 0.1$ *pārgludināja* novērojumus, un mazāks gludināšanas parametrs būtu atbilstošāks. Otrajā gadījumā (6.(c) un 6.(d) attēli) ieguvām neatkarīgus atlikumus, kas neatbilst simulētajiem datiem(7.(d) attēls). Šajā gadījumā pārāk mazs gludināšanas parametrs dod regresiju, kura ir konstruēta, neņemot vērā atlikumu atkarības struktūru. Šī piemēra ideja ņemta no [24].

No iepriekšējā piemēra ir skaidrs, ka nepareiza gludināšanas parametra izvēle var novest pie aplamiem rezultātiem. Tomēr ne vienmēr var paļauties uz atlikumu neatkarību pazīmi un secināt, ka izvēlētais modelis ir labs. Hart ([25]) parādīja, ka daudzas gludināšanas parametra izvēles metodes, ieskaitot korsvalidāciju, dod nepareizus rezultātus gadījumos, kad atlikumi no regresijas funkcijas ir korelēti.

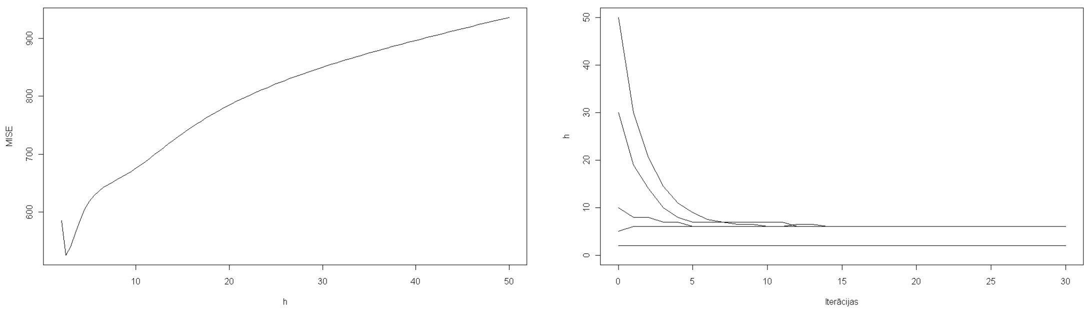
8.3. Beveridžas indeksu dati

Aplūkosim Beveridžas graudu cenu indeksus, kuri ir iegūstami no programmas R. Dati ir gada vidējie Eiropas graudu cenu indeksi no 1500. līdz 1869.gadam, skatīt 8. attēlā.



8. att. Beveridžas cenu indeksu dati.

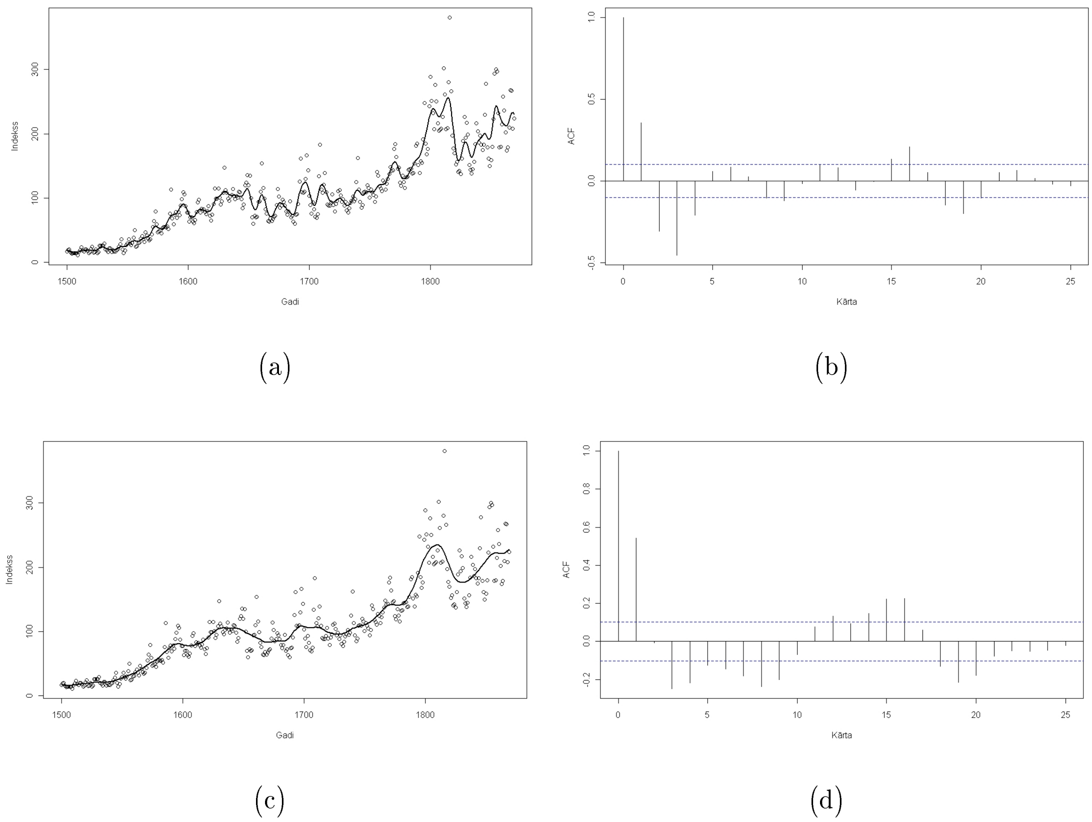
Pielietosim krosvalidācijas un butstrapa metodes, lai atrastu neparametriskās regresijas gludināšanas parametru. Krosvalidācijas *atstāt vienu ārā* metodei lietosim soli 0.5 un gludināšanas parametru meklēsim intervālā [0,50]. Atradīsim katra soļa novērtēto MISE un izvēlēsimies to h, pie kura $\widehat{\mathrm{MISE}}(h)$ sasniedz savu minimumu. Butstrapa procedūrā veiksim 30 iterāciju soļus ar pieciem dažādiem sākuma $h_0 = 2, 5, 10, 30, 50$. Katrā no iterācijas soļiem veidosim 2000 butstrapa izlases un meklēsim h_{opt} ar soli 0.5 pa intervālu [0,50].



(a) (b)

9. att.: Beveridžas indeksu datu regresijas optimālā gludināšanas parametra izvēle: (a) korsvalidācijas novērtētā MISE; (b) butstrapa iterāciju rezultāts (pie iterācijas soļa 0 h vērtības ir vienādas ar sākuma h_0 izvēli).

Pēc krosvalidācijas metodes esam ieguvuši $h_{opt} = 2.5$. Butstrapa metodei ar sākuma $h_0 = 2$ redzams, ka h izvēle paliek nemainīga visos iterāciju soļos. Pārējām iterācijām ar $h_0 = 5, 10, 30, 50$ gludināšanas parametrs iterāciju beigās dod vērtību 6. To arī izvēlēsimies par optimālo h_{opt}. Ar šiem parametriem veidosim regresijas un attēlosim regresiju atlikumu autokorelāciju grafikus.



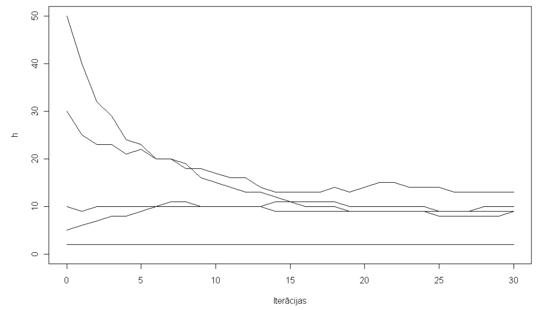
10. att.: Beveridžas indeksu dati: (a) ar krosvalidācijas metodi izvēlētais $h = 2.5$ un (b) atlikumu autokorelāciju grafiks; (c) ar butstrapa metodi izvēlētais $h = 6$ un (d) atlikumu autokorelāciju grafiks.

Regresija, kuras gludināšanas parametrs ir iegūts pēc krosvalidācijas metodes, ir tuva datu interpolācijai. Butstrapa metodes dotais gludināšanas parametrs $h_0 = 6$ dod regresiju, kura arī, šķiet, pārāk *seko* datiem. Redzams, ka abām regresijām, atlikumi uzrāda korelācijas pazīmi, pie tam, ar butstrapu veidotajai regresijai, korelācijas, šķiet, lielākas. Šis piemērs rāda, ka krosvalidācijas un butstrapa metode nedod labus rezultātus.

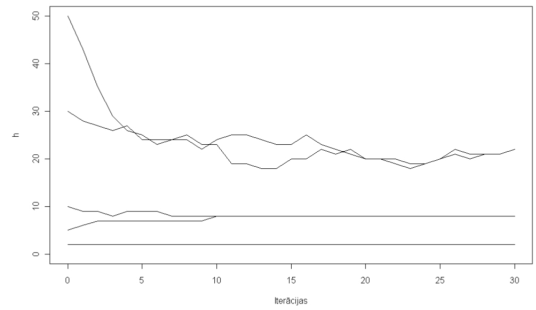
Risinājumu piedāvāja Hall, Lahiri un Polzehl ([21]), saistībā ar bloku butstrapu. Šī metode ir modificēta butstrapa metode, ko apskatījām iepriekš. Galvenā atšķirība ir tā, ka sākotnēji tiek lietoti divi sākuma h_{01} un h_{02}. Ar h_{01} iegūst atlikumu izlasi priekš butstrapa, bet ar h_{02} iegūst regresijas līkni, pie kuras pievieno butstrapotos atlikumus. Šajā gadījumā lieto bloku butstrapu, lai saglabātu atlikumu atkarības struktūru. Autori ir norādījuši, ka jāizvēlas $h_{02} = Ch_{01}^{5/9}$. Diemžēl konstante C ir aizklātā formā, kuru praktiski ir grūti iegūt.

Mūsu gadījumā lietosim $h_{01} = h_{02}$ un bloku butstrapa procedūrai izmantosim MBB metodi. Šajā gadījumā MBB metode, šķiet, ir vispiemērotākā, jo, kā zināms, tā uzliek mazākus svarus izlases sākuma un beigu elementiem pie butstrapošanas, bet neparametriskajā regresijā ir gadījumi, kad regresija pie robežām strādā slikti. Hall, Lahiri un

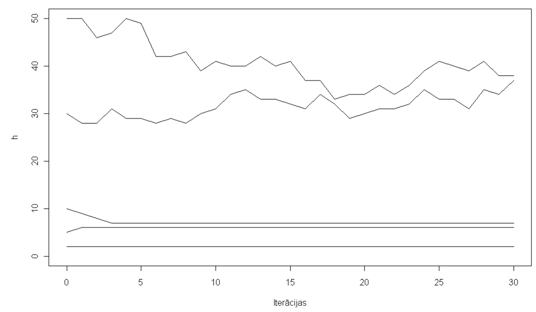
Polzehl pat iesaka atmest atlikumu izlases sākuma un beigu elementus. Ieviešot bloku butstrapu, parādās optimālā bloka garuma problemātika. Šim piemēram izvēlēsimies 3 bloku garumus $l = 5, 10, 20$, ar tiem veiksim 2000 butstrapa izlases. Sākumā fiksēsim $h_{01} = h_{02} = 2, 5, 10, 30, 50$. Meklēsim optimālo h_{opt} intervālā [0,50] ar soli 1. Visu trīs bloku garumu iterāciju rezultāti apskatāmi 11. attēlā.



(a)



(b)

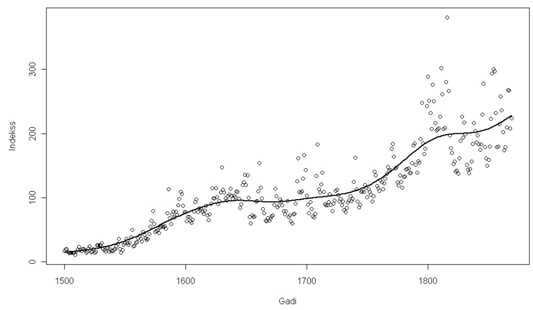


(c)

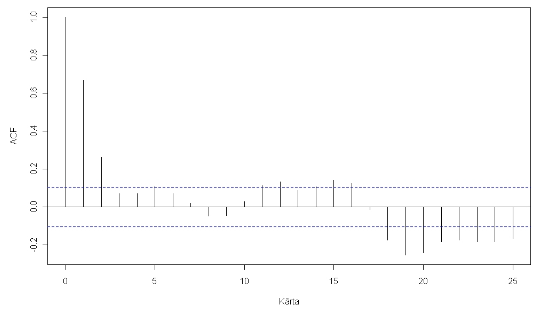
11. att.: Beveridžas indeksu datu iteratīvās bloku butstrapa procedūras: (a) ar bloku garumu $l = 5$; (b) ar bloku garumu $l = 10$; (c) ar bloku garumu $l = 20$.

Redzams, ka ar sākuma vērtībām $h_{01} = 2, 5, 10$, iterāciju laikā parametrs h nemainās, kas, iespējams, skaidrojams ar to, ka, iegūtie atlikumi slikti reprezentē patieso atkarības struktūru, kuru nav iespējams *atjaunot* ar butstrapu. Tas pats sakāms par iterācijas procesiem, kur bloku garums $l = 5$, kur iterāciju rezultāti krītas visiem h_{01}. Par optimāliem izvēlēsimies rezultātus, kur $l = 10, 20$ un $h_0 = 30, 50$. Pie $l = 10$ bloku butstrapa iterāciju process dod $h_{opt} = 22$, bet, ja $l = 20$, tad $h_{opt} = 38$. Tātad esam izvēlējušies divus optimālos gludināšanas parametrus.

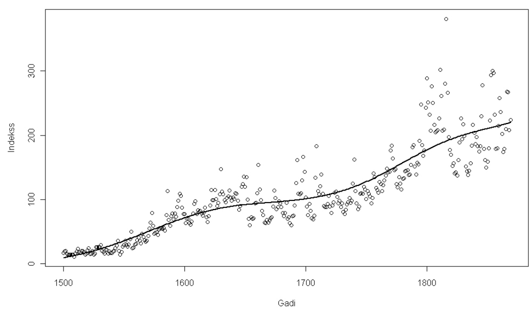
Attēlosim regresijas ar iegūtajiem gludināšanas parametriem.



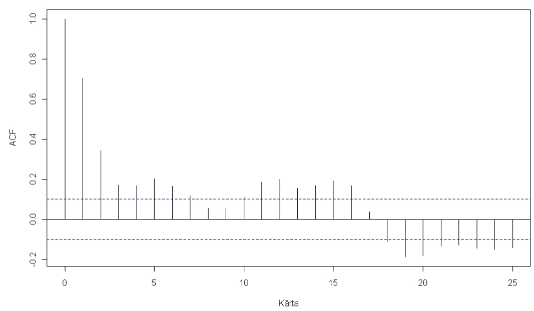
(a)



(b)



(c)



(d)

12. att.: Beveridžas indeksu dati: (a) ar bloku butstrapa metodi izvēlētais $h = 22$ un (b) atlikumu autokorelāciju grafiks, (c) ar bloku butstrapa metodi izvēlētais $h = 38$ un (d) atlikumu autokorelāciju grafiks.

Lai gan optimālie h_{opt} pie abiem bloku garumiem atšķiras gandrīz divas reizes, tomēr regresijas līknes ir *līdzīgas*. Tas skaidrojams ar to, ka datu apjoms ir pietiekami liels pret gludinošajiem parametriem. Iegūtās regresijas pret novērotajiem $Y_1, \ldots, Y_n$ dod atlikumus ar korelācijas pazīmēm.

Savās publikācijās Hart ([25] un [26]) deva metodi, kā novērtēt regresiju ar kodola gludinātājiem, kad dati ir korelēti. Šai metodē tiek veidota korelācijas struktūra, kura tiek novērtēta un ņemta vērā izvēloties optimālo h_{opt}. Praktiskajā daļā tika apskatīti Beveridžas indeksu dati. Datiem tika veidotas divas atkarības struktūras: AR(1) un AR(2). Atkarībā no izvēlētā modeļa, tika iegūti divi optimālie regresijas gludināšanas parametri $h_{opt,AR(1)} = 37$ un $h_{opt,AR(2)} = 24$, kas ir salīdzināmi ar mūsu bloku butstrapa metodes rezultātiem. Rezultātu līdzības ir izskaidrojamas, jo arī mēs, lietojot divus dažādus bloku garumus $l = 10$ un $l = 20$, pieņēmām dažādas atkarības struktūras.

9. Secinājumi

Butstrapa metode ir plaši pielietota praksē neatkarīgiem un vienādi sadalītiem datiem. Taču tā nav piemērota datiem ar atkarības struktūru. Apskatījām vienkāršu piemēru, kad šādas butstrapa metodes nav lietojamas atkarīgiem datiem, jo dod nepareizu dispersiju.

Tālāk parādījām bloku butstrapa metodes atkarīgiem datiem un devām tās īpašības dažos pielietojumos. No teorijas parādījām, ka CBB un MBB metodes dod labākos parametra dispersijas un novirzes novērtējumus nekā NBB un SB metodes, ko mēģinājām parādīt arī praktiski. Savukārt, ja mūsu interesējošā butstrapa statistika ir ar centrējošu lielumu tādu, kas vienāds ar butstrapa matemātisko cerību šādam lielumam, tad priekšrocība ir CBB un SB metodēm, jo centrējošā vērtība ir vienāda ar sākuma izlases attiecīgo parametru. MBB un NBB metodēm šis centrējošais parametrs ir speciālā formā, ne vienmēr praktiski viegli iegūstams. Tāpēc būtu uzsverama CBB metodes priekšrocība pār MBB metodi.

Neatkarīgi no izvēlētās bloku butstrapa metodes, eksistē divas butstrapa pieejas: ,,naivā" un ,,vienkāršā". Ja izvēlētā butstrapa statistika ir atkarīga ne tikai no atsevišķiem izlases elementiem, bet arī no to savstarpējām kombinācijām, tad ,,vienkāršā" metode būtu rekomendējama šādos gadījumos. Kā piemēru darbā parādījām gadījumu, kad mūsu interesējošā statistika ir izlases kovariācijas novērtējums ar kārtu 2. Ar datu simulāciju, palīdzību redzējām, ka ,,vienkāršās" pieejas rezultāti ir stabilāki pie dažādiem bloku garumiem, salīdzinot ar ,,naivo" pieeju.

Praktisku pielietojumu bloku butstrapa metodei apskatījām neparametriskās regresijas gadījumā, kad regresijas atlikumi ir korelēti. Šādos gadījumos populāras gludināšanas parametra izvēles procedūras nestrādā un dod joslas platuma novērtējumu, ar kuru regresija gandrīz interpolē datus. Savukārt butstrapa metodes pielietojums dot ticamus rezultātus, jo tiek ņemta vērā atlikumu atkarība. Šo daļu praktiski veicām ar Beveridžas cenu indeksu datiem un ieguvām divus gludināšanas parametrus, kuri ir salīdzināmi ar citās publikācijās iegūtajiem rezultātiem. Secinājumus par to, kurš no diviem parametriem un kura no regresijām būtu visprecīzākā, atstājām atvērtus, jo tie ir atkarīgi no datu interpretācijas un pieņēmumiem par to atkarību. Lai to noteiktu, būtu jāveido piemēri ar simulētiem datiem, kad īstā regresijas funkcija ir zināma. Iespējams, būtu jāpilnveido dotā metode heteroskedastisku atlikumu gadījumā. Kaut gan no piemēriem redzējām, ka arī šajos gadījumos metode strādā, taču tā sākotnēji tika veidota homoskedastisku atlikumu gadījumā. Šeit būtu izmantojamas butstrapa metodes, kas paredzētas heteroskedastisku atlikumu gadījumā, kombinētas ar bloku butstrapa metodēm.

Visbeidzot varam secināt par bloku butstrapa metožu praktisku pielietojumu un, iespējamu, tālāku attīstību. Pirmkārt ir jāuzsver tas, ka visas butstrapa metodes ir ļoti darbietilpīgas no datoru resursu viedokļa, tāpēc viens no praktiskās datorzinātnes un matemātikas izaicinājumiem būtu meklēt efektīvākus algoritmus, lai butstrapa metodes padarītu ikdienā pieejamākas. Šī problēma īpaši jūtama būtu finanšu jomā, kur parasti datu apjoms ir liels un pat mūsdienu datoram butstrapa procedūra būtu liels pārbaudījums. Otrs virziens ir pašu matemātisko modeļu izstrāde. Četras apskatītās bloku butstrapa metodes ir plaši izmantojamas, jo ir vāji nosacījumi to lietošanai, bet cena par to ir lielākas novirzes novērtētajiem parametriem. Jau pieminējām, ka eksistē bloku butstrapa metodes, piemēram, MaBB un TaBB, kuras dod precīzākus rezultātus, kā MBB, NBB, CBB un SB. Pie tam atkarīgu datu gadījumā eksistē arī citas butstrapa metodes, piemēram, sieta butstraps (no angļu valodas ,,sieve bootstrap"), kurš tāpat, kā MaBB un TaBB, dod labus rezultātus, bet ar papildus nosacījumiem.

10. Izmantotā literatūra un avoti

Literatūra

[1] B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics.* Vol. 7, 1 - 26, 1979.

[2] K. Singh. On the asymptotic accuracy of the Efron's bootstrap. *The Annals of Statistics.* Vol. 9, 1187 - 1195, 1981.

[3] P. Hall. Resampling a coverage pattern. *Stochastic Processes and Their Applications.* Vol. 20, 231 - 246, 1985.

[4] E. Carlstein. The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *The Annals of Statistics.* Vol. 14, 1171 - 1179, 1986.

[5] H. R. Kunsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics.* Vol. 17, 1217 - 1261, 1989.

[6] R. Y. Liu, K. Singh. Moving blocks jackknife and bootstrap capture weak dependence. *Exploring the Limits of the Bootstrap.* 225 - 248, 1992.

[7] S. N. Lahiri. *Resamplig Methods for Dependent Data.* Springler-Verlag, New York, 2003.

[8] R. H. Shumway, D. S. Stoffer. *Time Series Analysis and its Applications with R Examples.* Springer Science+Business Media, LLC, 2006.

[9] D. N. Politis, J. P. Romano. A circular block resampling procedure for stationary data. *Exploring the Limits of the Bootstrap.* 263 - 270, 1992.

[10] M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.* 43 - 47, 1956.

[11] P. Doukhan. *Mixing: Properties and Examples.* Springler-Verlag, New York, 1994.

[12] R. N. Bhattacharya, J. K. Ghosh. On the validity of the formal Edgeworth expansion. *The Annals of Statistics.* Vol. 7, 434 - 451, 1978.

[13] P. Billingsley. *Covergence of Probability Measures.* Wiley, New York, 1968.

[14] R. von Mises. On the asymptotic distribution of differentiable satistical functions. *The Annals of Mathematical Statistics.* Vol. 18, 309 - 348, 1947.

[15] D. J. Nordman. A note on the stationary bootstrap's variance. *The Annals of Statistics.* Vol. 37, No. 1, 359 - 370, 2009.

[16] D. N. Politis, H. White. Automatic block-lenght selection for the dependent bootstrap. *Econometric Reviews.* Vol. 23, No. 1, 2004.

[17] A. Patton, D. N. Politis, H. White. Correction to ,,Automatic block-lenght selection for the dependent bootstrap". *Econometric Reviews.* Vol. 28, Issue 4, 372 - 375, 2009.

[18] P. Buhlmann, H. R. Kunsch. Block lenght selection in the bootstrap for time series. *Seminar fur Statistik, ETH Zurich, Switzerland.* 1999.

[19] E. Carlstein, K. A. Do, P. Hall, T. Hesterberg, H. R. Kunsch. Matched-block bootstrap for dependent data. *Bernoulli.* Vol. 4, 305 - 328, 1998.

[20] E. Paparoditis, D. N. Politis. Tapered block bootstrap. *Biometrika.* Vol. 88, No. 4, 1105 - 1119, 2001.

[21] P. Hall, S. N. Lahiri, J. Polzehl. On bandwidth choice in nonparametric regression with both short- and long-range dependent errors. *The Annals of Statistics.* Vol. 23, No. 6, 1921 - 1936, 1995.

[22] J. J. Faraway. Bootstrap selection of bandwidth and confidence bands for nonparametric regression. *J. Statist. Comput. Simul.* Vol. 37, 37 - 44, 1990.

[23] L. Wasserman. *All of Nonparametric Statistics.* Springer Science+Business Media, Inc, 2006.

[24] J. Opsomer, Y. Wang, Y. Yang. Nonparametric regression with correlated errors. *Statistical Science.* Vol. 16, No. 2, 134 - 153, 2001.

[25] J. D. Hart. Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society. Series B (Methodological).* Vol. 53, No. 1, 173 - 187, 1991.

[26] J. D. Hart. Automated kernel smoothing of dependent data by using time series cross-validation. *Journal of the Royal Statistical Society. Series B (Methodological).* Vol. 56, No. 3, 529 - 542, 1994.

11. Pielikumi

Butstrapa neatbilstība atkarīgiem datiem:

```
#Sakuma parametri
nn<-50
b<-1000
l<-10
fi<-0.2
theta<--0.4

#Datu generesana
x1<-arima.sim(list(ar=c(fi),ma=c(theta)),n=nn)
plot(x1)

#Teoretiska vertiba
gamma<-c()
gamma[1]<-(1+theta*fi+theta^2)/(1-fi^2)
i<-2
while (abs(gamma[i-1])>0.001){
gamma[i]<-(1+theta*fi)*(fi+theta)/(1-fi^2)*fi^(i-2)
i<-i+1}
var.theor<-gamma[1]+2*sum(gamma[2:length(gamma)])

#Bloku butstrapa funkcija
nbb.mean<-function(fdata,fl,fbl){
mean(fdata[1:(fl*fbl)])}

nbb.fun<-function(fdata,fn,fl,fk0,fbn,fbl){
bootsamp<-c()
samp<-sample(fbl,fk0,replace=TRUE)
for (fi in 1:fk0){
bootsamp<-c(bootsamp,fdata[(fl*(samp[fi]-1)+1):(fl*(samp[fi]-1)+fl)])}
sqrt(fn)*(mean(bootsamp[1:fn])-nbb.mean(fdata,fl,fbl))}

#Butstrapa funkcija
iidb.fun<-function(fdata,fn){
bootsamp<-sample(fdata,fn,replace=TRUE)
sqrt(fn)*(mean(bootsamp[1:fn])-mean(fdata))}

#Monte-Carlo
nbb.sample<-sapply(1:b,function(i)(nbb.fun(x1,nn,l,ceiling(nn/l),ceiling(nn/l)*l,floor(nn/l))))
iddb.sample<-sapply(1:b,function(i)(iidb.fun(x1,nn)))

#Rezultati
mean(nbb.sample)
mean(iddb.sample)
var(nbb.sample)
var(iddb.sample)
```

,,Naivais" un ,,vienkāršais" butstraps:

```
#Sakuma parametri
nn<-102
k<-2
l<-25
b<-1000
fi<--0.5
theta<-0.8

#Datu generesana
x0<-arima.sim(list(ar=c(fi),ma=c(theta)),n=nn)
plot(x0)

#Simulacijas
cov.k<-(1+theta*fi)*(fi+theta)/(1-fi^2)*fi^(k-1)
acf(x0,type="covariance")[k]

n.sim<-10000

disp<-function(){
sim.disp<-arima.sim(list(ar=c(fi),ma=c(theta)),n=nn)
covv.disp<-function(j){
sim.disp[j+k]*sim.disp[j]-mean(sim.disp[1:(nn-k)])^2}
sqrt(nn-k)*(mean(sapply(1:(nn-k),covv.disp))-cov.k)}

cov.disp<-mean(sapply(1:n.sim,function(i)disp())^2)

#Butsrapi
#Ordinary
ord.k0<-ceiling((nn-k)/l)
ord.bn<-ord.k0*l

x<-array(,c(nn-k,2))

x[,1]<-x0[1:(nn-k)]
x[,2]<-sapply(1:(nn-k),function(i)(x0[i]*x0[i+k]))

smooth<-function(a,b){
b-a^2}

#MBB
ord.mbb.bl<-nn-k-l+1

ord.mbb.mju1<-sum(sapply(1:ord.mbb.bl,function(i)((ord.k0-1)*sum(x[i:(i+l-1),1])+sum(x[i:(i+nn-k-(ord.k0-1)*l-1),1]))))/ord.mbb.bl/(nn-k)
ord.mbb.mju2<-sum(sapply(1:ord.mbb.bl,function(i)((ord.k0-1)*sum(x[i:(i+l-1),2])+sum(x[i:(i+nn-k-(ord.k0-1)*l-1),2]))))/ord.mbb.bl/(nn-k)

ord.mbb.fun<-function(){
bootsamp1<-c()
bootsamp2<-c()
```

```
samp<-sample(ord.mbb.bl,ord.k0,replace=TRUE)
for (fi in 1:ord.k0){
bootsamp1<-c(bootsamp1,x[samp[fi]:(samp[fi]+l-1),1])
bootsamp2<-c(bootsamp2,x[samp[fi]:(samp[fi]+l-1),2])}
sqrt(nn-k)*(smooth(mean(bootsamp1[1:(nn-k)]),mean(bootsamp2[1:(nn-k)]))-smooth(ord.mbb.mju1,ord.mbb.mju2))}

ord.mbb.fi<-sum(sapply(1:b,function(i)ord.mbb.fun())^2)/b

#NBB
ord.nbb.bl<-floor((nn-k)/l)

ord.nbb.mju1<-mean(x[1:(ord.nbb.bl*l),1])
ord.nbb.mju2<-mean(x[1:(ord.nbb.bl*l),2])

ord.nbb.fun<-function(){
bootsamp1<-c()
bootsamp2<-c()
samp<-sample(ord.nbb.bl,ord.k0,replace=TRUE)
for (fi in 1:ord.k0){
bootsamp1<-c(bootsamp1,x[(l*(samp[fi]-1)+1):(l*(samp[fi]-1)+l),1])
bootsamp2<-c(bootsamp2,x[(l*(samp[fi]-1)+1):(l*(samp[fi]-1)+l),2])}
sqrt(nn-k)*(smooth(mean(bootsamp1[1:(nn-k)]),mean(bootsamp2[1:(nn-k)]))-smooth(ord.nbb.mju1,ord.nbb.mju2))}

ord.nbb.fi<-sum(sapply(1:b,function(i)ord.nbb.fun())^2)/b

#CBB
ord.cbb.bl<-nn-k

ord.cbb.mju1<-mean(x[,1])
ord.cbb.mju2<-mean(x[,2])

ord.cbb.fun<-function(){
bootsamp1<-c()
bootsamp2<-c()
samp<-sample(ord.cbb.bl,ord.k0,replace=TRUE)
for (fi in 1:ord.k0){
bootsamp1<-c(bootsamp1,x[samp[fi]:min(samp[fi]+l-1,nn-k),1],x[0:max(0,samp[fi]+l-1-(nn-k)),1])
bootsamp2<-c(bootsamp2,x[samp[fi]:min(samp[fi]+l-1,nn-k),2],x[0:max(0,samp[fi]+l-1-(nn-k)),2])}
sqrt(nn-k)*(smooth(mean(bootsamp1[1:(nn-k)]),mean(bootsamp2[1:(nn-k)]))-smooth(ord.cbb.mju1,ord.cbb.mju2))}

ord.cbb.fi<-sum(sapply(1:b,function(i)ord.cbb.fun())^2)/b

#SB
ord.sb.bl<-nn-k

ord.sb.mju1<-mean(x[,1])
ord.sb.mju2<-mean(x[,2])

ord.sb.fun<-function(){
lk<-c()
while(sum(lk)<ord.sb.bl){
lk<-c(lk,rgeom(1,1/l))
lk<-lk[lk>0]}
bootsamp1<-c()
bootsamp2<-c()
samp<-sample(ord.sb.bl,length(lk),replace=TRUE)
for(fj in 1:length(lk)) {
bootsamp1<-c(bootsamp1,x[samp[fj]:min(samp[fj]+lk[fj]-1,nn-k),1],x[0:max(0,samp[fj]+lk[fj]-1-(nn-k)),1])
bootsamp2<-c(bootsamp2,x[samp[fj]:min(samp[fj]+lk[fj]-1,nn-k),2],x[0:max(0,samp[fj]+lk[fj]-1-(nn-k)),2])}
sqrt(nn-k)*(smooth(mean(bootsamp1[1:(nn-k)]),mean(bootsamp2[1:(nn-k)]))-smooth(ord.sb.mju1,ord.sb.mju2))}

ord.sb.fi<-sum(sapply(1:b,function(i)ord.sb.fun())^2)/b

#Naive
nai.k0<-ceiling(nn/l)
nai.bn<-nai.k0*l

#MBB
nai.mbb.bl<-nn-l+1

nai.mbb.mju<-mean(sapply(1:(nn-k),function(i)(x0[i]*x0[i+k])))-mean(x0[1:(nn-k)])^2

nai.mbb.fun<-function(){
bootsamp<-c()
samp<-sample(nai.mbb.bl,nai.k0,replace=TRUE)
for (fi in 1:nai.k0){
bootsamp<-c(bootsamp,x0[samp[fi]:(samp[fi]+l-1)])}
sqrt(nn-k)*(mean(sapply(1:(nn-k),function(i)(bootsamp[i]*bootsamp[i+k])))-mean(bootsamp[1:(nn-k)])^2-nai.mbb.mju)}

nai.mbb.fi<-sum(sapply(1:b,function(i)nai.mbb.fun())^2)/b

#NBB
nai.nbb.bl<-floor(nn/l)

nai.nbb.mju<-mean(sapply(1:(nai.nbb.bl*l-k),function(i)(x0[i]*x0[i+k])))-mean(x0[1:(nai.nbb.bl*l-k)])^2

nai.nbb.fun<-function(){
bootsamp<-c()
samp<-sample(nai.nbb.bl,nai.k0,replace=TRUE)
for (fi in 1:nai.k0){
bootsamp<-c(bootsamp,x0[(l*(samp[fi]-1)+1):(l*(samp[fi]-1)+l)])}
sqrt(nn-k)*(mean(sapply(1:(nn-k),function(i)(bootsamp[i]*bootsamp[i+k])))-mean(bootsamp[1:(nn-k)])^2-nai.nbb.mju)}

nai.nbb.fi<-sum(sapply(1:b,function(i)nai.nbb.fun())^2)/b

#CBB
nai.cbb.bl<-nn

nai.cbb.mju<-mean(sapply(1:(nn-k),function(i)(x0[i]*x0[i+k])))-mean(x0[1:(nn-k)])^2

nai.cbb.fun<-function(){
bootsamp<-c()
samp<-sample(nai.cbb.bl,nai.k0,replace=TRUE)
```

```
for (fi in 1:nai.k0){
bootsamp<-c(bootsamp,x0[samp[fi]:min(samp[fi]+l-1,nn-k)],x0[0:max(0,samp[fi]+l-1-(nn-k))])}
sqrt(nn-k)*(mean(sapply(1:(nn-k),function(i)(bootsamp[i]*bootsamp[i+k])))-mean(bootsamp[1:(nn-k)])^2-nai.cbb.mju)}

nai.cbb.fi<-sum(sapply(1:b,function(i)nai.cbb.fun())^2)/b

#SBB
nai.sb.bl<-nn

nai.sb.mju<-mean(sapply(1:(nn-k),function(i)(x0[i]*x0[i+k])))-mean(x0[1:(nn-k)])^2

nai.sb.fun<-function(){
lk<-c()
while(sum(lk)<nai.sb.bl){
lk<-c(lk,rgeom(1,1/l))
lk<-lk[lk>0]}
bootsamp<-c()
samp<-sample(nai.sb.bl,length(lk),replace=TRUE)
for(fj in 1:length(lk)){
bootsamp<-c(bootsamp,x0[samp[fj]:min(samp[fj]+lk[fj]-1,nn-k)],x0[0:max(0,samp[fj]+lk[fj]-1-(nn-k))])}
sqrt(nn-k)*(mean(sapply(1:(nn-k),function(i)(bootsamp[i]*bootsamp[i+k])))-mean(bootsamp[1:(nn-k)])^2-nai.sb.mju)}

nai.sb.fi<-sum(sapply(1:b,function(i)nai.sb.fun())^2)/b

#Rezultati
ord.mbb.fi
ord.nbb.fi
ord.cbb.fi
ord.sb.fi

nai.mbb.fi
nai.nbb.fi
nai.cbb.fi
nai.sb.fi
```

MBB, NBB, CBB un SB salīdzinājums:

```
#Biblioteka
library(MASS)

#Sakuma nosacijumi
nn<-400
k<-2000
b<-500
l<-20
phi<-0.6
theta<--0.1

#Datu generesana
x4<-arima.sim(list(ar=c(phi),ma=c(theta)),n=nn)

#Butstrapa funkcijas
#MBB
mbb.mean<-function(fdata,fl,fn){
1/(fn-fl+1)*(fn*mean(fdata)-1/fl*sum(sapply(1:(fl-1),function(i)((fl-i)*(fdata[i]+fdata[fn-i+1]))))))}

mbb.fun<-function(fdata,fn,fl,fk0,fbn,fbl){
bootsamp<-c()
samp<-sample(fbl,fk0,replace=TRUE)
for (fi in 1:fk0){
bootsamp<-c(bootsamp,fdata[samp[fi]:(samp[fi]+fl-1)])}
sqrt(fn)*(mean(bootsamp[1:fn])-mbb.mean(fdata,fl,fn))}

#NBB
nbb.mean<-function(fdata,fl,fbl){
mean(fdata[1:(fl*fbl)])}

nbb.fun<-function(fdata,fn,fl,fk0,fbn,fbl){
bootsamp<-c()
samp<-sample(fbl,fk0,replace=TRUE)
for (fi in 1:fk0){
bootsamp<-c(bootsamp,fdata[(fl*(samp[fi]-1)+1):(fl*(samp[fi]-1)+fl)])}
sqrt(fn)*(mean(bootsamp[1:fn])-nbb.mean(fdata,fl,fbl))}

#CBB
cbb.fun<-function(fdata,fn,fl,fk0,fbn,fbl){
bootsamp<-c()
samp<-sample(fbl,fk0,replace=TRUE)
for (fi in 1:fk0){
bootsamp<-c(bootsamp,fdata[samp[fi]:min(samp[fi]+fl-1,fn)],fdata[0:max(0,samp[fi]+fl-1-fn)])}
sqrt(fn)*(mean(bootsamp[1:fn])-mean(fdata))}

#SB
sb.fun<-function(fdata,fn,fl,fk0,fbn,fbl){
lk<-c()
while(sum(lk)<fbl){
lk<-c(lk,rgeom(1,1/fl))
lk<-lk[lk>0]}
bootsamp<-c()
samp<-sample(fbl,length(lk),replace=TRUE)
for(fj in 1:length(lk)) {
bootsamp<-c(bootsamp,fdata[samp[fj]:min(samp[fj]+lk[fj]-1,fn)],fdata[0:max(0,samp[fj]+lk[fj]-1-fn)])}
sqrt(fn)*(mean(bootsamp[1:fn])-mean(fdata))}

#Butstrapa procedura
x4.fi.mbb<-array(,c(k,l-1))
x4.fi.nbb<-array(,c(k,l-1))
x4.fi.cbb<-array(,c(k,l-1))
x4.fi.sb<-array(,c(k,l-1))

for (ll in 2:l){
k0<-ceiling(nn/ll)
bn<-k0*ll
```

```
mbb.bl<-nn-ll+1
nbb.bl<-floor(nn/ll)
cbb.bl<-nn
sb.bl<-nn
for (kk in 1:k){
st<-Sys.time()
x4.fi.mbb[kk,ll-1]<-mean(sapply(1:b,function(i)mbb.fun(x4,nn,ll,k0,bn,mbb.bl))^2)
x4.fi.nbb[kk,ll-1]<-mean(sapply(1:b,function(i)nbb.fun(x4,nn,ll,k0,bn,nbb.bl))^2)
x4.fi.cbb[kk,ll-1]<-mean(sapply(1:b,function(i)cbb.fun(x4,nn,ll,k0,bn,cbb.bl))^2)
x4.fi.sb[kk,ll-1]<-mean(sapply(1:b,function(i)sb.fun(x4,nn,ll,k0,bn,sb.bl))^2)
bt<-Sys.time()
print(c(ll,kk,bt-st))}}

#Teoretiskais fi
gamma<-c()
gamma[1]<-(1+theta*phi+theta^2)/(1-phi^2)
i<-2
while (abs(gamma[i-1])>0.001){
gamma[i]<-(1+theta*phi)*(phi+theta)/(1-phi^2)*phi^(i-2)
i<-i+1}

var.theor<-gamma[1]+2*sum(gamma[2:length(gamma)])

#Grafiki
mse.mbb<-c()
mse.nbb<-c()
mse.cbb<-c()
mse.sb<-c()

for (i in 2:l){
mse.mbb[i-1]<-mean((x4.fi.mbb[,i-1]-var.theor)^2)
mse.nbb[i-1]<-mean((x4.fi.nbb[,i-1]-var.theor)^2)
mse.cbb[i-1]<-mean((x4.fi.cbb[,i-1]-var.theor)^2)
mse.sb[i-1]<-mean((x4.fi.sb[,i-1]-var.theor)^2)}

plot(2:l,mse.mbb,type="l",xlim=c(0,23),ylab="MSE",xlab="l")
text(21,0,"MBB")
lines(2:l,mse.nbb,type="l")
text(21,2,"NBB")
lines(2:l,mse.cbb,type="l")
text(21,0.5,"CBB")
lines(2:l,mse.sb,type="l")
text(21,1,"SB")
```

Atlikumu analīze:

```
#Bibliotekas
library(RandomFields)
library(KernSmooth)

#Sakuma parametri
n<-100
x<-(1:n)/n
sigma<-1
lamda<-70

#Atkarigi atlikumi
g.sim<-GaussRF(x=lamda*x,grid=F,model="exponential",param=c(mean=0,variance=sigma,nugget=0,scale=1))

y<-sapply(x,function(i)(sin(3*pi*i)))
plot(x,y,type="l")

e<-y+g.sim
sm.m<-locpoly(x,e,bandwidth=0.02,gridsize=n)

plot(x,e,ylab="y")
lines(x,sm.m$y)
lines(x,y)
acf(e-sm.m$y,main="")
acf(g.sim,main="")

#Neatkarigi atlikumi
g.sim<-rnorm(100,0,sqrt(0.01))

y<-sapply(x,function(i)(sin(3*pi*i)))

e<-y+g.sim
sm.m<-locpoly(x,e,bandwidth=0.1,gridsize=n)

plot(x,e,ylab="y")
lines(x,sm.m$y)
lines(x,y)
acf(e-sm.m$y,main="")
acf(e-y,main="")
```

Butstrapa un krosvalidācijas metodes:

```
#Bibliotekas
library(KernSmooth)
library(tseries)
library(boot)

#Butstraps
hb1<-function(fx,fy,fh0,fhmin,fhmax,fhsolis,fb,fl){
fn<-length(fx)
m0<-locpoly(fx,fy,bandwidth=fh0,gridsize=fn)
epsilon<-0

e<-((fy-m0$y-mean(fy-m0$y))/(1-1/sqrt(2*pi)/fn/fh0))[(epsilon+1):(fn-epsilon)]
```

```
ne<-length(e)
mh<-array(,c((fhmax-fhmin)/fhsolis+1,2))

for (hh in seq(fhmin,fhmax,by=fhsolis)){
bootfun<-function(bfdati){
ysamp<-m0$y+bfdati[1:fn]
(locpoly(fx,ysamp,bandwidth=hh,gridsize=fn)$y-m0$y)^2}
msamp<-tsbootstrap(e,bootfun,nb=fb,b=fl,type="block",m=1)
mh[hh/fhsolis-fhmin/fhsolis+1,1]<-hh
mh[hh/fhsolis-fhmin/fhsolis+1,2]<-mean(msamp$statistic)}
mh}

#MCV
kod<-function(fx){
1/sqrt(2*pi)*exp(-fx^2/2)}

skj<-function(fx,fj,fl,fk,fdati,fh){
aaa<-c(((fj-fl)*sign(sign(fj-fl)+1)):(fj+fl))
sum(kod((fx-fdati[-aaa])/fh)*(fx-fdati[-aaa])^fk)}

vij<-function(fx,fi,fj,fl,fdati,fh){
kod((fx-fdati[fi])/fh)*(skj(fx,fj,fl,2,fdati,fh)-(fx-fdati[fi])*skj(fx,fj,fl,1,fdati,fh))}

wi<-function(fx,fi,fj,fl,fdati,fh){
fn<-length(fx)
aaa<-c(((fj-fl)*sign(sign(fj-fl)+1)):(fj+fl))
vij(fx,fi,fj,fl,fdati,fh)/(sum(vij(fx,-aaa,fj,fl,fdati,fh))+1/fn^2)}

mxj<-function(fx,fj,fl,fxdati,fydati,fh){
aaa<-c(((fj-fl)*sign(sign(fj-fl)+1)):(fj+fl))
sum(wi(fx,-aaa,fj,fl,fxdati,fh)*fydati[-aaa])}

hmcv<-function(fxdati,fydati,fl,fhmin,fhmax,fhsolis){
mise<-array(,c(2,round((fhmax-fhmin)/fhsolis)+1))
fn<-length(fxdati)
for (j in seq(fhmin,fhmax,by=fhsolis)){
ss<-mean((sapply(1:fn,function(i)(mxj(fxdati[i],i,fl,fxdati,fydati,j)))-fydati)^2)
mise[1,(j-fhmin)/fhsolis+1]<-j
mise[2,(j-fhmin)/fhsolis+1]<-ss
print((j-fhmin)/fhsolis+1)}
mise}

#IDD butstraps
hi<-2
l<-1
h.sadal<-c()
for (i in 1:30)
{
a<-Sys.time()
mise.h<-hb1(x,y,hi,2,50,1,2000,l)
h.sadal[i]<-mise.h[order(mise.h[,2])[1],1]
hi<-h.sadal[i]
b<-Sys.time()
print(c(i,b-a))
}

#Bloku butstraps
hi<-2
l<-5
h.sadal<-c()
for (i in 1:30)
{a<-Sys.time()
mise.h<-hb1(x,y,hi,2,50,1,2000,l)
h.sadal[i]<-mise.h[order(mise.h[,2])[1],1]
hi<-h.sadal[i]
b<-Sys.time()
print(c(i,b-a))}
```

Maģistra darbs ,,Butstrapa metodes analīze atkarīgiem datiem" 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: Mārcis Bratka

____________________ ________________
(paraksts) (datums)

Rekomendēju darbu aizstāvēšanai.
Vadītājs: Dr. Math. Jānis Valeinis

____________________ ________________
(paraksts) (datums)

Recenzents: Dr. Math. Nadežda Siņenko

____________________ ________________
(paraksts) (datums)

Darbs iesniegts Matemātikas nodaļā ________________
(datums)

____
(darbu pieņēma)

Darbs aizstāvēts maģistra darba gala pārbaudījuma komisijas sēdē

________________ prot. Nr. ________, vērtējums________
(datums)

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